



Mémoire de recherche/juin 2015

Diplôme national de master

Domaine - sciences humaines et sociales

Mention - sciences de l'information et des bibliothèques

Spécialité - archives numériques

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

Ariane Néroulidis

Sous la direction de Céline Guyon
Maître de conférence - ENSSIB



Remerciements

Je tiens à remercier tout particulièrement Raphaëlle Bats qui m'a introduite à la notion d'archives participatives et encouragée à poursuivre dans cette voie.

J'adresse également mes remerciements à Véronique Ginouvès et à l'équipe de la phonothèque de la MMSH, qui m'ont soutenue et donné de précieux conseils.

Un grand merci à mes interlocuteurs Véronique Pouyadou, Valéry Vesson, Sandrine Aufray, Hélène Cavalié, Isabelle Josse ainsi qu'à Antoine Courtin et Lionel Maurel qui se sont rendus disponibles pour répondre à mes questions.

Enfin, je remercie ma directrice de mémoire Céline Guyon pour son accompagnement.

Résumé :

Un des impacts de la numérisation des archives est l'intégration des pratiques participatives issues du web social. Le crowdsourcing permet de solliciter les internautes pour l'enrichissement des fonds d'archives, et de développer de nouveaux services à partir de ces métadonnées sociales.

Descripteurs : crowdsourcing, archives participatives, web participatif, métadonnées sociales, sciences citoyennes, usages numériques

Abstract :

One of the impacts of digitalization of archives is the integration of participatory uses coming from the social web. The crowdsourcing enables users to enhance and enrich collections and to develop new services based on these social metadatas.

Keywords : *crowdsourcing, participatory archives, participative web, user-generated content, citizen science, digital practices*

Droits d'auteurs



Cette création est mise à disposition selon le Contrat :
« **Paternité-Pas d'Utilisation Commerciale-Pas de Modification 2.0 France** »
disponible en ligne <http://creativecommons.org/licenses/by-nc-nd/3.0/deed.fr> ou par
courrier postal à Creative Commons, 171 Second Street, Suite 300, San Francisco,
California 94105, USA.

Sommaire

SIGLES ET ABBREVIATIONS	7
INTRODUCTION.....	8
NOTION D'ARCHIVES PARTICIPATIVES	10
Elements de terminologie	10
<i>Human Computation vs Wisdom of Crowds</i>	<i>10</i>
<i>Web collaboratif vs web participatif.....</i>	<i>15</i>
<i>Du crowdsourcing au nichesourcing</i>	<i>18</i>
Comprendre la démarche des archives	21
<i>Politiques des services d'archives</i>	<i>21</i>
<i>Spécificités des fonds d'archives</i>	<i>27</i>
<i>Valeurs des documents d'archives</i>	<i>31</i>
Un intérêt croissant pour les métadonnées sociales	33
<i>Etat des lieux en France.....</i>	<i>33</i>
<i>Publics participatifs.....</i>	<i>38</i>
<i>Forces et faiblesses de l'indexation sociale</i>	<i>41</i>
TYPOLOGIE DES PRATIQUES PARTICIPATIVES	44
Les activités	44
<i>Correction & transcription</i>	<i>44</i>
<i>Contextualisation.....</i>	<i>47</i>
<i>Enrichissement</i>	<i>49</i>
<i>Classification.....</i>	<i>51</i>
Les plateformes.....	53
<i>Les solutions logicielles</i>	<i>54</i>
<i>Les sites de partage de contenu.....</i>	<i>57</i>
<i>Les initiatives en marge des GLAM</i>	<i>60</i>
Les stratégies	63
<i>Des événements ponctuels</i>	<i>63</i>
<i>Une expérience ludique.....</i>	<i>66</i>
<i>Une time-machine</i>	<i>69</i>
LES MOTEURS DE LA PARTICIPATION	72
Du côté des publics, les facteurs de motivation	72
<i>Motivations intrinsèques</i>	<i>72</i>
<i>Motivations extrinsèques.....</i>	<i>74</i>
Du côté des institutions, les facteurs de réussite	76
<i>Ingrédients d'un site réussi</i>	<i>76</i>

<i>Recommandations</i>	78
<i>Retours d'expérience</i>	81
Les enjeux du crowdsourcing	84
<i>Enjeux juridiques</i>	84
<i>Enjeux éthiques</i>	87
<i>Enjeux économiques</i>	90
CONCLUSION	93
SOURCES	94
BIBLIOGRAPHIE	96
TABLE DES ANNEXES	100
LISTE DES ENTRETIENS	106
TABLE DES ILLUSTRATIONS	107
TABLE DES MATIERES	108

Sigles et abréviations

ANDP, Australian Newspaper Digitisation Program

BnF, Bibliothèque nationale de France

CGU, Conditions Générales d'Utilisation

CNIL, Commission Nationale de l'Informatique et Libertés

DEPS, Département des Etudes, de la Prospective et des Statistiques

FEVIS, Fédération des Ensembles Vocaux et Instruments Spécialisés

FSC, Fondation Sciences Citoyennes

GLAM, Galleries, Libraries, Archives and Museums

ICARUS, International Centre for Archival Research

JORF, Journal Officiel de la République Française

LOC, Library of Congress

NARA, National Archives and Records Administration

OCLC, Online Computer Library Center

OCR, Optical Character Recognition

SIAF, Service Interministériel des Archives de France

SNCI, Services Numériques Culturels Innovants

UCL, University College of London

INTRODUCTION

Crowdfunding, crowd labor, crowd voting, crowd collaboration, crowd competition, autant de principes dérivés du crowdsourcing faisant appel à la participation des internautes.

Pris au sens générique, le terme de crowdsourcing (de l'anglais *crowd* qui désigne la foule) n'a pas d'équivalent en français et est traduit par « externalisation ouverte » ou « production participative », suivant le contexte¹. Ce néologisme calqué sur le concept d'*outsourcing* connaît avant tout une acception économique, l'enjeu étant cette valeur ajoutée générée par la foule. Toutefois, on ne saurait limiter le crowdsourcing à cette dimension marchande : à preuve, son émergence dans les institutions culturelles.

Héritières du web 2.0, ces pratiques participatives sont en plein essor dans le secteur des GLAM (Galleries, Libraries, Archives and Museums), pour reprendre la terminologie anglo-saxonne. Parmi ces secteurs culturels, nous nous attacherons au domaine des archives au sens large, en ne nous limitant pas aux seuls services d'archives publics, mais en prenant en compte la gestion des documents d'archives par d'autres institutions.

Dans le premier temps de notre réflexion, nous explorerons le concept d'« archives participatives » tel qu'il a été défini par l'archiviste américaine Kate Theimer² :

« Un organisme, un site ou une collection auxquels des personnes qui ne sont pas des professionnels des archives apportent leur connaissance ou ajoutent des contenus, généralement dans un contexte numérique en ligne. Il en résulte une meilleure compréhension des documents d'archives. »

Une nuance sera également apportée entre les notions de participation et de collaboration.

Après cette phase de définition, nous proposerons une typologie des pratiques participatives, illustrées par des projets de plateforme francophones, mais aussi

¹ « Crowdsourcing — Wikipédia ». Consulté le 9 juin 2015. <http://fr.wikipedia.org/wiki/Crowdsourcing>.

² « Exploring the Participatory Archives | ArchivesNext ». Consulté le 9 juin 2015. <http://www.archivesnext.com/?p=2319>.

internationales, afin de pouvoir établir des points de comparaison. Notons que les Etats-Unis ont été les principaux initiateurs de ces pratiques.

Enfin, dans un troisième temps, nous amorcerons une réflexion sur les moteurs du crowdsourcing en nous plaçant du point de vue des contributeurs pour ce qui est des facteurs de motivation et du point de vue des institutions en ce qui concerne les facteurs de réussite. Nous nous appuierons pour cela sur des entretiens menés auprès d'institutions ayant mis en place de type de projet, afin de pouvoir en tirer un retour d'expérience.

Ce nouveau paradigme nous amènera à nous poser la question suivante : « En quoi l'émergence du crowdsourcing introduit-il un nouveau rapport au document d'archive ? »

NOTION D'ARCHIVES PARTICIPATIVES

Afin de mieux cerner le principe d'archives participatives, il s'agira tout d'abord de maîtriser les notions clés liées à ce concept, mais aussi de bien saisir les spécificités de la démarche des archives, l'objectif étant de comprendre l'intérêt des métadonnées sociales dans ce contexte de production.

ELEMENTS DE TERMINOLOGIE

La notion d'archives participatives permet d'aborder des concepts tels que la sagesse des foules, le concept humanoïde, mais aussi de marquer la différence entre web collaboratif et web participatif, entre crowdsourcing et nichesourcing.

Human Computation vs Wisdom of Crowds

Le crowdsourcing relève-t-il davantage de la puissance de calcul du cerveau humain, ou bien de la sagesse des foules ?

Le calcul humanoïde

Abordons tout d'abord le concept d'Human Computation, que nous traduirons par « calcul humanoïde ». Comme son nom l'indique, le concept d'Human Computation assimile l'être humain à un ordinateur, car il est capable comme une machine, de traiter des données, et d'en tirer des résultats. Pour résoudre des problèmes de grande envergure, la puissance du cerveau humain peut être exploitée à la manière d'un processeur. On peut tirer parti de cette capacité de traitement pour améliorer le traitement de données culturelles. Ce concept de *computing* est exploité dans de nombreuses applications web.

A commencer par les tests de type « CAPTCHA », mis en place afin de se prémunir contre les robots, qu'il s'agisse de *crawlers* (robot d'indexation) ou bien de programmes malfaisants. Au-delà de cette fonction de sécurité, ce test peut être une manière de produire des données. Par exemple, le système *reCAPTCHA*³ a été mis au point par *Google* pour compléter son outil de reconnaissance optique des caractères. En effet, la

³ « reCAPTCHA: Easy on Humans, Hard on Bots ». Consulté le 9 juin 2015.
<https://www.google.com/recaptcha/intro/index.html>.

reconnaissance automatique étant insuffisante, on fait appel à l'œil humain pour déchiffrer des caractères. L'objectif : améliorer la recherche plein texte des collections du *New York Times* et de *Google Books*. *reCAPTCHA* obéit donc à une double fonction car il récolte de l'information en même temps qu'il assure la sécurité de l'internaute : « *reCAPTCHA* est un service d'antibot gratuit qui aide à la numérisation des livres ».

Pour améliorer son moteur de recherche *Google Images*, *Google* a développé en 2006 l'outil *Google Image Labeler*⁴ qui a cessé de fonctionner en 2011. Le principe était le suivant : décrire des images à l'aide de mots-clés dans un temps limité, sachant qu'à chaque descripteur déjà signalé par un internaute, le joueur marquait un point. Ce module était donc pensé sur un mode ludique.

Ce module de *Google* a d'ailleurs inspiré une plateforme de jeux : *Metadata games*⁵, à partir de 40 collections issues de 9 institutions dont la British Library, le Boston Public Library et la Digital Public Library of America. L'objectif étant d'assurer la description la plus fine possible des images.

Le *computing* peut également constituer une activité rémunérée, comme dans le cas du projet d'*Amazon : Mechanical Turk Project*⁶. Les internautes ont le choix entre différentes micro-tâches comme la traduction de textes, la réalisation de sondages ou encore le renommage de fichiers. Ces tâches demandent peu d'expertise mais ne peuvent pas être automatisées, c'est pourquoi elles sont reléguées aux cerveaux humains. Grâce à cette sous-traitance numérique, *Amazon* gagne une main d'œuvre « à la demande ». Du côté du travailleur numérique en revanche se pose le problème de la cyber-exploitation. En effet, le *digital worker* est rémunéré à la tâche et non pas à l'heure. Or en moyenne, pour une tâche qui peut prendre jusqu'à 60 minutes, l'internaute est rémunéré 0,02 euros. On parle alors de *digital sweatshop*, en référence aux ateliers misérables dans les pays en développement. Dans la même idée, il est aussi question de « prolétariat du web ».

L'instrumentalisation des foules sur le web est donc bien connue, tandis que l'engouement pour les *citizen science* se fait de plus en plus fort. Dans le cas des

⁴ « Google Image Labeler - Wikipedia, the free encyclopedia ». Consulté le 9 juin 2015. http://en.wikipedia.org/wiki/Google_Image_Labeler.

⁵ « Metadata Games ». Consulté le 9 juin 2015. <http://www.metadatagames.org/>.

⁶ « Amazon Mechanical Turk - Welcome ». Consulté le 9 juin 2015. <https://www.mturk.com/mturk/welcome>.

sciences citoyennes, ce n'est pas l'appât du gain qui attire, mais la volonté de contribuer à la science. C'est le cas du projet *GalaxyZoo*⁷, qui propose aux astronomes amateurs d'identifier des galaxies diffusées en direct de l'espace. La motivation des publics n'est donc pas forcément liée à la compensation financière, comme nous le verrons dans le troisième temps de notre exposé.

Si la puissance de calcul du cerveau humain n'est plus à prouver, l'émulation des cerveaux entre eux peut encore nous surprendre, c'est en tout cas la thèse avancée par le journaliste américain James Surowiecki.

La sagesse des foules

La théorie de l'« intelligence collective » est défendue par l'économiste James Surowiecki, dans son essai intitulé *La sagesse des foules* publié en 2004⁸. En prenant appui sur des expérimentations, l'auteur va à l'encontre du préjugé selon lequel la foule est synonyme d'ignorance, à l'image des moutons de Panurge. Cette vision négative de la masse est partagée par nombre d'intellectuels du XXe siècle, dont Gustave le Bon qui déclare dans la *Psychologie des foules*⁹ :

« Dans les foules, c'est la stupidité et non la sagesse mère qui s'accumule ».

Chez Gustave le Bon, les foules sont souvent perçues comme « inférieures intellectuellement à l'individu isolé ». James Surowiecki démontre au contraire que le plus grand nombre est souvent à l'origine des meilleures décisions, que ce soit sur le plan politique, économique ou encore du point de vue du management. Pour cela, il s'appuie sur des expériences menées sur les foules, comme celle du statisticien britannique Francis Galton qui avait demandé à un groupe de personnes d'évaluer à vue d'œil le poids d'un bœuf de boucherie. En faisant la moyenne des réponses, il a obtenu le poids exact de la pièce de viande, à une livre près ! Cette anecdote comme tant d'autres nous montre à quel point les prédictions de la foule peuvent tomber juste. En effet, ce phénomène s'explique de manière presque mathématique : en faisant la moyenne des prédictions d'un grand nombre de personnes, les erreurs commises se compensent mutuellement. Tandis que dans la plupart des cas, la moyenne est

⁷ « Galaxy Zoo ». Consulté le 9 juin 2015. <http://www.galaxyzoo.org/?lang=fr>.

⁸ Surowiecki, James. *La sagesse des foules*. Paris: Lattès, 2008.

⁹ Cité p23 in Surowiecki

synonyme de médiocrité, en matière de décision, la moyenne mène presque à l'excellence. Les illustrations de cette théorie sont nombreuses : des paris sportifs aux marchés de pronostics, la foule est en capacité de prédire les résultats avec une exactitude déconcertante.

Pour appuyer sa démonstration, James Surowiecki¹⁰ identifie trois types de problèmes qui peuvent être résolus par la foule : la cognition, qui consiste à trouver une solution définitive à une question complexe ; la coordination, qui vise à agir en fonction du comportement des autres ; la coopération, qui consiste à faire travailler ensemble des personnes aux intérêts personnels divergents.

Pour que la foule fasse preuve de sagesse, il énonce trois conditions. La foule doit se caractériser par une diversité d'opinion, autrement dit, chaque personne doit détenir des informations qui lui sont propres. Elle doit également faire preuve d'indépendance pour ne pas se laisser influencer par les opinions des autres. Enfin, elle doit être décentralisée, afin de s'appuyer sur un savoir local. Ces caractéristiques garantissent une certaine fiabilité des prédictions.

Outre ces conditions intrinsèques, l'auteur met l'accent sur le rôle essentiel joué par l'essor d'internet dans la constitution de cette sagesse collective. En effet, le principe même du moteur de recherche *Google* est démocratique puisqu'il permet d'accéder aux sites les plus cités, grâce à son algorithme *PageRank*. La structure du web est donc anti-hiérarchique par nature : elle n'est dominée par aucune autorité. De plus, elle simplifie l'agrégation d'informations de provenances diverses. Le secret pour une bonne recherche d'information est donc d'élargir au maximum le champ de la requête, plutôt que de suivre des filtres d'informations restrictifs. S'en remettre à la foule d'internautes plutôt qu'à un individu maximise donc les chances de trouver des informations pertinentes.

Toutefois James Surowiecki ne sous-estime pas la valeur de l'expertise car plus un groupe dispose d'informations, plus son jugement collectif est garanti. La foule doit donc compter des experts, pour garantir un certain équilibre. En revanche isoler un décideur par sa qualité d'expert ne garantit pas le succès, d'une part parce que les véritables experts sont difficiles à identifier, d'autre part parce que même ces experts-là

¹⁰ Ibid.

ne sont pas infaillibles. Encore une fois, se fier à une foule de décideurs augmente les chances de prendre une bonne décision.

Si nous devons exclure l'hypothèse d'un décideur unique, il y a paradoxalement besoin d'un leader pour guider les foules. En effet, les foules n'ont pas encore pleinement conscience de leur potentiel, c'est pourquoi il leur faut un guide. L'auteur attire toutefois l'attention sur les risques d'une instrumentalisation de la sagesse collective, avec l'écueil du «prolétariat du web» que nous avons évoqué précédemment.

Une combinaison intelligente

Basé sur l'intelligence collective, le crowdsourcing repose donc à la fois sur la sagesse des foules (*wisdom of crowds*) et la puissance de calcul humain (*human computation*). Dans son blog consacré aux technologies numériques, Trevor Owens¹¹ dresse un comparatif de ces deux concepts récapitulés dans le tableau ci-dessous, traduit en français :

	Human Computation	Wisdom of Crowds
Difficulté de l'outil	Sophistiqué	Simple
Nature des tâches	Structurées	Libres
Durée de l'engagement	Sporadique	Long terme
Interaction sociale	Minimale	Fédératrice
Nature des règles	Techniques	Sociales

Figure 1 : Comparatif entre les concepts Human Computation et Wisdom of Crowds par Trevor Owens

Plus concrètement, le concept d'*Human Computation* trouve son application l'outil *Google Image Labeler*¹², tandis que les principes de *Wisdom of Crowds* sont exploités dans des projets tels que *Wikipédia*.

Le module *Google Image Labeler* est un outil sophistiqué qui nous invite à réaliser des tâches structurées dans un laps de temps court. L'engagement sur la durée est

¹¹ « Human Computation and Wisdom of Crowds in Cultural Heritage | Trevor Owens ». Consulté le 9 juin 2015. <http://www.trevorowens.org/2012/06/human-computation-and-wisdom-of-crowds-in-cultural-heritage/>.

¹² *op. cit.*

quasiment nul, de même que l'investissement social, si bien que le niveau d'interaction est relativement bas.

L'encyclopédie *Wikipédia* au contraire est un outil relativement facile d'accès du point de vue de l'utilisateur. En revanche, la grande amplitude de la tâche à réaliser demande un investissement à long terme, ainsi qu'un engagement sur le plan social. Ce type d'activité est donc fédérateur de liens sociaux.

De par la diversité de ses activités, le crowdsourcing combine donc ces deux aspects, proposant tantôt l'exécution de tâches structurées tantôt des contributions plus libres. Ces tâches peuvent être effectuées par différents types de publics, suivant le temps et les compétences de chacun. En effet, l'engagement est un facteur déterminant, comme nous le verrons lorsque nous analyserons les motivations des usagers.

Web collaboratif vs web participatif

Les expressions de «web collaboratif» et de «web participatif» sont souvent employées l'une pour l'autre : quelle nuance peut-on apporter entre ses synonymes du Web 2.0 ?

Le web 2.0

Le crowdsourcing s'inscrit directement dans la sphère du web 2.0, aussi qualifié de « web collaboratif ». Ce concept a été théorisé par Tim O'Reilly dans son article-manifeste : « What is Web 2.0 »¹³.

Parmi les sept principes énoncés dans ce texte fondateur, on retrouve l'idée de « tirer parti de l'intelligence collective ». A partir des années 2000, se développent de véritables espaces de collaboration sur le Web. Tout d'abord les blogs qui sont des espaces de publication personnels, mais également les wiki, plateformes d'écriture collaborative. Le wiki a été popularisé à travers le projet encyclopédique de *Wikipédia*. Sans oublier les outils de partage -tels que *Flickr* pour les photographies- qui permettent une dissémination des contenus. Enfin notons que les réseaux sociaux numériques sont l'expression de cette entrée dans le web social.

¹³ O'Reilly, Tim. « What Is Web 2.0 - O'Reilly Media », 2005. Consulté le 9 juin 2015.
<http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.

Ces outils divers ont permis l'émergence de nouvelles pratiques participatives dont le crowdsourcing. A cet égard, il serait intéressant d'analyser dans quelle mesure ces plateformes numériques conditionnent les échanges des internautes.

Une « production participative »

Au 15 août 2014, le terme de crowdsourcing est officiellement traduit dans la langue française par l'expression de « production participative ». L'avis est publié dans le Journal Officiel de la République Française (JORF) au numéro 179.

La définition, que l'on retrouve sur Légifrance¹⁴ est la suivante et s'applique à tous les domaines :

« Mode de réalisation d'un projet ou d'un produit faisant appel aux contributions d'un grand nombre de personnes, généralement des internautes. »,

A cette définition s'ajoutent deux notes.

« 1. On peut, par exemple, recourir à la production participative pour concevoir un logiciel ou élaborer une encyclopédie ».

Il s'agit de deux applications possibles du crowdsourcing.

Puis on nous signale un équivalent du terme :

« 2. On trouve aussi l'expression « production collaborative ».
Equivalent étranger : crowdsourcing. »

D'après cette définition, les termes de participation et de collaboration sont introduits comme des synonymes. Or, il convient d'apporter une nuance sémantique entre les deux termes :

Un projet de type collaboratif est un projet pour lequel les contributeurs sont à la fois les concepteurs et les bénéficiaires, comme c'est le cas de *Wikipédia*.

Un projet de type participatif implique deux types d'acteurs : les concepteurs, à l'initiative du projet, et les internautes, qui sont les contributeurs.

¹⁴ « JORF n°0179 | Legifrance », 5 août 2014. Consulté le 9 juin 2015.
<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000029331922>.

En somme, la participation nécessite d'accepter certaines règles préétablies, tandis que la collaboration implique une responsabilité plus importante, en ce qui concerne la gestion et l'accomplissement des tâches : il s'agit par exemple de partager les responsabilités éditoriales d'un blog ou d'un wiki.

La famille des « sciences citoyennes »

L'émergence de la « production participative » dans les institutions culturelles rejoint le concept plus large des « sciences citoyennes », aussi appelées « sciences participatives ».

Ce mode de recherche, qui s'appuie sur la participation citoyenne est en pleine expansion, comme l'a démontré la Fondation Sciences Citoyennes (FSC)¹⁵, dans son rapport paru en avril 2013 sur les pratiques de recherche en sciences participatives. Dans ce rapport, la FSC étudie deux approches en sciences participatives : d'une part les partenariats existant entre la recherche et des regroupements de citoyens et d'autre part la participation du citoyen lambda en dehors de toute organisation.

Dans le premier cas, on remarque que les projets qui sont le fruit de collaboration entre les chercheurs et les acteurs de la société civile répondent bien souvent à une problématique sociétale. C'est le cas par exemple du projet *Observatoire des Saisons*, en partenariat avec des associations comme *Tele Botanica*¹⁶. Les chercheurs ont ainsi pu créer un réseau d'observations phénologiques. Dans son étude la FSC tire la conclusion suivante :

« Ce processus de production de savoirs [...] a mené à une plus grande implication des citoyens dans les orientations de la recherche, notamment concernant ses implications éthiques et environnementales ».

Du point de vue du citoyen lambda, ces initiatives ne sont plus réservées aux seuls chercheurs et se démocratisent peu à peu. A preuve, le projet d'*Observatoire des*

¹⁵ Storup, Bérangère. « La recherche participative comme mode de production de savoirs », FSC, 2011-2012. Consulté le 9 juin 2015. http://sciencescitoyennes.org/wp-content/uploads/2013/04/FSC_final_recherche-participative_FdF.pdf.

¹⁶ « ODS | Observatoire des Saisons ». Consulté le 9 juin 2015. <http://www.obs-saisons.fr/>.

*papillons de jardin*¹⁷, qui vise à étudier l'évolution de la diversité à l'échelle de son jardin. Le citoyen est alors mis au centre de la recherche. Dans son ouvrage intitulé *Des sciences citoyennes ? La question de l'amateur dans les sciences naturalistes*¹⁸, Florian Charvolin met en avant cette révolution dans l'histoire des sciences :

« La Science participative nous invite à décentrer nos façons de penser la science traditionnelle ».

Si l'objet de notre étude ne concerne pas la portée scientifique du crowdsourcing, ce concept de décentralisation constituera un élément clé dans notre réflexion. En effet, dans le cas des archives participatives, la relation entre usagers et producteurs d'archives se retrouve profondément bouleversée, d'où l'idée d'un nouveau rapport au document d'archives.

Du crowdsourcing au nichesourcing

Interrogeons-nous à présent sur le terme de crowdsourcing : est-il vraiment approprié, ou bien vaut-il mieux parler de nichesourcing ?

Une terminologie discutable

Le terme de « crowdsourcing » qui est un mot composé, pose une double question.

Tout d'abord la notion de *crowd* qui désigne la foule n'est pas toujours appropriée dans le sens où les institutions font souvent appel à des publics initiés, et non à des foules anonymes. En effet, on se rend compte que les principaux intéressés sont déjà familiers de ces institutions. Il s'agit donc davantage d'un groupe de volontaires motivés que de foules à proprement parler. L'intérêt personnel pour ces collections constitue souvent un facteur de motivation. Par exemple, le projet *What's on the Menu ?*¹⁹ initié par la New York Public Library attire de nombreux gourmets, qu'il s'agisse de chefs étoilés ou d'amateurs de bonne chère. La mission consiste à transcrire des menus de restaurants new-yorkais, depuis 1840 jusqu'à nos jours. Notons que cette activité

¹⁷ « NOÉ biodiversité | Observatoire des Papillons des Jardins ». Consulté le 9 juin 2015.

<http://www.noeconservation.org/index2.php?rub=12&srub=31&ssrub=98&goto=contenu>.

¹⁸ Charvolin, Florian. *Des sciences citoyennes ? La question de l'amateur dans les sciences naturalistes*. Paris: Editions de l'Aube, 2007.

¹⁹ « New York Public Library | What's on the menu? ». Consulté le 9 juin 2015. <http://menus.nypl.org/>.

existait déjà avant l'avènement du web 2.0, puisque les volontaires cataloguaient déjà ces menus dans la salle de lecture des archives. La notion de contribution n'est pas un phénomène nouveau. On note son apparition dans le dictionnaire Robert en 1580, avec pour définition : « collaboration à une œuvre commune »

En ce sens, le terme de « sourcing » pose aussi question puisqu'il assimile ces activités à du travail (*labor*).

La définition donnée par *Wikipédia*²⁰ va dans ce sens :

« Le crowdsourcing est un processus qui implique l'externalisation des tâches vers un groupe de personnes ».

Or, cette valeur économique n'a pas de sens dans le contexte des institutions culturelles, qui ne tirent pas de profit de ce type d'activités. La participation des publics relève donc davantage du volontariat, entendu comme un don de sa personne et de son temps. A la manière des bâtisseurs de cathédrales, chacun apporte sa pierre à l'édifice de la connaissance.

Le crowdsourcing dans le contexte des GLAM est donc caractérisé par un certain désintéressement. Il est avant tout une affaire d'amateurs, au sens originel du terme -du latin « amare » qui signifie aimer-. Il n'est donc pas question de foules anonymes mais bien de publics volontaires, animés par la passion et un profond désir d'œuvrer pour le bien commun.

Culture snackers vs culture vultures

Au sein de ces publics volontaires, nous pouvons introduire une distinction entre deux catégories de contributeurs : les experts d'un côté et le grand public d'un autre.

Dans son étude "Design for User Engagement on Europeana Channels"²¹, Chenchen Shen propose deux appellations : la catégorie des « culture snackers » et la catégorie des « culture vultures ». Tandis que le terme « snackers » renvoie à l'habitude de grignoter, le mot « vulture » peut être associé soit au vautour, soit au vampire. Ces deux images renvoient donc à deux modes de consommation : le premier à petites doses et le second

²⁰ *op. cit.*

²¹ Shen, Chenchen. « Design for User Engagement on Europeana Channels ». *Delft University and Europeana, Master of Science Graduation Project-Design for Interaction*, 2014.

en grosse quantité. Il en va de même pour les publics, pour certains assoiffés de culture et pour d'autres consommateurs occasionnels. Dans son étude, C.Shen²² dresse les profils de ces deux catégories de personnes :

Les « Culture snackers » :

« Ils aiment consulter des contenus culturels sur internet et les partager avec leurs amis et followers, mais cette recherche le patrimoine culturel est juste une part de leur consommation quotidienne (...) Pour eux, le côté plaisant de l'expérience a parfois plus d'importance que la qualité du contenu ».

D'après l'étude menée par C.Shen, 49% des participants ont souligné leur désir de « partager leur savoir, leur expertise et leurs idées », tandis que d'autres ont été intéressés par « créer ou participer à un sujet de discussion », « traduire des contenus dans leur langue » et « recommander des contenus en lien ».

Les « Culture vultures » :

« Ils sont professionnels dans la culture ou passionnés. Ils ont un grand intérêt pour le patrimoine culturel et certainement des connaissances dans un certain domaine. Ils travaillent probablement dans le secteur culturel, ou sont depuis toujours passionnés de culture. Il peut s'agir de chercheurs, d'étudiants, de professionnels et de non-initiés. »

Les experts ne sont donc pas nécessairement des professionnels du domaine, et peuvent être de simples amateurs, tels que nous les avons définis précédemment.

Ces deux publics pourront donc être sollicités à différents niveaux, selon le degré de compétences exigé.

Une stratégie de niche

Pour impliquer un public de type « culture vulture », le crowdsourcing doit s'appuyer sur une stratégie de niche que l'on nomme « nichesourcing ».

Selon une étude menée par Schroer et Hertel sur l'engagement volontaire au sein de la communauté de *Wikipédia*, la définition du nichesourcing est la suivante :

« Le nichesourcing est un type spécifique de crowdsourcing où les tâches complexes sont confiées à communauté restreinte d'experts-amateurs,

²² Ibid.

plutôt qu'à une foule anonyme. Une niche est constituée soit d'experts spécialistes d'un sujet donné, soit d'un réseau préexistant fondé autour d'une culture, d'un lieu ou d'un sujet commun. Dans les deux cas, les membres possèdent un domaine de connaissance et une motivation intrinsèque qui les poussent à contribuer et produire des résultats de haute qualité ».²³

La distinction entre crowdsourcing et nichesourcing se situe au niveau des tâches et de leur complexité. En somme, le crowdsourcing consiste en d'importantes tâches complexes, qu'il s'agit de diviser en sous-tâches. Les sous-tâches ne nécessitant pas de savoirs ou de compétences particulières peuvent alors être prises en charge par des individus appartenant à la foule (*crowd*). Les sous-tâches qui exigent un certain niveau de connaissances peuvent être déléguées à des niches, dont les membres possèdent une expertise dans un domaine.

Enfin, il ne faut pas négliger l'effet de vase clos de ces communautés de niche, qui constituent de véritables viviers de connaissance (en anglais *resource pool of users*) En effet, les membres peuvent tirer parti de leurs propres relations pour attirer de nouveaux contributeurs. Tout l'enjeu est donc d'impliquer ces communautés et de maintenir leur intérêt dans le temps.

Le nichesourcing est donc bien un aspect du crowdsourcing mais ne s'y limite pas puisqu'il ne prend en considération qu'une catégorie minoritaire de la foule.

COMPRENDRE LA DEMARCHE DES ARCHIVES

Afin de bien comprendre l'intérêt des métadonnées sociales, il nous faut tout d'abord comprendre le positionnement des archives en considérant les politiques des services d'archives, les spécificités des fonds ainsi que les valeurs des documents.

Politiques des services d'archives

Politique des publics

Pour analyser la politique des publics dans les services d'archives, nous pouvons nous appuyer sur l'enquête de la sociologue Laure Ciosi datant de 2012²⁴. Son rapport a été

²³ Schroer, Joachim, et Guido Hertel. « Voluntary engagement in an open web based encyclopedia: Wikipedians, and why they do it ». *Media Psychology* 12, 1–25, 2009.

²⁴ Ciosi, Laure. *Politique des publics dans les services d'archives*. CIAF. Paris, 2013.

publié par le SIAF, Service Interministériel des Archives de France. Cette étude porte sur la politique des publics et ses partenariats fonctionnels dans le réseau des archives municipales, départementales et régionales en France métropolitaine. Parmi les positionnements des services d'archives, nous nous intéresserons plus particulièrement au nouveau rapport au public instauré par Internet. En effet, l'outil numérique est une révolution pour les services d'archives, comme l'atteste ce directeur d'archives départementales²⁵ :

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

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

En ce qui concerne les modalités participatives, elle laisse certains archivistes perplexes, comme ce directeur d'archives départementales²⁶ qui reconnaît ses réticences dans un entretien :

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

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

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

²⁵ Les entretiens sont anonymisés : aucun nom n'est cité

²⁶ Idem

image, une anecdote historique, un complément historique, enfin tout ce que l'on veut »

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

Pratiques des publics

Après avoir analysé la politique des services d'archives, interrogeons à présent les pratiques des publics, et tout particulièrement leur rapport aux services numériques. Pour cela, nous prendrons appui sur l'*Enquête nationale auprès des publics dans les services d'archives* publiée en 2014²⁷. Cette étude a été menée par les Archives de France, en collaboration avec le Département de la politique des publics de la Direction générale des patrimoines. Il s'agit d'une triple enquête qui s'est déroulée de septembre 2013 à janvier 2014. Elle a été soumise à trois publics distincts : les internautes, les lecteurs en salle et les visiteurs des Journées Européennes du patrimoine édition 2013. Ces échantillons nous permettent d'observer une diversité des pratiques.

En premier lieu, il s'agit de déterminer si les internautes fréquentant les sites internet des institutions côtoient aussi les salles de lectures des archives. Selon l'étude, il semblerait que plus de la moitié des internautes (63%) ne fréquentent pas physiquement le service d'archives. A côté de cela, la grande majorité des internautes interrogés (89,4%) ont consulté d'autres sites d'archives au cours des 12 derniers mois. Cette démarche dénote donc un certain intérêt pour les services proposés en ligne.

De manière assez naturelle, la majeure partie des lecteurs interrogés est constituée d'habitues : toutefois 26,3% se déplacent pour la première fois pour consulter dans la salle de lecture. Parmi les habitués, plus de la moitié se rendent aux archives plusieurs fois par an (à 57,5%) voir plusieurs fois par mois (à 27,9%).

Enfin, pour ce qui est des visiteurs dans le cadre des Journées du Patrimoine, il s'agit bien souvent de « primovisiteurs », puisque 77,8% visitent pour la première fois un service d'archives. Pour la plupart, la visite a répondu à leurs attentes et ils envisagent pour certains de revenir, que ce soit pour consulter des documents en salle de lecture (à

²⁷ Guigueno, Brigitte, et Anne Jonchery. « Enquête nationale sur les publics dans les services d'archives ». Paris: SIAF, 2014.

50,4%) ou bien participer à une activité (à 26,8%). Ces personnes sont donc potentiellement intéressées par des activités collaboratives autour des archives.

En ce qui concerne les activités proposées sur le site des institutions, sans surprise, ce sont les internautes qui sont les plus actifs en ce qui concerne l'enrichissement de contenus -indexation et commentaires collaboratifs- avec 23,3% de participation. Les lecteurs quant à eux atteignent un taux de participation de 12,6%. Enfin, les visiteurs qui ne sont pas des usagers réguliers atteignent un score encourageant de 7,1%.

L'intérêt des publics pour l'enrichissement de contenus est donc bien réel, malgré les craintes de certains directeurs d'archives.

Cadre légal

D'après les derniers chiffres, communiqués par les Archives de France en avril 2015 (voir figure 2), les instruments de recherche en ligne dans les services d'archives publics ont tendance à se généraliser, grâce aux opérations d'inventaire et d'état des fonds.

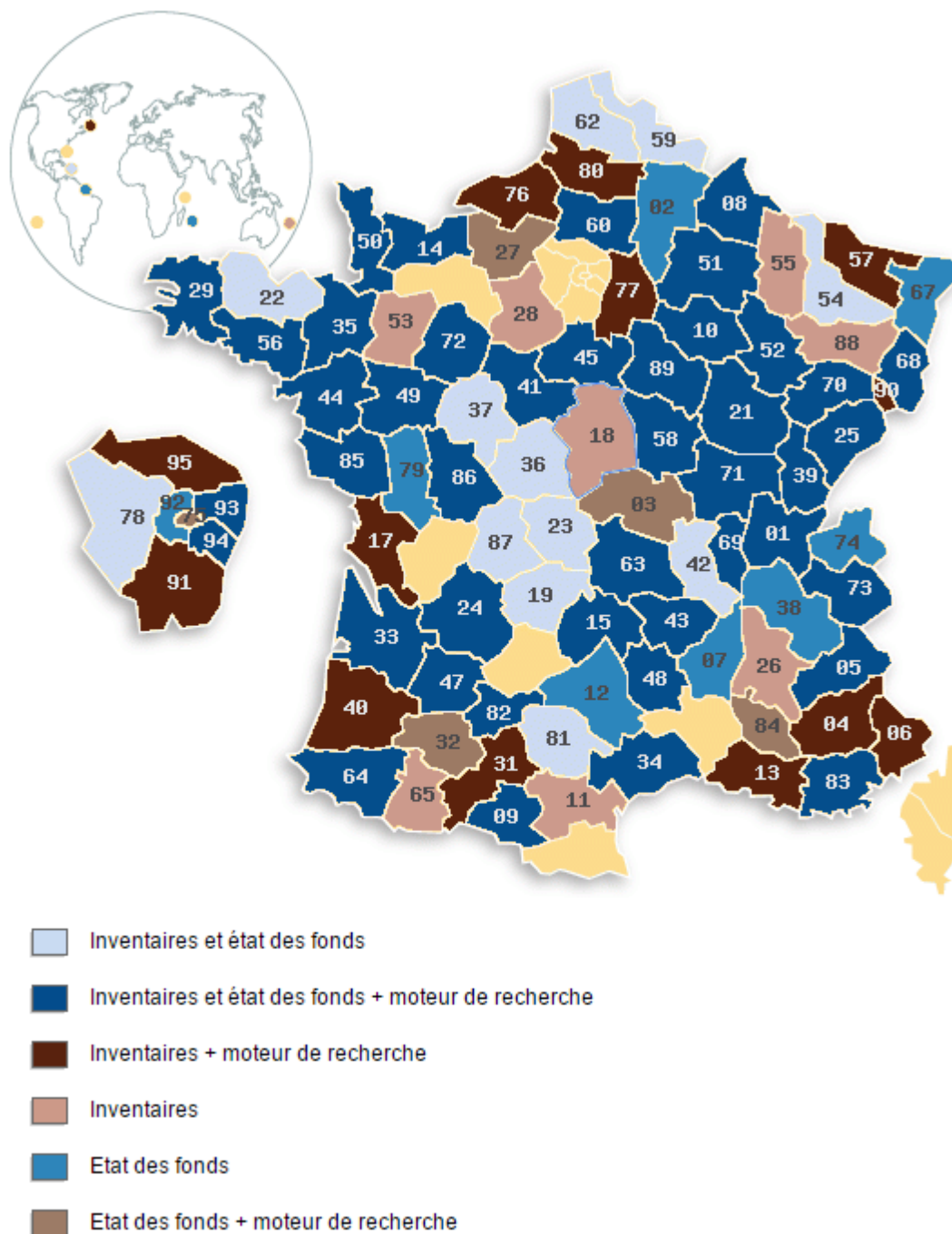


Figure 2 Cartographie des instruments de recherche en ligne dans les services d'archives publiques, réalisée par les Archives de France (avril 2015)²⁸

Or ces services publics d'archives sont soumis à la législation, en ce qui concerne la mise en ligne des documents dans un premier temps, et l'indexation par des moteurs de recherche dans un deuxième temps.

²⁸ « Archives de France | Carte des instruments de recherche en ligne ». Consulté le 10 juin 2015. <http://www.archivesdefrance.culture.gouv.fr/chercher/organiser/carte-ir/>.

En effet, les délais de diffusion ne sont pas les mêmes suivant les deux cas. A noter que ces délais sont différents des délais de communicabilité des documents en salle de lecture, définis par le Code du patrimoine (article L.213-2)²⁹.

Les délais concernant la mise en ligne sur internet ont été définis par la délibération 2012-113 du 12 avril 2012 par la Commission Nationale de l'Informatique et Libertés (CNIL)³⁰ portant « autorisation unique de traitements de données à caractère personnel contenues dans des informations publiques aux fins de communication et de publication par les services d'archives publiques ». Ce texte tend à protéger les données personnelles définies comme « relatives à la vie privée », contrairement aux données dites « sensibles », qui sont :

« Des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses, ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci (art. 8 de la loi 78-17 du 16 janvier 1978).³¹ »

Pour l'indexation des données nominatives, le délai est donc plus long que pour la simple diffusion des documents. Cette précaution s'explique par la finalité de l'indexation, qui facilite l'accès aux données personnelles. L'objectif de la CNIL est donc de veiller à protéger les informations de toute personne vivante.

²⁹ « Code du patrimoine - Article L213-2 | Legifrance ». Consulté le 9 juin 2015.

<http://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006074236&idArticle=LEGIARTI000020566964>.

³⁰ « Délibération n° 2012-113 du 12 avril 2012 de la CNIL | Legifrance ». Consulté le 9 juin 2015.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000025753449>.

³¹ « Loi n° 78-17 du 6 janvier 1978 | Legifrance ». Consulté le 9 juin 2015.

<http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT00000886460>.

Suivant le type de documents et leur contenu, les délais légaux sont les suivants :

Documents	Mise en ligne	Indexation
Etat civil : naissances	100 ans	120 ans
Etat civil : mariages	75 ans	100 ans
Etat civil : décès	25 ans	75 ans
Autres données personnelles	100 ans	120 ans
Données sensibles	100-150 ans	A définir

Figure 3 : Les délais de diffusion et d'indexation applicables à certains documents mis en ligne par les services publics d'archives

D'un point de vue juridique, en France, les activités d'indexation collaborative ne sont pas rendues possibles dès la mise en ligne des documents.

Spécificités des fonds d'archives

Afin de mieux comprendre le traitement collaboratif des archives numériques, il nous faut tout d'abord saisir la nature des documents d'archives à travers ses spécificités. Pour cela nous nous appuyerons sur l'exposé de Pauline Moirez, spécialiste en France de la notion d'« archives participatives »³².

La masse de documents

L'une des caractéristiques première des documents d'archives est leur volume : c'est pourquoi on parle en « kilomètres d'archives ». La priorité des services d'archives à l'ère du numérique est donc la mise en ligne de documents. Tandis que les institutions culturelles telles que les musées ou les bibliothèques vont plutôt miser sur des politiques de médiation, les services d'archives eux investissent dans la numérisation de leurs fonds.

Cette entreprise est toutefois de taille, si l'on calcule le temps nécessaire pour numériser ne serait-ce qu'une partie d'un fonds. A titre d'exemple, on estime que pour

³² Moirez, Pauline. « Archives participatives ». Bibliothèques 2.0 à l'heure des médias sociaux, dir. Muriel Amar et Véronique Mesguich, Editions du Cercle de la Librairie, p. 187-97. Paris, 2012.

numériser l'équivalent de 30 kilomètres d'archives à un rythme de 10 000 pages par semaine, il faudrait 406 ans pour en venir à bout! Une solution consiste à solliciter les publics à travers un appel à contribution.

En 2012, les Archives Nationales des Etats-Unis se sont associées à *Wikimedia* pour organiser des séances de numérisation collectives sur plusieurs jours, dans le cadre d'un événement baptisé « ExtraSCANza »³³. Le public peut également être sollicité pour effectuer un travail de curation, qui consiste à sélectionner les fonds à numériser, comme cela avait été proposé par les Archives Municipales d'Amsterdam dans le cadre du projet « You ask, we scan »³⁴ en 2009. En effet, les services d'archives ne peuvent pas se permettre de numériser l'ensemble de leurs fonds, c'est pourquoi ils doivent réaliser un choix, ici confié aux publics.

A l'heure actuelle, les archives départementales concentrent leurs efforts sur la mise en ligne des registres paroissiaux et d'états civils réalisés dans la quasi-intégralité des départements (cf figure 4)

Au 13 août 2013, les Archives de France ont recensé 281 millions de pages d'archives en ligne et 11 millions d'images en ligne, ce qui aurait attiré 50 millions de visites et plus de 2 milliards de pages vues sur les sites des services d'archives. Parmi ces 50 millions de visites, plus de 40 000 concernent les sites des archives départementales.

³³ « Wikipedia:Meetup/NARA 4 - ExtravaSCANza ». Consulté le 9 juin 2015.
http://en.m.wikipedia.org/wiki/Wikipedia:Meetup/NARA_4.

³⁴ The Amsterdam City Archives and the Archiefbank. « You Ask We Scan ». Consulté le 9 juin 2015.
<http://fr.slideshare.net/ktheimer/you-ask-we-scan-amsterdam-city-archives-high-res>.

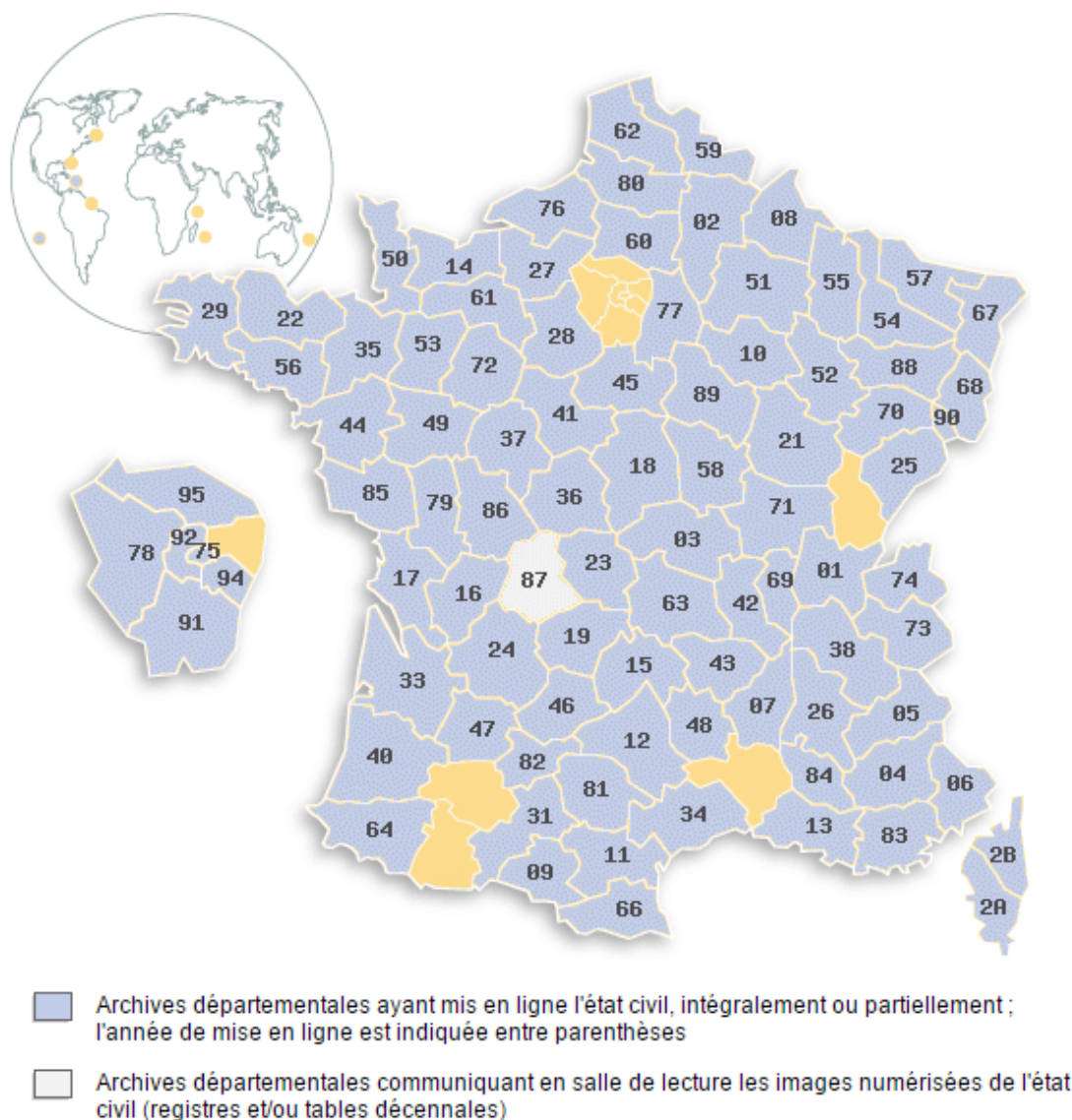


Figure 4 Cartographie de la numérisation et mise en ligne des registres paroissiaux et état civil dans les services d'archives publiques, réalisée par les Archives de France (avril 2015)³⁵

Unicité

Au-delà de leur volume, les documents d'archives se démarquent par leur caractère unique. En effet, contrairement aux ouvrages qui sont disponibles en plusieurs exemplaires dans des bibliothèques, un document d'archives n'est détenu que par une

³⁵ « Archives de France | Etat civil en ligne ». Consulté le 10 juin 2015. <http://www.archivesdefrance.culture.gouv.fr/ressources/en-ligne/etat-civil/>.

seule institution. Rappelons ici la définition des archives, tel qu'il a été défini par le Code du Patrimoine dans l'article L211-1³⁶ :

« Les archives sont l'ensemble des documents, quels que soient leur date, leur forme et leur support matériel, produits ou reçus par toute personne physique ou morale et par tout service ou organisme public ou privé dans l'exercice de leur activité. »

Le document d'archive est donc unique, car il est le produit d'une activité à un instant t.

C'est pourquoi la dimension originale du document est primordiale, et qu'il faut pouvoir s'assurer de la fiabilité, de l'authenticité et de l'intégrité du document.

Ce caractère unique du document rend son traitement fastidieux, puisqu'il ne peut pas être répertorié dans un catalogue collectif. Dans la pratique, les documents d'archives papier sont décrits au niveau de l'« article », qui correspond généralement au dossier ou au registre. Or, ce niveau de granularité médian, adapté aux requêtes en salle de lecture, n'est pas suffisant pour une recherche sur le web. En effet, pour mettre en place un moteur de recherche, il faut pouvoir accéder à des informations à un degré plus fin. Or ce travail d'indexation est très chronophage, ce qui rend difficile la description exhaustive des documents, quelle que soit leur nature. Si nous prenons le cas particulier des archives sonores, on compte environ quatre heures pour traiter une heure d'enregistrement. L'unité minimale de description est l'item, puis on distingue l'enquête, le corpus et le fonds : au total, nous obtenons donc quatre niveaux d'analyse. Au niveau textuel, la description exhaustive du document passe par la transcription du contenu, ce qui permet la recherche plein texte. Les activités collaboratives afin de permettent donc de décrire les documents au niveau de la pièce, et non plus de l'article.

Numérisés

Le format numérique permet de nouveaux usages. L'environnement numérique apporte de nouvelles perspectives aux services d'archives, qui peuvent rendre accessibles leurs documents. Les documents, une fois sortis de leurs boîtes d'archives, doivent alors être recontextualisés, mis en perspective les uns par rapport aux autres. Le

³⁶ « Code du patrimoine : LIVRE II ARCHIVES | Legifrance ». Consulté le 9 juin 2015.
<http://www.legifrance.gouv.fr/affichCode.do?idArticle=LEGIARTI000006845559&idSectionTA=LEGISCTA000006159940&cidTexte=LEGITEXT000006074236&dateTexte=20080716>.

format numérique des archives fait donc émerger de nouvelles pratiques, permettant aux internautes de se les réapproprier et de leur donner une nouvelle vie. Notons quelques réutilisations originales de documents d'archives sur les médias sociaux :

Un bel exemple est le blog *Orwell Diaries*³⁷ qui reprend le journal de George Orwell depuis 1938. Chaque jour, un nouveau billet est publié, qui correspond au texte écrit par l'auteur 70 ans auparavant. Ces publications basées sur des documents d'archives permettent donc de faire revivre la personnalité à travers les moyens d'expression modernes. Sur le même principe, le *Massachusetts Historical Society* a utilisé *Twitter* pour publier les mémoires de John Quincy Adams, ambassadeur des Etats-Unis en Russie au début du XIXème. Ce mode d'expression via le *tweet* crée donc un décalage. Les réseaux sociaux peuvent même aller jusqu'à donner la parole à des archives, comme ça a été le cas en 2009 sur le compte @Schaghenletter créé par les Archives Nationales des Pays-Bas, permettant de suivre les aventures d'un document prêté pour exposition. Les réseaux sociaux tels que *Twitter* sont donc investis par les services d'archives pour créer une nouvelle relation aux usagers, et encourager la création de communautés. C'est sur ces communautés qu'ils pourront ensuite s'appuyer pour développer des projets collaboratifs.

Valeurs des documents d'archives

Les documents d'archives véhiculent également un ensemble de valeurs : valeur historique, valeur administrative et citoyenne, et valeur pédagogique.

Valeur historique

La valeur historique est intrinsèque au document d'archives, qui a acquis au cours du temps une dimension patrimoniale, en entrant dans son troisième cycle de vie. En cela, l'archives se distingue chez les anglo-saxon du *record*, qui bénéficie d'une valeur probante.

³⁷ « Orwell Diaries 1938-1942 ». Consulté le 9 juin 2015. <https://orwelldiaries.wordpress.com/>.

Le document d'archives est le témoin d'une époque et de son contexte de production, d'où l'importance de maîtriser les codes de cette époque et tout particulièrement la graphie.

Les archives historiques peuvent être contextualisés à l'aide de témoignages, qui apportent un éclairage nouveau sur une période donnée. Nous pouvons aussi donner l'exemple du site des Archives Nationales d'Australie, *Destination Australia*³⁸, qui rassemblent des récits de migrants après la guerre. Ces récits écrits sont illustrés par des photographies, ce qui donne un visage aux migrants. Ces collectes de témoignages sont donc des projets collaboratifs, en ce qu'ils impliquent des volontaires qui acceptent de livrer leurs récits de vie. La petite histoire rejoint alors la grande Histoire comme le suggère la devise du projet : « Your story, our history ».

Valeur citoyenne

Outre la dimension historique souvent associée aux documents d'archives, les archives peuvent aussi être à l'origine de projets citoyens.

Aux Etats-Unis tout particulièrement émerge le concept du *citizen archivist*. Les Archives Nationales encouragent « l'archiviste citoyen » à s'investir dans des activités collaboratives diverses, proposées sur le « tableau de bord » de son site institutionnel. Cette interface est appelée « Citizen Archivist Dashboard »³⁹ et propose des missions diverses comme la transcription de documents sur le site ou Wikisource, la rédaction d'articles scientifiques sur le wiki *Our Archives*, le *tagging* sur *Flickr*. Le crowdsourcing est ici présenté comme un acte citoyen.

Un projet des Archives Nationales du Royaume-Uni, a permis d'aider le travail des météorologues : *Old Weather*⁴⁰. La mission du citoyen : transcrire et géolocaliser des relevés météorologiques manuscrits, réalisés par la Marine royale au début du XXème siècle. Ces données météorologiques ont permis aux scientifiques de mieux appréhender les évolutions du climat durant deux siècles. En ce sens, ce projet relève aussi des sciences citoyennes.

³⁸ « Destination: Australia - Sharing our post war migrant stories ». Consulté le 9 juin 2015. <https://www.destinationaustralia.gov.au/site/>.

³⁹ « Citizen Archivist Dashboard ». Consulté le 9 juin 2015. <http://www.archives.gov/citizen-archivist/>.

⁴⁰ « Old Weather - Our Weather's Past, the Climate's Future ». Consulté le 9 juin 2015. <http://www.oldweather.org/>.

Valeur éducative

Cette dimension citoyenne va de pair avec la valeur éducative des archives, qui peuvent donner lieu à des activités ludiques, à destination des enfants mais pas exclusivement. On note le développement de jeux interactifs qui permettent un apprentissage ludique auprès des jeunes publics.

Aux Archives municipales de Saint Denis, l'atelier pédagogique *Copains de Banlieue*⁴¹, permet de découvrir l'histoire de la banlieue nord de Paris au XXème siècle, à travers 7 petites fictions animées. Ce jeu a obtenu le prix des « Adélouzes Online » en 2006, à l'occasion de la cinquième édition du festival Les Adélouzes⁴².

Les Archives départementales de l'Aube ont développé le *Mystère de la Cordelière*⁴³, un jeu mettant en scène des *Archiphages*, créatures se nourrissant d'archives. Le but du jeu est de résoudre des énigmes, posées dans chaque pièce du manoir, et dont les réponses se trouvent dans des documents d'archives. Ces ateliers interactifs permettent d'appréhender les archives sous un angle plus ludique, ce qui suscite l'intérêt des jeunes internautes.

Ces valeurs fondamentales inhérentes aux archives permettent donc d'imaginer des pratiques participatives innovantes.

UN INTERET CROISSANT POUR LES METADONNEES SOCIALES

Etat des lieux en France

En 2013, un « Etat de l'art en matière de crowdsourcing » a été produit par la BnF⁴⁴, en vue d'un projet de plateforme, et dresse un bilan des pratiques dans les bibliothèques numériques.

⁴¹ « Copains de banlieue ». Consulté le 9 juin 2015. <http://www.copainsdebanlieue.com/gene.html>.

⁴² Prix multimédia jeunesse

⁴³ « Le Mystère de la Cordelière ». Consulté le 9 juin 2015. <http://www.archives-aube.fr/a/155/le-mystere-de-la-cordeliere/>.

⁴⁴ Moirez, Pauline, Jean-Philippe Moireux, et Isabelle Josse. « État de l'art en matière de crowdsourcing dans les bibliothèques numériques ». Bibliothèque nationale de France, 2013.

Participation des Archives Départementales

Pour estimer le degré de participation des services d'archives en France dans les activités de crowdsourcing, nous prendrons comme échantillon représentatif les Archives Départementales. Nous nous appuyerons pour cela sur les chiffres fournis par Edouard Bouyé, Directeur des Archives départementales du Cantal. En 2012 il publie un article intitulé « Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics »⁴⁵ où il recense les services d'archives impliqués dans la dynamique du web collaboratif, tout en dressant un bilan encourageant de ces initiatives.

En 2012, on compte 19 services d'archives départementales proposant un module d'indexation collaborative : 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. Début 2015, selon notre recensement, 25 archives départementales ont intégré l'indexation collaborative à leurs pratiques et trois autres institutions proposent une autre activité de type collaborative. Au total, plus d'un quart des institutions françaises sont concernées :

⁴⁵ Bouyé, Edouard. « Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics ». *La Gazette des Archives*, Association des Archivistes français, n° 227 (2012).

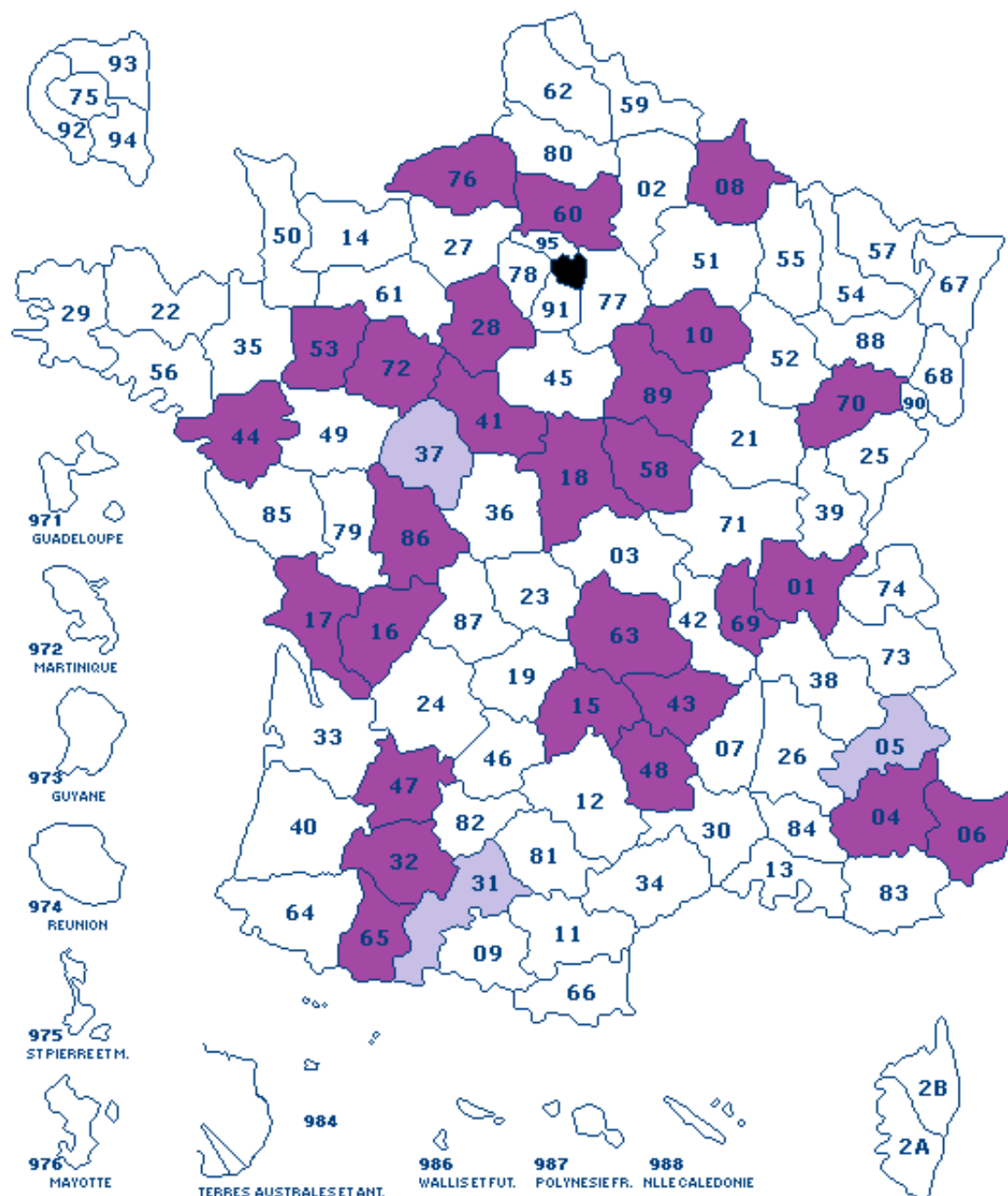


Figure 5 : Archives départementales pratiquant l'indexation collaborative début 2015⁴⁶

Nature des fonds concernés

L'étude menée par Edouard Bouyé porte également sur la nature des fonds d'archives faisant l'objet d'indexation collaborative. Sachant qu'un même service d'archives peut exploiter plusieurs de ses fonds, on observe la répartition suivante :

⁴⁶ Réalisé à partir d'un fond de carte téléchargé sur <http://coloriages.ws/?s=D%C3%A9partements>

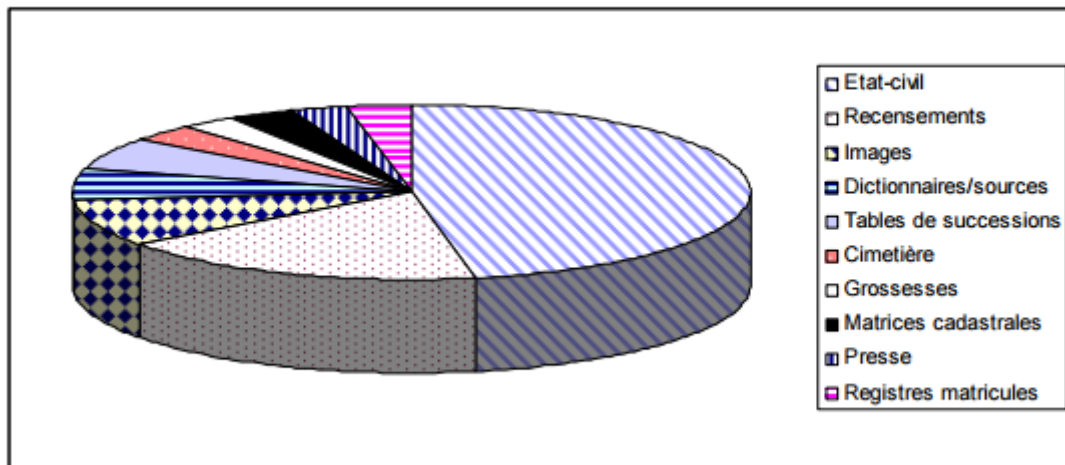
Fonds faisant l'objet d'indexation collaborative en janvier 2012

Figure 6 : Fonds d'archives faisant l'objet d'indexation collaborative en janvier 2012 (diagramme réalisé par Edouard Bouyé)

En première position, on trouve les registres d'état civil, qui sont les fonds numérisés prioritairement par les services d'archives. Dans l'étude, ces registres représentent seize fonds, soit presque 50% des fonds indexés par les internautes. En effet, les informations fournies par l'état civil sont précieuses car elles permettent de retrouver des actes de naissance, de mariage et de décès remontant jusqu'en 1539. L'état civil constitue donc une mine de renseignements pour les généalogistes.

En deuxième position, on trouve les listes de recensement, qui sont également des documents nominatifs. Dans l'étude, ces listes représentent six fonds, soit 18% des fonds indexés par les internautes. Les listes de recensement ont été établies de manière systématique depuis 1836, à intervalle de 5 ans. Elles indiquent pour chaque commune les noms et prénoms des habitants, leurs professions, leur place dans le ménage, et comportent, selon les époques, des informations complémentaires telles que l'âge, l'année et/ou lieu de naissance, la nationalité de la personne. En 1851, on y trouve même la religion et les infirmités éventuelles. Ces informations permettent de mener des études sociologiques sur une population donnée.

En troisième position, on trouve les images, qui servent de matériau à de nombreux projets de crowdsourcing. Dans l'étude, ces images représentent trois fonds, soit 9% des fonds indexés par les internautes. En effet, une des activités collaboratives consiste en l'identification de photographies. La nature de ces images est diverse : il peut s'agir de cartes postales, de photographies de lieux ou de personnes, qui proviennent de dons de particuliers ou bien de l'héritage des institutions.

Enfin, l'indexation collaborative peut concerner -dans une moindre mesure- d'autres catégories de documents : dictionnaires, tables de succession, cimetière, grossesses, matrices cadastrales, presse et registres matricules.

Des services culturels innovants

Le Ministère de la Culture et de la Communication s'inscrit depuis 2010 dans une démarche de soutien à l'innovation numérique en encourageant le développement de nouveaux usages culturels numériques. En 2014, 60 projets ont été sélectionnés suite à l'appel à projet SNCI, « Services numériques culturels innovants »⁴⁷. Ces projets impliquent une grande variété d'acteurs, dont des services d'archives. Parmi ces services innovants, on relève de nombreux projets basés sur le principe de la contribution. Nous avons sélectionné trois projets de type crowdsourcing, dans trois domaines produisant des documents d'archives. Dans le secteur des musées, notons le projet des *Muséonautes*, lancé par un regroupement de 12 musées de Normandie⁴⁸. Il s'agit de créer un outil de partage collaboratif et interactif entre les internautes et les institutions culturelles. En effet, la *Museobase* permet désormais de contribuer à l'enrichissement des données relatives aux images et objets des collections, organisées en parcours thématiques. Au niveau musical, la Fédération des Ensembles Vocaux et Instruments Spécialisés (FEVIS)⁴⁹, en collaboration avec 40 ensembles musicaux, propose une ontologie musicale et indexation sémantique à partir d'*OMNI*, Objets Musicaux Non Identifiés. Enfin, dans le domaine du spectacle vivant, on relève le projet mené par la Scène nationale de Valenciennes : *Rekall-online*⁵⁰. Cette interface permettrait aux internautes d'annoter des captations vidéo de spectacle, agrégeant différents types de documents liés à l'œuvre (photos, croquis, plans, vidéos, sons). Un module de « spectacle enrichi » sur tablette permettrait de diffuser ce travail collaboratif.

⁴⁷ « Résultats de l'appel à projets "Services numériques culturels innovants" 2014 ». Consulté le 9 juin 2015. <http://www.culturecommunication.gouv.fr/Politiques-ministerielles/Recherche-Enseignement-superieur-Technologies/Innovation-numerique/Services-numeriques-culturels-innovants/Resultats-de-l-appel-a-projets-Services-numeriques-culturels-innovants-2014>.

⁴⁸ « Muséobase - Collections muséographiques en Basse-Normandie ». Consulté le 9 juin 2015. <http://museobase.fr/>.

⁴⁹ « Fevis : Fédération des ensembles vocaux et instrumentaux spécialisés ». Consulté le 9 juin 2015. <http://www.fevis.com/>.

⁵⁰ « Le phénix, scène nationale Valenciennes ». Consulté le 9 juin 2015. <http://www.lephenix.fr/>.

Ces projets qui verront le jour en 2015 sont donc la preuve d'une dynamique en faveur du crowdsourcing dans les institutions culturelles en général.

Publics participatifs

Généalogistes

D'après les « chiffres-clefs de la Culture » du Département des études, de la prospective et des statistiques (DEPS) des Archives de France⁵¹, 50 à 60% des personnes qui fréquentent les salles de lecture des Archives Départementales se présentent comme des généalogistes.

Il en va de même pour les internautes, qui sont majoritairement composés de passionnés de généalogie. C'est pourquoi les services d'archives sont attentifs à ces publics et développent à leur destination des outils pour faciliter leurs recherches. Des formations leurs sont également proposées comme des cours de paléographie en ligne. Enfin, les opérations de numérisation concernent en priorité des documents nominatifs.

En retour, les communautés de généalogistes contribuent à améliorer l'accès à ces documents. Ce sens de l'entraide est un héritage ancien, la généalogie étant considérée comme une pratique intellectuelle collective. En effet depuis des siècles, cette pratique repose sur la transmission et le partage. De plus, la recherche généalogique connaît des évolutions et ne se limite plus au seul arbre généalogique ; désormais, de plus en plus de généalogistes élargissent leur champ d'étude au territoire de leurs ancêtres.

Grâce à internet, la diffusion massive de données généalogiques est facilitée. Les modules d'indexation collaborative permettent d'enrichir les bases de données, ce qui améliore la qualité des requêtes. Les généalogistes en contribuant à ces activités servent aussi leurs propres intérêts et par extension ceux leurs pairs. On appelle cela le « cercle vertueux ».

⁵¹ « Archives de France | Chiffres clés de la Culture ». Consulté le 9 juin 2015.
<http://www.archivesdefrance.culture.gouv.fr/archives-publiques/chiffres-clefs-rapports-et-etudes/chiffres-clefs-de-la-culture/>.

Historiens

Si le succès des projets de crowdsourcing est en grande partie assuré par l'investissement des généalogistes, n'oublions pas l'apport des passionnés d'histoire, qu'ils soient érudits ou amateurs. Ce public éclairé appartient à la catégorie des « culture vulture », pour reprendre la terminologie de Chenchen Shen⁵². En effet, de nombreux projets collaboratifs s'appuient sur des connaisseurs, dans des domaines pointus. Dans un certain nombre de projet de crowdsourcing, l'historien est un profil recherché.

Prenons par exemple le cas des Archives départementales des Alpes-Maritimes⁵³ qui ont mis en place un projet pour la transcription de documents bien spécifiques : des visites pastorales datant du XVIIIème siècle. Ces documents, permettent de dater les visites de l'évêque dans un diocèse. Ils peuvent être utiles dans différents cas, par exemple pour aider à la datation de rues portant le nom d'un évêque ou de fresques couvertes d'enduit. Ces documents manuscrits sont particulièrement difficiles à déchiffrer, d'où l'importance d'avoir des notions en paléographie. Cet exercice constitue donc un bon entraînement pour les paléo-historiens, curieux de l'histoire de leur région sous l'Ancien Régime.

Au-delà de la transcription de manuscrits, les chercheurs sont aussi sollicités pour l'édition scientifique et critique de documents, comme c'est le cas avec le projet européen Monasterium⁵⁴, adressé aux médiévistes. Il s'agit du portail numérique du réseau ICARus (International Centre for Archival Research). A ce jour il propose près de 229 000 documents, issus de 100 institutions d'archives à travers l'Europe.

Via ces plateformes, l'historien a l'opportunité de travailler directement à partir de sources historiques inédites, ce qui valorise son travail de recherche.

⁵² *op. cit.*

⁵³ « Wikisource:Partenariats/Archives Départementales des Alpes-Maritimes ». Consulté le 9 juin 2015. http://fr.wikisource.org/wiki/Wikisource:Partenariats/Archives_D%C3%A9partementales_des_Alpes-Maritimes.

⁵⁴ « Monasterium.net ». Consulté le 9 juin 2015. <http://monasterium.net/mom/home>.

Wikimédiens

Enfin, ne négligeons pas la force des communautés pré-existantes comme les contributeurs de *Wikipédia*. En effet, la célèbre encyclopédie en ligne est sans doute l'expression la plus aboutie du web collaboratif, d'où l'intérêt pour les institutions de s'y associer.

De nombreux services d'archives ont d'ores et déjà signé un partenariat avec *Wikimedia France*, comme les Archives départementales des Alpes-Maritimes⁵⁵ qui utilisent cet outil collaboratif depuis janvier 2012.

En juin 2013, un partenariat a été signé avec les Archives de France, afin de les accompagner dans cette démarche de contribution, et de favoriser la mise en ligne de documents d'archives sur *Wikimedia Commons*.

Pour renforcer les liens avec les institutions a été inauguré le concept du « wikipédien en résidence ». Le principe : envoyer un membre de *Wikimedia* au sein de l'institution pour instaurer un véritable dialogue.

Il s'agit d'une idée de Liam Wyatt, vice-président de *Wikimedia Australia*, qui a organisé en 2009 une conférence intitulée « GLAM-WIKI : find the common ground »⁵⁶, proposant de trouver un terrain d'entente entre les institutions culturelles et *Wikipédia*.

Le premier pas a été franchi en 2010 par le British Museum, qui a reçu Liam Wyatt en résidence pour une période de 5 semaines. Les missions énoncées par *Wikimedia* étaient les suivantes⁵⁷ :

- Créer et améliorer les articles concernant des objets ou sujets notables relatifs aux collections et à l'expertise du musée ;
- Soutenir les wikipédiens qui éditent des articles liés au British Museum à la fois au niveau local et international ;

⁵⁵ *op. cit.*

⁵⁶ « GLAM-WIKI - Wikimedia Australia ». Consulté le 9 juin 2015.
<http://wikimedia.org.au/wiki/GLAM-WIKI>.

⁵⁷ « Un wikipédien en résidence au British Museum ». Consulté le 9 juin 2015.
<http://blog.wikimedia.fr/wikipedia-residence-british-museum-1388>.

- Travailler avec les employés du musée pour expliquer le fonctionnement de *Wikipédia* et comment ils peuvent contribuer aux projets directement

Pour la première fois en France, en février 2011, un wikipédien a été accueilli pendant six mois au château de Versailles, afin de faire fructifier les relations entre ces deux acteurs. L'objectif du partenariat étant de permettre une meilleure diffusion des richesses historiques, artistiques et architecturales du domaine de Versailles.

Le bénéfice de cet échange est double : enrichissement en contenu pour *Wikipédia*, et meilleure visibilité pour l'institution culturelle. Quant aux internautes, c'est l'occasion pour eux de se familiariser avec le principe d'édition collaborative.

Forces et faiblesses de l'indexation sociale

Par indexation sociale nous entendons les informations générées par les usagers, désignées en anglais par l'expression *user-generated content*. Ces informations s'inscrivent en complément de l'indexation professionnelle, ce qui pose la question de la fiabilité de ces données. En effet, quels sont les bénéfices de ces métadonnées sociales et les risques qu'elles soulèvent ?

Bénéfices de l'indexation sociale

En premier lieu, l'indexation sociale permet de mieux répondre aux besoins des internautes. En effet, les données produites par les usagers permettent de développer des services différents adaptés à leurs usages. L'objectif est d'améliorer l'accès aux documents, grâce à une description plus fine de ces documents, ce qui permet la recherche avancée. Des outils peuvent alors être mis en place comme des bases de données nominatives, la géolocalisation des documents, ou encore au mode de recherche plein texte. En effet, l'Optical Character Recognition (OCR) est insuffisant, c'est pourquoi on a recourt à l'intelligence humaine. De plus, à ce jour, le système de reconnaissance optique de caractères ne reconnaît ni les manuscrits anciens, ni les livres imprimés qui datent d'avant le XVII^{ème} siècle. Par conséquent l'œil humain reste le seul encore apte à lire et transcrire certains types de documents. L'océrisation des textes

permet non seulement la recherche en texte intégral, mais rend aussi possible la synthèse vocale pour les non-voyants, ou encore la conversion en livre numérique.

Hormis ces nouveaux usages, l'indexation sociale est aussi un apport sur le plan de la connaissance. Dans le cas de documents « muets » dont on ne connaît pas la provenance, l'utilisateur peut être d'un grand secours. Il s'agit bien souvent de photographies, dont on souhaiterait retrouver le lieu, et éventuellement les personnes photographiées. Les internautes se retrouvent dans la position d'un détective, chargé de résoudre une enquête. Enfin, pour les documents qui requièrent une certaine expertise, l'indexation sociale permet de faire appel à des compétences qui échappent aux professionnels de l'information.

Ecueils liés à l'indexation sociale

Si l'indexation sociale présente de nombreux avantages, elle nécessite aussi certaines précautions. Pour commencer, méfions-nous de toute activité collaborative « cosmétique » qui ne donne pas lieu à une amélioration des services. En effet, l'indexation sociale n'a de sens que si elle est ensuite valorisée, et mise au service des internautes. Il faut donc veiller à ce que les contenus générés par les usagers soient réintégrés dans un outil de recherche, comme la base de données.

Un autre écueil à éviter est lié au principe de « folksonomie ». Ce concept a été analysé par Olivier Le Deuff dans un article intitulé « Les usagers indexent le web »⁵⁸. Forgé par Thomas Vander ce terme a été construit à partir du mot « taxinomie » qui désigne les règles de classification et « folk » qui désigne les usagers. La « folksonomie » se définit donc comme des règles établies par les usagers pour indexer le web. Ce mode de classification pose problème dans la mesure où il n'est pas normalisé, contrairement au langage des professionnels. En effet, ce processus de classification collaborative repose sur l'ajout libre de mots-clés. Or, l'usage d'un vocabulaire non contrôlé pour indexer peut causer des problèmes de polysémie (plusieurs mots-clés pour une même idée), d'absence de hiérarchie entre les termes, sans parler des problèmes liés au vocabulaire ou encore à l'orthographe. En somme, ces mots-clés peuvent générer une forme d'« info-pollution », pour reprendre l'expression

⁵⁸ Le Deuff, Olivier. « Folksonomies : les usagers indexent le web ». *Bulletin des bibliothèques de France [en ligne]*, n° 4 (2006).

d'Eric Sutter⁵⁹. En somme, « une folksonomie représente en même temps ce qu'il y a de meilleur et de pire dans l'organisation de l'information », comme conclut très bien Adam Mathes dans son article consacré aux métadonnées sociales publié en 2004.⁶⁰

Etant donné les spécificités des documents d'archives, les métadonnées sociales représentent un nouvel enjeu dans leur traitement, ce qui explique le développement des modules d'indexation collaborative par les services d'archives publics et l'engouement de différents types de public. Malgré la valeur de ces informations générées par les usagers, l'indexation sociale est toutefois à manipuler avec précaution, pour éviter les risques liés aux usages des folksonomies.

⁵⁹ Sutter, Eric. « Pour une écologie de l'information ». *Documentaliste - Sciences de l'information* vol. 35, n° 2 (1998), p 83-86.

⁶⁰ Mathes, Adam. « Folksonomies. Cooperative Classification and Communication Through Shared Metadata », 2004.

TYOLOGIE DES PRATIQUES PARTICIPATIVES

Après avoir inscrit la notion d'archives participatives dans son contexte, établissons à présent une typologie des pratiques. En effet, nous pouvons distinguer plusieurs types de pratiques participatives, impliquant les internautes à échelle variable.

LES ACTIVITES

Pour cet exercice de classification, nous reprendrons la typologie proposée par Oomen et Aroyo dans leur manifeste : *Crowdsourcing in the Cultural Heritage Domain*⁶¹, qui recense les pratiques participatives dans le domaine du patrimoine culturel. Ils établissent les six catégories suivantes : correction/transcription, contextualisation, enrichissement des collections, classification, co-curation et crowdfunding. Pour notre exposé, nous ne retiendrons que les quatre premières catégories, qui sont les plus représentées dans le domaine des archives participatives. Pour chacune des catégories, nous nous appuyerons sur des réalisations concrètes, en alternant études de cas français et internationaux.

Correction & transcription

Correction d'OCR

En raison de la perfectibilité des systèmes de reconnaissance optique que nous avons pu évoquer, la relecture humaine s'avère précieuse. Seule la phase de correction permet d'atteindre un niveau de reconnaissance de 100%. En effet, l'amélioration du mode texte constitue un enjeu pour l'accès au texte intégral.

De nombreux projets de correction ont été mis en place par des institutions à la suite d'une océrisation massive. Début 2015, la BnF (Bibliothèque nationale de France) a inauguré sa plateforme *Correct*⁶², pour la correction et l'enrichissement collaboratifs de textes. Ce projet a été lancé en janvier 2012, dans le cadre de l'appel à projet FUI 12 Ozalid, en collaboration avec 8 autres partenaires : Orange, Jamespot, Urbilog, I2S, ISEP, INSA, l'université Lyon 1 et l'université Paris 8. Depuis décembre 2014, la

⁶¹ Oomen, Johan, et Lora Aroyo. *Crowdsourcing in the Cultural Heritage Domain : Opportunities and Challenges*. 5th International Conference on Communities and Technologies. New York, 2011.

⁶² « Réseau Correct ». Consulté le 10 juin 2015. <http://www.reseau-correct.fr/>.

plateforme existe en version bêta, pour une phase de test qui prendra fin en juin 2015. Comme son nom l'indique, cette plateforme est destinée à la correction collaborative de documents, afin de permettre la réédition, notamment pour les personnes malvoyantes. Ce projet concerne les monographies issues de Gallica, la bibliothèque numérique de la BnF. A terme, les objectifs sont les suivants :

- Améliorer le mode texte et l'indexation des documents
- Diffuser les documents au format ePub
- Produire une version accessible aux synthétiseurs vocaux
- Concevoir une édition critique d'un ouvrage numérique

Cette plateforme a été conçue sur le modèle d'un réseau social, afin de générer un lien entre les contributeurs. Le lancement du projet est encourageant, selon la responsable du projet Isabelle Josse⁶³ : en à peine 4 mois d'existence, la plateforme compte déjà 531 inscrits, et 5560 pages corrigées⁶⁴.

Sur le plan international, nous pouvons citer le projet *Trove*⁶⁵ de la Bibliothèque nationale d'Australie ou encore le projet de la California Digital Newspaper Collection aux Etats-Unis.

Transcription de manuscrits

A un degré de participation supérieur, l'activité de transcription consiste en la saisie d'un texte *ex nihilo*, directement à partir du manuscrit.

Un projet emblématique de transcription collaborative est le projet *Transcribe Bentham*⁶⁶ lancé par l'University College of London (UCL) en avril 2010. Il est consacré au philosophe britannique Jeremy Bentham, dont on a retrouvé 60 000 manuscrits inédits datant de la fin du XVIIIème siècle. En bon visionnaire, il écrit en 1793 « Many hands make light work. Many hands together make merry work »,

⁶³ Réponse à notre questionnaire en date du 09/03/15

⁶⁴ Chiffres collectés le 9 mars 2015 sur la plateforme Correct

⁶⁵ « National Library of Australia - Trove ». Consulté le 11 juin 2015. <http://trove.nla.gov.au/>.

⁶⁶ « Transcribe Bentham ». Consulté le 10 juin 2015. http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham.

sentence que nous pouvons traduire par le proverbe « A plusieurs ouvriers, la tâche devient aisée ».

Cinq ans après son lancement, ce projet connaît une participation constante. Régulièrement, des statistiques sont communiquées sur le blog rattaché au projet⁶⁷. En la date du 27 mars 2015, on compte 27,222 inscrits et 12,798 manuscrits transcrits, dont 91% ont été vérifiés par les modérateurs.

En effet, les transcriptions passent par une étape de vérification par l'équipe de l'UCL. Avant d'être éditée, une transcription doit être soumise à plusieurs utilisateurs, ce qui permet un croisement des sources. C'est seulement après cette comparaison que la version finale peut être verrouillée dans la base de données. Les participants doivent être inscrits pour accéder au tableau de bord du projet qui permet l'accès aux manuscrits. L'interface permet de sélectionner un manuscrit selon plusieurs critères – thématiques ou chronologiques, niveau de difficulté- ou encore sur un mode aléatoire. La plateforme accompagne l'utilisateur dans la transcription, en lui proposant les balises XML correspondantes dans la barre d'outils. Pour les débutants, un guide d'utilisation est mis à disposition sous forme de tutoriel vidéo.

En mai 2011, ce projet a reçu le prix *Ars Electronica* dans la catégorie des Digital Communities, tout comme *Wikipédia* en 2004. Son succès repose sur l'originalité de cette entreprise, qui permet aux internautes de découvrir les textes inédits d'un grand philosophe.

Transcription de fragments

Si l'on monte encore d'un cran dans la difficulté, l'exercice de la transcription collaborative peut s'appliquer à des fragments de manuscrits. C'est le cas du projet *Ancient Lives*⁶⁸ lancé en 2001 et coordonné par l'Université d'Oxford, en collaboration avec des chercheurs spécialisés en papyrologie. Il s'inscrit dans le programme *Zooniverse* qui repose sur le principe de « sciences citoyennes ». Si la plupart du temps

⁶⁷ « UCL Transcribe Bentham ». Consulté le 10 juin 2015. <http://blogs.ucl.ac.uk/transcribe-bentham/>.

⁶⁸ « Ancient Lives | Transcribe ». Consulté le 10 juin 2015. <http://www.ancientlives.org/tutorial/transcribe>.

les projets du programme concernent les sciences dures, comme *Galaxy Zoo*⁶⁹, le projet *Ancient Lives* concerne lui des collections patrimoniales. Ces documents sont d'une grande valeur historique puisqu'il s'agit d'une centaine de milliers de fragments de papyrus de l'Égypte gréco-romaine. Ils proviennent de la ville antique de l'Oxyrhynque, la cité du « poisson au nez pointu ».

Ces manuscrits sont inscrits en grec ancien, ce qui nécessite une connaissance minimale de l'alphabet. Toutefois, on constate que de nombreux contributeurs sont des hellénistes puisqu'ils arrivent à deviner des mots difficilement déchiffrables. En effet la difficulté de l'exercice réside dans le morcellement de ces documents, dont on n'a que des bribes et pas d'éléments contextuels à part l'époque. Toutefois il est possible de lancer une recherche de correspondance entre documents, afin d'identifier des corpus analogues qui présentent les mêmes groupes de lettres. Outre le module de transcription, le site propose un module consistant à mesurer les marges et l'espacement des colonnes. Ces mesures aident à l'identification du contenu des documents grâce aux règles de structuration du papyrus.

En décembre 2012, le blog⁷⁰ fait état de 1,5 million de transcriptions réalisées, qui ont permis l'identification d'une centaine de textes, dont des écrits de Plutarque et de Simonide. A l'avenir, les manuscrits devraient être élargis à d'autres alphabets comme l'égyptien ancien, l'araméen, l'arabe ou encore le copte.

Contextualisation

Au-delà du contenu, l'activité collaborative peut concerner le contexte d'élaboration des archives. L'internaute peut fournir des éléments de contexte sur le fonds d'archives, afin de permettre une meilleure compréhension de ces documents. Cette contextualisation peut être objective ou subjective et prendre des formes variées.

⁶⁹ *op. cit.*

⁷⁰ « Ancient Lives | A Zooniverse Project Blog ». Consulté le 10 juin 2015. <http://blog.ancientlives.org/>.

Encyclopédie collaborative

Certains services d'archives peuvent encadrer des travaux de recherche autour de leurs collections. Ces travaux peuvent amener à l'élaboration d'encyclopédies collaboratives par exemple. La Bibliothèque numérique de Toulouse a lancé sa propre encyclopédie : la *Rosalipédie*⁷¹. Cette œuvre collective contribue à la valorisation de ses collections. Sur l'espace collaboratif, chaque internaute peut publier un article à propos d'un document, à condition qu'il soit libre de droit. Cet article, une fois validé, vient alimenter l'encyclopédie. L'accès aux articles se fait par sujet, par auteur ou bien selon une liste alphabétique des documents. Le site compte à ce jour une cinquantaine de contributeurs, dont chacun a publié entre 1 et 100 articles. Outre des articles, on peut également trouver dans la Rosalipédie des conférences audio, expositions virtuelles et dossiers pédagogiques, tous élaborés à partir des collections de la bibliothèque.

Dictionnaire collaboratif

Outre le format encyclopédique, les internautes peuvent être à l'origine d'autres outils collaboratifs comme des dictionnaires. C'est le projet des Archives départementales de Vendée qui ont lancé en 2011 un module intitulé *Le L@boratoire des internautes*⁷². Sur cette plateforme, les internautes ont la possibilité d'alimenter trois dictionnaires collaboratifs : le dictionnaire des Vendéens, le dictionnaire historique des communes et le dictionnaire des toponymes. Ces dictionnaires ont été élaborés grâce à la contribution massive de communautés de chercheurs, et continuent d'évoluer au fil de la numérisation de documents d'archives. Ils comportent des notices détaillées sur l'histoire de la région, avec la possibilité d'effectuer des recherches cartographiques parmi ces notices.

⁷¹ « Rosalis - Rosalipédie ». Consulté le 10 juin 2015.

<http://rosalis.bibliotheque.toulouse.fr/index.php?pages/rosalipedie#.VXhaGkY2V-o>.

⁷² « Laboratoire des internautes ». Consulté le 10 juin 2015. <http://laboratoire-archives.vendee.fr/>.

Plateforme d'écriture collaborative

Enfin, les activités de contextualisation peuvent aboutir à de véritables plateformes d'écriture collaborative, sur le modèle des wiki. L'un des wiki les plus emblématiques dans ce domaine a fermé en 2012 et s'intitulait *Your Archives*⁷³. Ce projet avait été initié en 2007 par les Archives nationales du Royaume-Uni autour de documents portant sur l'histoire et le patrimoine britannique et conservés dans différents centres d'archives du pays. En 2012, le wiki comptait 31 000 contributeurs et 21 000 articles publiés. Après la fermeture de la plateforme dédiée, les contenus ont été directement réintégrés au sein du nouveau catalogue de l'institution, et ce afin de permettre une valorisation de cette production collaborative. Les modules collaboratifs sont désormais rattachés au site institutionnel. Les technologies évoluent donc, mais les usages du crowdsourcing, eux, perdurent.

Les activités de contextualisation génèrent donc des outils de recherches collaboratifs innovants.

Enrichissement

Les institutions culturelles peuvent également tirer parti du web social pour enrichir leurs collections d'archives privées via des opérations de collecte.

Collecte autour d'un événement

Une collecte peut s'organiser dans le cadre d'un projet en particulier, comme la commémoration d'un événement historique. C'est le cas du projet *Europeana 1914-1918*⁷⁴, lancé à l'occasion du centenaire de la première guerre mondiale. Le site regroupe des documents officiels provenant de bibliothèques et services d'archives du monde entier ainsi que des histoires inédites qui constituent des souvenirs de famille.

⁷³ « Your Archives ». Consulté le 10 juin 2015. http://webarchive.nationalarchives.gov.uk/+/http://yourarchives.nationalarchives.gov.uk/index.php?title=home_page.

⁷⁴ « Europeana 1914-1918 - histoires inédites et histoires officielles de la Première Guerre mondiale ». Consulté le 10 juin 2015. <http://www.europeana1914-1918.eu/fr>.

L'internaute est invité à partager son histoire familiale en lien avec la Grande Guerre, à travers des photographies, des cartes postales ou bien des échanges épistolaires. Cette opération qui a pris fin en juin 2013 a permis de numériser plus de 60 000 pièces, provenant d'une douzaine de pays européens.

En partenariat, de nombreux services d'archives ont participé à la « Mission du Centenaire de la Première Guerre Mondiale », mise en place par le gouvernement, qui a rejoint cette collecte de grande envergure. Cette Grande Collecte a été organisée du 9 au 16 novembre 2013 par les Archives de France et la Bibliothèque Nationale de France, en collaboration avec *Europeana*. Cette collecte a été suivie d'événements commémoratifs dans la France entière.

Cette collecte auprès du grand public a aussi permis de fédérer les pays européens autour de cet épisode historique marquant du XXème siècle.

Appel aux collectionneurs

Les opérations de collecte peuvent également faire appel aux collectionneurs de documents d'archives, intéressés par l'histoire de leur région. Un exemple emblématique est celui du projet des archives départementales de Lot-et-Garonne qui a constitué une cartotheque⁷⁵ grâce aux dons de particuliers. Ce fonds iconographique est aujourd'hui composé de 3 000 cartes postales. Pour déposer ses cartes postales, trois critères sont imposés : la carte doit être suffisamment ancienne pour être libre de droits, inédite sur la base de données et d'une qualité minimale de 300 dpi. En même temps que de déposer ses archives, le contributeur a la possibilité de les indexer, avec l'obligation de renseigner au moins la commune. Selon la responsable du fonds iconographique, Véronique Pouyadou⁷⁶, la plupart des contributeurs sont originaires de la région. Au lancement du projet en 2010, 300 dépôts ont été effectués durant la première année, sachant qu'un même internaute a pu déposer entre 1 et 40 cartes. L'inscription est obligatoire pour pouvoir déposer ses cartes postales. La dimension collaborative est toutefois limitée, puisqu'il n'est pas possible d'indexer les fonds déposés par d'autres internautes.

⁷⁵ « Le Lot-et-Garonne en cartes postales ». Consulté le 10 juin 2015. <http://www.cartotheque47.fr/>.

⁷⁶ Entretien téléphonique du 06/03/15

Classification

Pratiquée en masse, l'activité de classification vise à récolter des métadonnées descriptives liées aux documents d'une collection.

Tagging social

L'activité de *tagging social* rejoint le principe de « folksonomie » défini plus haut, qui désigne un processus d'indexation libre. L'internaute est invité à indexer des documents à l'aide de mots-clés, qu'on appelle des « tags ». Ces mots-clés sont exprimés en langage naturel et choisis librement par l'utilisateur. Ce procédé est très répandu pour l'indexation de photographies et peut s'effectuer via les médias sociaux de partage de contenus comme *Flickr*, qui propose cette fonctionnalité. A noter qu'en février 2015, seuls cinq services d'archives français possèdent un compte officiel sur *Flickr* : les Archives départementales des Alpes-Maritimes, les Archives Municipales de Beaune, les Archives Municipales d'Evreux, les Archives Municipales de Lyon et les Archives Municipales de Toulouse. Or l'identification des documents iconographiques rend possible l'indexation par un moteur de recherche, ce qui facilite l'accès à ces photographies.

Toutefois, les documents iconographiques ne sont pas les seuls soumis au *tagging*. Aux Pays-Bas, l'Institut néerlandais pour le Son et l'Image a proposé l'indexation de documents audiovisuels dans un projet intitulé *Waisda* ⁷⁷, traduit en anglais par *What's that ?* Lancé en mars 2009, il se présente comme un jeu vidéo (*video labelling game*). La consigne est simple : « Saisissez les mots que vous voyez et entendez ». Le joueur marque des points lorsqu'il saisit le même mot-clé que son adversaire dans un laps de temps de 10 secondes. Il s'agit d'un moyen de contrôler le vocabulaire employé par les joueurs. Les vidéos sont tirées des archives de la télévision. Cette activité ludique a remporté un succès étonnant de la part du public : lors des six premiers mois d'essai (de mars à novembre 2009), 2 000 joueurs ont contribué à l'ajout de 34 000 tags sur 600 images d'archives. Sur ces 34 000 tags, 40,3% étaient des mots-clés « gagnants », car employés par plusieurs joueurs.

⁷⁷ « Waisda | GitHub ». Consulté le 10 juin 2015. <https://github.com/beeldengeluid/waisda>.

Le projet *Waisda ?* a réussi trois défis : utiliser le jeu comme une méthode d'indexation du patrimoine télévisuel, collaborer avec des communautés liées à ce type de contenu, et intégrer le vocabulaire des utilisateurs en complément de l'indexation professionnelle.

Indexation collaborative

Contrairement au *tagging* social, l'activité d'indexation collaborative fait l'objet d'un encadrement plus strict puisqu'elle exige l'utilisation d'un vocabulaire contrôlé. En effet, les informations saisies par les internautes doivent pouvoir servir à la constitution de bases de données structurées. Pour cela, le contributeur doit renseigner des champs avec des informations précises : indexation du nom, prénom, date, lieu... Ces champs sont signalés par des « signets », qui désignent où se trouve l'information sur le document. Les documents sont souvent attribués aux internautes par « lot », afin de permettre une cohérence des documents entre eux.

Un grand nombre de services d'archives disposent déjà d'un module d'indexation collaborative, afin de permettre un accès aux documents numérisés. Cette activité collaborative est largement suivie par les publics, à en juger par les résultats de ces dernières années. Selon les chiffres communiqués par Bouyé⁷⁸, au 31 janvier 2012, 2.487.000 signets avaient été posés au total, dont 1.804.000 dans les dix-neuf services d'archives départementales. Parmi ces institutions, les archives départementales d'Aube et du Cantal se détachaient nettement du lot, dépassant à elles deux les 400.000 signets. Pour donner une indication, aux Archives départementales du Cantal, 1 000 micro-tâches sont réalisées chaque jour. Malgré tout, le volume des documents nominatifs à indexer reste important : à elles seules, les Archives Départementales de l'Ain proposent 2.200.000 noms sur leur base. En deux ans, 500.000 pages ont déjà été indexées. Enfin, deux services d'Archives Communales ont réussi à atteindre un taux de participation très important, avec un total de 405.000 signets posés pour Nantes et 278.000 pour Rennes. Cet engouement dans cette zone de l'est de la France peut s'expliquer par le mouvement d'émigration des bretons dès la fin du XIX^{ème} siècle, d'où leur intérêt pour la généalogie.

⁷⁸ Bouyé, Edouard. *op. cit.*, p. 5.

De ce fait, l'indexation collaborative est l'activité de *crowdsourcing* la plus représentée dans les services d'archives publiques.

Catalogage collaboratif

Au-delà de l'activité d'indexation, le catalogage permet de fournir des informations plus détaillées sur un document d'archives. En effet, il s'agit de décrire le contenu du document à l'aide de métadonnées descriptives.

Le catalogage peut s'appliquer à des documents textuels, mais pas exclusivement. Par exemple, la Bodleian Library, une des plus prestigieuses universités d'Oxford, propose le catalogage de partitions musicales. Le projet est intitulé *What's the score at the Bodleian ?*⁷⁹ A travers ce projet, la bibliothèque de Bodleian souhaite améliorer l'accès à sa collection musicale, qui comporte 4 000 partitions numérisées. Il s'agit principalement de partitions pour piano datant de la période victorienne, dont certaines comportent des couvertures illustrées. Les partitions sont présentées dans un mode aléatoire, afin de donner une importance égale à tous les documents. Pour chaque partition, le contributeur doit remplir des métadonnées structurées qui concernent le morceau de musique (titre, compositeur, arrangeur...), la publication (éditeur, lieu de publication...) ou les illustrations (type de couverture, illustrateur...). Une connaissance musicale minimale est requise pour pouvoir contribuer, car il faut également préciser la notation musicale (clé musicale, tempo...) En avril 2015, 40.017 pages étaient déjà cataloguées, soit presque la moitié de la collection en ligne (46%).

De même que l'activité de *tagging*, le catalogage se révèle aussi très utile pour l'identification de documents iconographiques.

LES PLATEFORMES

Afin de développer des fonctionnalités collaboratives, les services d'archives doivent faire un choix, entre soit développer leur site institutionnel à l'aide d'un prestataire de service, soit disséminer leurs contenus sur des sites de partage.

⁷⁹ « Whats the score at the Bodleian ». Consulté le 10 juin 2015. <http://www.whats-the-score.org/>.

Les solutions logicielles

Opérateurs du marché

Pour effectuer leur choix parmi les différents services opérateurs, les services d'archives sont appelées à réaliser une étude de marché. En effet, le web collaboratif est soumis à la loi du marché imposant la concurrence entre différents acteurs économiques. Ils font alors appel à des sociétés privées pour comparer leurs offres. Dans son étude sur l'indexation collaborative par les services d'archives, Edouard Bouyé⁸⁰ dresse une analyse de ce marché, en janvier 2012.

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

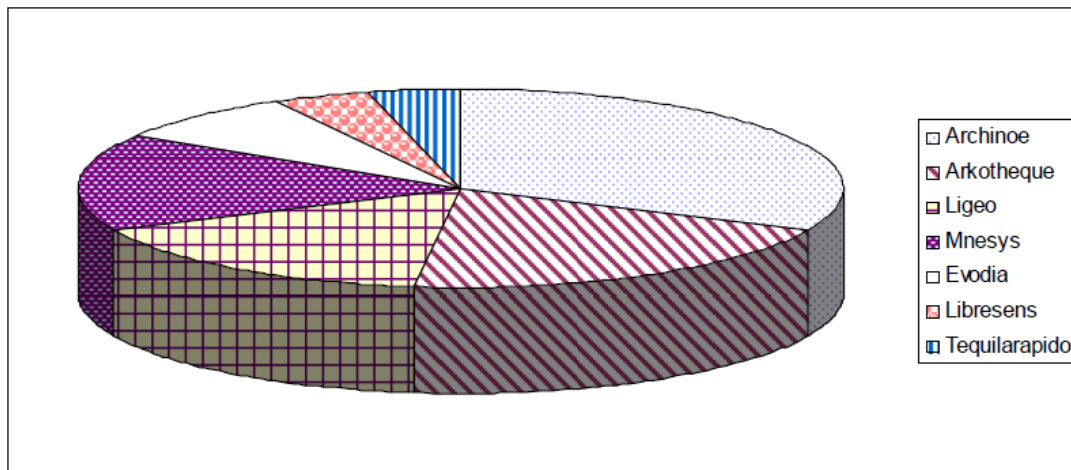


Figure 7 : Systèmes d'indexation collaborative utilisés par les services d'archives en janvier 2012 (diagramme réalisé par Edouard Bouyé)

De cette étude, on retient que quatre principaux opérateurs dominent le marché :

- 1) *Archinoë*, de la société *Archimaine*, sollicité par six services d'archives, soit un tiers du marché.
- 2) *Arkotheque*, de la société *1 égale 2*, sollicité par cinq services d'archives
- 3) *Ligeo*, de la société *V-Technologies*, sollicité par quatre services d'archives *ex aequo* avec
- 4) *Mnesys*, de la société *Naoned Systèmes*, sollicité par quatre services d'archives également

⁸⁰ Bouyé, Edouard. *op. cit.*, p. 5.

A noter que quatre services d'archives ont souhaité coupler deux opérateurs pour leur projet.

Les trois opérateurs suivants correspondent à des cas particuliers :

- *Evodia*, société créée en 2005 à Rennes qui est en cours de liquidation
- *Libresens*, agence localisée à Agens qui a développé la cartotheque des Archives Départementales du Lot-et-Garonne
- *Tequilarapido*, société niçoise et parisienne qui a été sollicité par les Archives Départementales des Alpes-Maritimes pour son espace collaboratif *Le L@boratoire des internautes*.

Analysons à présent l'offre des deux principaux prestataires en matière d'indexation collaborative : *Archinoë* et *Ligeo*.

Archinoë

Ce leader de la numérisation a numérisé les registres paroissiaux et les registres d'état civil de plus de trente départements et de dizaine de mairies. Sur son site web, il propose un portail d'indexation accessible en version beta⁸¹. Sur cette plateforme collaborative sont accessibles 44 fonds numérisés, issus de 13 services d'archives. Les services participant au dépôt sur *Archinoë* sont les suivants : 9 services d'Archives départementales (AD de Loire Atlantique, AD du Lot-et Garonne, AD de la Mayenne, AD de Meurthe-et-Moselle, AD de l'Oise, AD de la Sarthe, AD des Deux-Sèvres, AD de Vendée, AD de la Vienne), et 3 services d'Archives Municipales (AM de Romans-sur-Isère, AM de Nantes, AM de Lyon).

L'utilisateur peut participer librement à des campagnes d'indexation, à condition de s'être inscrit au préalable sur le site. L'internaute peut malgré tout visualiser les documents numérisés sans ouvrir de compte. Le module de visualisation est ergonomique et permet une lecture confortable, à l'aide d'outils tels que le zoom, la sélection de zone, la rotation, le réglage de l'image –luminosité, contraste-, et même un convertisseur de dates, qui permet d'établir une correspondance entre les calendriers grégorien et républicain. Le module collaboratif quant à lui permet une indexation facilitée : il suffit de remplir les champs indiqués (Nom, Prénom, Sexe, Date, Type

⁸¹ « Archinoë - Portail d'indexation collaborative ». Consulté le 10 juin 2015.
<http://www.archinoe.fr/portail/>.

d'acte, Numéro d'acte, Commune, Subdivision). Enfin, une fenêtre permet de signaler des erreurs éventuelles.

Ligeo Archives

La solution proposée par la société *Ligeo*⁸² est conçue par des archivistes pour des archivistes. Elle prend en compte différents aspects du métier, dont des modules collaboratifs, en vue de la valorisation des fonds.

A un premier niveau de participation, *Ligeo* propose un module de « commentaires sur les notices », qui permet d'associer les internautes à la description des fonds. Par exemple, les archives municipales de Saint-Denis ont utilisé cette fonctionnalité sur des notices de recensement militaire, ce qui a permis aux contributeurs de témoigner en apportant des précisions sur la vie de soldat.⁸³

A un deuxième niveau, l'indexation collaborative s'effectue via un visualiseur, qui sert d'interface, de la même manière que pour *Archinoë*. Toutefois, une restriction de plus est imposée au contributeur : l'attribution d'un niveau, qui lui permet d'accéder à un lot d'images numérisées selon le nombre de contributions à son actif. Les formulaires d'indexation sont configurables selon les cas.

Outre l'indexation collaborative, *Ligeo* propose ce qu'il appelle les « enquêtes participatives », qui permettent l'identification d'images. Il s'agit de photos mystères, pour lesquelles on manque d'informations : on parle alors de photos « muettes ». Le module se divise entre « enquêtes en cours » et « enquêtes résolues », l'objectif étant de résoudre un maximum d'énigmes. Ce module « SOS Détective » a été mis en place dans quatre services d'archives : les archives départementales de l'Ain, avec « SOS Archives »⁸⁴, les archives départementales de Lozère, avec « Photos mystères »⁸⁵ les

⁸² « Ligeo archives - Progiciel de gestion ». Consulté le 10 juin 2015. <https://www.ligeo-archives.com/>.

⁸³ « Engagements volontaires et recensements militaires- Archives municipales de Saint-Denis ». Consulté le 10 juin 2015. <http://archives.ville-saint-denis.fr/archive/recherche/recensementmilitaire/n:18>.

⁸⁴ « SOS Détective - Archives départementales de l'Ain ». Consulté le 10 juin 2015. http://www.archives-numerisees.ain.fr/archives/enquete/enquetes_en_cours/.

⁸⁵ « Photos mystères - Archives départementales de la Lozère ». Consulté le 10 juin 2015. <http://archives.lozere.fr/enquetes/sos-detectives/enquetes-en-cours/27>.

archives municipales de Bordeaux, avec « Aidez-nous à identifier »⁸⁶ et les archives municipales de Saint-Denis avec « Appel à vous »⁸⁷. Depuis 2012, plus de 200 enquêtes ont déjà été publiées sur le site de Saint-Denis, dont la moitié a été résolue.

Les sites de partage de contenu

Les services d'archives peuvent également faire le choix de disséminer leurs contenus sur des plateformes consacrées, comme les médias sociaux. En élargissant leur champ, ils offrent une meilleure visibilité à leurs collections.

Flickr the Commons

Si les amateurs de photographie numérique connaissent la plateforme *Flickr*, il existe depuis 2008 un volet spécialement réservé aux organismes publics. Lancé par la Bibliothèque du Congrès, le projet *Flickr the Commons*⁸⁸ a été mis en place pour permettre aux acteurs institutionnels de diffuser leurs photographies d'archives. Depuis son lancement le 16 janvier 2008, 92 institutions participent à cette vaste opération, d'ampleur mondiale⁸⁹. Cette entreprise permet une « dissémination » des archives, pour reprendre le concept anglo-saxon, employé en sciences de l'information.

Rappelons qu'initialement, le réseau *Flickr* avait été conçu en 2002 par la société canadienne *Ludicorp*, comme un jeu vidéo en ligne, où les photographies étaient utilisées comme supports. Puis, le téléchargement des photographies prenant de l'importance, la plateforme s'est développée dans ce sens. Elle permet désormais la recherche, le stockage et le partage de photos sur sa page personnelle. Notons que le terme *flickr* renvoie au verbe anglais *to flick through* qui signifie feuilleter.

La plateforme de *Flickr the Commons* fonctionne sur le modèle du réseau social en proposant trois types d'interactions. Tout d'abord les « tags », choisis librement par les usagers, qui permettent un enrichissement continu du thésaurus. Chaque image comporte au minimum 3 tags (nom de la bibliothèque et 2 numéros machine

⁸⁶ « De quel événement s'agit-il? - Archives municipales de Bordeaux ». Consulté le 10 juin 2015. <http://archives.bordeaux.fr/enquetes/sos-detectives/enquetes-en-cours/11>

⁸⁷ « Appel à vous - Archives municipales de Saint-Denis ». Consulté le 10 juin 2015. <http://archives.ville-saint-denis.fr/enquetes/sos-detectives/enquetes-en-cours/290>.

⁸⁸ « Flickr: The Commons ». Consulté le 10 juin 2015. <https://www.flickr.com/commons>.

⁸⁹ « Flickrland 3 | Les carnets de la phonothèque ». Consulté le 10 juin 2015. <http://phonothèque.hypotheses.org/14364>.

d'identification) et maximum 75, sachant que la moyenne tourne autour de 23 tags. Puis le module de « commentaire », qui permet aux internautes d'interagir. Enfin, les « notes », qui sont des annotations intervenant directement sur l'image, mais qui ont disparu pour des raisons de surcharge de l'image.

Le réseau de *Flickr the Commons* est donc en perpétuelle évolution et accueille un nombre de plus en plus importants d'institutions, en quête de visibilité.

History Pin

Tout comme *Flickr*, *History Pin*⁹⁰ est un média social qui repose sur les archives iconographiques, avec comme caractéristique la géolocalisation des images. La navigation se fait via *Google Map*, où sont épinglées les photographies. Comme pour *Flickr*, ces photographies peuvent être ajoutées par des utilisateurs ou bien des institutions. Le 9 avril 2015, on compte 409.619 contenus, dont 2.308 ont été ajoutés par des institutions. Parmi les institutions internationales présentes sur *History Pin*, on note : les Archives nationales des Etats-Unis, les Archives nationales de Grande-Bretagne et les Archives de Vancouver au Canada.

Une fois son profil créé, l'utilisateur peut créer un nombre illimité de collections photographiques autour de sujets spécifiques. Chaque photo épinglée doit comporter des métadonnées, sans lesquelles elle ne pourra pas être visible. Les renseignements de date et de lieu sont obligatoires. Ces informations peuvent être plus ou moins précises : date précise ou approximative, adresse ou nom de la ville. Ces métadonnées permettent de contextualiser les images.

Enfin, l'originalité du site est de proposer une fonction de *geomapping* des photos, qui permet de superposer des vieilles photos sur le paysage actuel, grâce à l'outil de *Streetview* de *GoogleMap*⁹¹. Ce module permet de confronter plusieurs époques et de constater les transformations opérées par le temps.

⁹⁰ « Historypin | Home ». Consulté le 10 juin 2015. <http://www.historypin.org/>.

⁹¹ « Street View - Google Maps ». Consulté le 10 juin 2015. <https://www.google.com/maps/views/streetview?gl=fr&hl=fr>.

Wikisource

En ce qui concerne l'édition collaborative, elle peut se faire directement grâce à l'outil de *Wikisource*⁹². Rappelons qu'au départ *Wikisource* est un projet de publication pour les textes libres de droit. Un atout non négligeable est que cet outil est déjà pensé pour la correction d'OCR et comporte des fonctionnalités intéressantes comme l'affichage côte à côte de l'image et du texte, des indications pour signaler si la page a été transcrite, corrigée ou validée, un historique des modifications et la possibilité de commenter. Enfin, *Wikisource* dispose d'un bon *PageRank* sur Google, ce qui assure une visibilité des collections mises en lignes.

En raison de ses avantages, les archives départementales des Alpes-Maritimes ont signé en 2012 un partenariat avec *Wikisource* pour proposer une transcription collaborative de ses visites paroissiales⁹³. Ce projet est coordonné par la directrice adjointe des Archives, Mme Cavalié qui nous a apporté des précisions sur les publics visés.⁹⁴ Le public prioritairement visé est celui des étudiants en paléographie de l'université de Nice, qui dans le cadre de leur licence professionnelle souhaitent s'entraîner. Il peut s'agir aussi d'anciens élèves de ce cursus, qui souhaitent continuer à pratiquer. Enfin, certains érudits sollicitent aussi ce corpus, en transcrivant des passages qui concernent leur municipalité. On retrouve aussi professeurs de paléographie, à la recherche d'exercices pratiques, mais aussi des thésards spécialisés dans cette période, ou même des évêques. Bien que modeste, ce projet a permis d'obtenir en 2012, un total de 136 pages transcrites et corrigées, grâce à une augmentation croissante : 95 pages de transcrites et corrigées en 2013, soit un total de 231 pages et 152 pages transcrites et corrigées en 2014, soit un total de 383.

Sur le même modèle, les archives départementales des Alpes-Maritimes se sont associées en 2014 à d'autres départements pour mettre en ligne et éditer les Carnets de guerre de poilus⁹⁵, à l'occasion de la « Grande collecte 1914-1918 ». Il s'agit de carnets de guerre inédits, conservés par les familles et redécouverts 100 ans plus tard. La page *Wikisource* consacrée au projet cite les institutions partenaires : Archives départementales d'Indre-et-Loire, Archives départementales du Pas-de-Calais et

⁹² « Wikisource ». Consulté le 10 juin 2015. <https://fr.wikisource.org/wiki/Wikisource:Accueil>.

⁹³ *op. cit.*

⁹⁴ Réponse à notre questionnaire du 09/03/15

⁹⁵ « Carnets de guerre - Wikisource ». Consulté le 10 juin 2015. https://fr.wikisource.org/wiki/Wikisource:Carnets_de_guerre.

Archives départementales de Saône-et-Loire, et donne la liste actualisée des carnets mis en ligne (qui étaient au nombre de six en mars 2015). Un travail de valorisation de ces carnets de guerre a été réalisé à travers une sélection d'extraits, respectant l'orthographe des manuscrits dans un souci d'authenticité.

Les initiatives en marge des GLAM

En marge des institutions de type GLAM, d'autres projets collaboratives émergent à travers des initiatives associatives, ministérielles ou bien individuelles

Geneanet

L'association *Geneanet*⁹⁶ rassemble près de deux millions de membres passionnés de généalogie, depuis sa création en 1996. Il s'agit du 6^{ème} site mondial de généalogie. Cet outil collaboratif est gratuit, mais comporte des fonctionnalités réservées aux abonnés *Premium*. Il offre la possibilité de reconstituer l'arbre généalogique de sa famille, en interrogeant une base de données puissante -par nom, lieu, période-, avec un module comparatif. Des « projets contributifs » sont également proposés aux internautes. Pour l'heure, quatre projets sont en cours, classés par difficulté, par période et par localisation. Ils sont détaillés sur *GeneaWiki*⁹⁷, encyclopédie de la généalogie gratuite, éditée de manière coopérative.

Parmi les projets en cours, l'opération « Sauvons nos tombes »⁹⁸ consiste à photographier des pierres tombales, pour ensuite en relever les noms. Le contributeur peut soit réaliser lui-même ses prises de vue dans un cimetière de son choix, soit participer directement au relevé des noms de chez lui. Ce projet, lancé en septembre 2014 a déjà permis de relever les noms de 265 000 tombes.

L'opération « matricules napoléoniens 1802-1815 »⁹⁹, elle a pour objectif d'indexer les registres matricules de la garde impériale et de l'infanterie de ligne datant

⁹⁶ « Geneanet ». Consulté le 10 juin 2015. <http://www.geneanet.org/>.

⁹⁷ « GeneaWiki ». Consulté le 10 juin 2015. <http://fr.geneawiki.com/index.php/Accueil>.

⁹⁸ « Sauvons nos tombes - GeneaWiki ». Consulté le 10 juin 2015.

http://fr.geneawiki.com/index.php/Sauvons_nos_tombes.

⁹⁹ « Matricules Napoléoniens 1802-1815 - GeneaWiki ». Consulté le 10 juin 2015.

http://fr.geneawiki.com/index.php/Matricules_Napol%C3%A9oniens_1802-1815.

de l'époque napoléonienne. Jusqu'à présent, 55 000 soldats ont déjà été indexés, ce qui peut permettre aux généalogistes de retrouver leurs ancêtres.

Dans le projet « Livres d'or des Morts pour la France »¹⁰⁰, l'objectif est de recenser les soldats ayant bénéficié du titre « Mort pour la France ». Ces livres d'or ont été réalisés en 1929 dans chaque commune de France. Ils permettent de compléter les sources que constituent les inscriptions sur les monuments aux morts, et les fiches des Morts pour la France » réalisées par l'armée. Pour l'heure, 33 communes ont déjà été indexées.

Enfin, un projet porte sur les « Archives notariales »¹⁰¹, qui constituent le Livre des vies de nos ancêtres. Ils permettent de compléter les sources classiques que sont les registres paroissiaux ou encore l'état civil. Ces documents constituent une mine de renseignements en ce qui concerne les précisions historiques, les détails de la vie quotidienne, les anecdotes sur le patrimoine. Malgré leur importance, trop peu d'archives de ce type ont été numérisées par les archives départementales. Un mode de participation consiste donc en la numérisation de ces archives, ou bien en l'indexation des documents déjà en ligne. A ce jour, 10 000 liasses d'archives notariales ont été numérisées dans le cadre de ce projet. Depuis 2005, les généalogistes peuvent aussi renseigner la fiche de leur commune sur *Geneawiki*, afin de partager leurs connaissances du lieu, d'un point de vue historique ou plus anecdotique et éventuellement des documents liés à ce lieu. En 10 ans, 39 705 contributeurs ont participé à l'enrichissement de ce projet.

Mémoire des Hommes

Le projet *Mémoire des Hommes*¹⁰² a été lancé en novembre 2013 par le Ministère de la Défense, et plus particulièrement par la Direction de la mémoire, du patrimoine et des archives. Il propose un programme d'indexation collaborative de la base des Morts pour la France de la Première guerre mondiale. Le module d'indexation est accessible

¹⁰⁰ « Livres d'or des Morts pour la France - GeneaWiki ». Consulté le 10 juin 2015.
http://fr.geneawiki.com/index.php/Livres_d%27or_des_Morts_pour_la_France.

¹⁰¹ « Projet Archives notariales - GeneaWiki ». Consulté le 10 juin 2015.
http://fr.geneawiki.com/index.php/Projet_Archives_notariales.

¹⁰² « Ministère de la Défense - Mémoire des Hommes ». Consulté le 10 juin 2015.
<http://www.memoiredeshommes.sga.defense.gouv.fr/article.php?laref=1>.

aux internautes disposant d'un espace personnel. Au niveau des règles de saisie, la visionneuse permet une indexation guidée par des info-bulles. Les champs d'indexation sont les suivants : grade, unité, lieu de naissance, bureau de recrutement, classe, matricule au recrutement, date de décès, lieu de décès (département, pays), transcription du décès (lieu, département). Aucun de ces champs d'indexation n'est obligatoire. En effet, les index sont interrogeables indépendamment les uns des autres. Selon Sandrine Aufray, responsable TIC du ministère¹⁰³, ce projet a été pensé comme un hommage à la mémoire de nos ancêtres, et comme une participation à la politique mémorielle et historique de notre territoire et du conflit. Le bilan est très positif : en mars 2015, 102 205 indexations ont déjà été réalisées, en complément des 47 609 indexations réalisées par le webmestre, soit 10% de la base, qui compte 1,4 million de fiches. Les dix premiers annotateurs cumulent à eux seuls 31 485 indexations et les deux premiers le chiffre de 10 448. Quant aux cinquante premiers, ils ont réalisé au moins 551 indexations. Le 29 janvier 2015, le compte Twitter « 1 jour, 1 poilu » (@1J1P) annonce un pic d'indexation record : « Indexations sur *Mémoire des Hommes* : 2^e meilleur résultat depuis le lancement du programme : 487 fiches/jour »¹⁰⁴, avec un premier pic le 26 novembre 2014 avec 522 indexations ce jour.

PhotosNormandie

Le projet PhotosNormandie¹⁰⁵ est une initiative privée, qui ne dépend pas des archives départementales. Il est porté par Michel Le Querrec et Patrick Peccatte, auteur du Carnet de recherche Déjà vu¹⁰⁶. Ce projet collaboratif repose sur la plateforme *Flickr*, qui héberge plus de 3000 photos de la bataille de Normandie (6 juin-fin août 1944). La plupart de ces photos proviennent des Archives Nationales des Etats-Unis (NARA) et du Canada. Elles ont été téléchargées librement à partir de la base de données Archives Normandie 1939-1945¹⁰⁷, du Conseil général de Basse-Normandie, à

¹⁰³ Réponse à notre questionnaire du 09/03/15

¹⁰⁴ « 1 Jour - 1 Poilu sur Twitter : "Indexations sur Mémoire des Hommes : 2e meilleur résultat depuis le lancement du programme : 487 fiches/jour CC #1J1P <http://t.co/YlBprGSlmg>". » Consulté le 10 juin 2015. <https://twitter.com/1J1Poilu/status/560886350669709315/photo/1>.

¹⁰⁵ « PhotosNormandie | Flickr ». Consulté le 10 juin 2015.

<https://www.flickr.com/photos/photosnormandie/>.

¹⁰⁶ « Déjà Vu | Carnet de recherche de Patrick Peccatte ». Consulté le 10 juin 2015.

<http://dejavu.hypotheses.org/>.

¹⁰⁷ Le site « Archives Normandie 1939-1945 » n'est plus en ligne, mais il est archivé par la *Wayback machine*

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

LES STRATEGIES

Quel que soit le type d'activité et les outils sollicités pour dans la mise en place d'un projet de crowdsourcing, la stratégie envers les contributeurs est aussi un facteur déterminant. Elle peut prendre plusieurs formes : événement ponctuel, activité à dimension ludique ou bien instrument à remonter le temps.

Des événements ponctuels

Challenge

Lancer un défi peut constituer une bonne stratégie pour stimuler les contributeurs. Il suffit de déterminer un objectif et de fixer une durée pour l'atteindre. Dans le cas des activités de transcription, cet objectif se traduit en nombre de pages à transcrire. C'est la stratégie adoptée cette année par les Archives nationales des Etats-Unis, avec la *Sunshine Week*, qui a eu lieu du 16 au 23 mars 2015¹⁰⁸. Durant cette semaine, le but était de transcrire 1000 pages d'archives. Ce « transcription challenge » a été soumis aux *Citizen archivists*, autrement dit des contributeurs déjà investis dans ce type d'activité. L'accroche était « Help us unlock history », en français : « Aidez-nous à déverrouiller l'histoire ». Parmi les missions proposées, des types de documents très divers comme des lettres d'amour entre Lyndon et Lady Bird Johnson, des archives sonores sur l'attaque du 11 septembre, des publicités pour des marques d'alcool.. Pour chaque mission, le niveau de difficulté était précisé. L'objectif a été largement atteint

¹⁰⁸ « Sunshine Week | Citizen Archivist ». Consulté le 10 juin 2015. <http://www.archives.gov/citizen-archivist/transcribe/#sunshine>.

puisqu'il y a plus de 2500 pages ont été transcrites en l'espace d'une semaine, et 10 000 tags ajoutés au catalogue des Archives nationales. Ce challenge a été suivi par le réseau social Twitter grâce au hashtag #1000pages, invitant les contributeurs à partager leur expérience de contribution sur le compte @USNatArchives

Concours

La dynamique du concours est également un bon moyen de drainer des participants. Elle peut permettre de développer des projets innovants, tels que le concours de photographie *History Happens Here !* initié par les Archives nationales des Etats-Unis en 2010. Le projet repose sur le principe de réalité augmentée, basé sur la superposition en 2D de deux images. L'effet recherché est un *mashup* des deux images : une photographie d'archives intégrée au paysage moderne. Pour participer au concours, il suffisait de choisir une photographie d'archives disponible sur la banque d'images du compte *Flickr : History Happens Here !*¹⁰⁹, puis de se rendre sur le lieu où cette photo a été prise. Ces photos, qui datent de 30, 80 ou 150 ans en arrière sont issues des collections des archives nationales : il peut s'agir de paysages, de monuments, mais aussi de rassemblements. A noter que les photographies proposées sont toutes géotaguées, ce qui permet de retrouver plus facilement le lieu. Une fois en place, la photographie doit venir se superposer à la réalité, ce qui crée un décalage, parfois étonnant, entre l'ancien et le moderne. Le photographe devait alors poster sa composition sur le groupe dédié de *Flickr*, en n'oubliant pas de mentionner les références de la photographie d'archives ni de citer son url pérenne dans la description. Le concours a pris fin le 21 janvier 2011. Parmi les 70 photos postées sur le groupe, 20 gagnants ont été sélectionnés.

Edit-a-thon

Enfin, le crowdsourcing peut également s'organiser comme des « journées contributives », aussi appelées « edit-a-thon ».

Cette expression amusante est un mot-valise créé à partir du mot téléthon, qui désigne la course d'endurance, pratiquée dès l'antiquité. Par extension le suffixe -thon

¹⁰⁹ « History Happens Here! | Flickr ». Consulté le 10 juin 2015.
<https://www.flickr.com/photos/usnationalarchives/sets/72157624674340229/>.

renvoie à tout concours ou événement d'une durée plus ou moins longue, souvent organisé dans un but précis, à l'image du *hackathon*, destiné aux développeurs, qui propose de la programmation informatique collaborative.

Ces sessions de contribution sont ouvertes à tous : wikipédiens, étudiants ou encore néophytes, et permettent d'améliorer l'encyclopédie libre de *Wikipédia*, et plus largement les projets *Wikimedia*.

Ces rencontres ont généralement lieu dans des centres d'information et de documentation, ou structure analogue, suivant l'organisme porteur du projet. L'accès à des ressources, numériques ou non permettant aux participants d'effectuer des recherches sur place, sur un sujet défini au préalable.

Les bénéfices de ces journées contributives sont nombreux. Au niveau des participants, il les aide à approfondir leurs connaissances sur un sujet donné. Du point de vue des internautes, il améliore significativement la qualité des articles, qui sont enrichis de bibliographies et de références. Au niveau de l'organisme, ce type de projet permet la valorisation d'une ou plusieurs collections d'archives, mais est aussi l'occasion de tisser des liens avec des contributeurs locaux. Enfin, au-delà de l'enrichissement des contenus de *Wikimedia*, ces ateliers sont aussi l'occasion d'initier de nouveaux contributeurs à la multicontribution dans l'encyclopédie libre.

Pour ce qui est de la nature des archives exploitées, il peut s'agir d'ouvrages sur un sujet donné, mais aussi d'images, de son et de médias appartenant au domaine public pouvant être versées dans *Wikimedia Commons*, une plateforme analogue à *Flickr The Commons* pour les photographies d'archive. Dans le champ des archives sonores, le premier projet de ce type a eu lieu le 15 janvier 2015, dans le cadre du lancement d'*Europeana sounds*. La session avait pour thème universel le chant des oiseaux.¹¹⁰ La journée a été organisée par l'Institut de l'Image et du Son des Pays-Bas (Netherlands Institute for Sound and Vision) en partenariat avec Xeno-Canto, une base de données internationale spécialisée dans le son des oiseaux.

¹¹⁰ « Europeana Sounds - Early Christmas gift for bird watchers and Wikipedians.. » Consulté le 10 juin 2015. <http://www.europeanasounds.eu/news/december-16-early-christmas-gift-for-bird-watchers-and-wikipedians>.

Ce genre d'initiative peut être mis en relation avec le mouvement du Libre Accès, en tant qu'il contribue à la mise en ligne de contenus en accès gratuit et libre.

Une expérience ludique

Les metadata games

On constate l'émergence de modules de jeux en ligne, dont l'objectif est de collecter des métadonnées sur des images : les *metadata games*¹¹¹. Le but du jeu est de renseigner un maximum de tags en un minimum de temps possible -2 minutes en moyenne-. Certains jeux sont basés sur l'affrontement avec un adversaire (*partner*) parmi les internautes, ce qui permet de comparer les réponses et donc de renforcer la fiabilité des données, comme pour le jeu « Guess What ». Le mini-jeu intitulé « Stupid robot » propose lui, d'apprendre de nouveaux mots à un robot, en respectant une longueur progressive : d'abord 4 lettres puis 5 et 6 lettres etc...Ce critère de longueur permet de varier le vocabulaire employé pour désigner une même image. Chaque jeu se déroule en plusieurs manches, et permet de marquer des points : le contributeur peut ainsi suivre son évolution. Certains jeux existent également sous forme d'application sur smartphone, disponible sur Google Play ou Apple Store.

Outre les photographies, ces modules peuvent également servir à identifier des archives audiovisuelles comme le mini-jeu « Nextag », qui est le pendant du jeu « Zen tag » pour les images. Du côté des institutions, la British Library s'est approprié ce concept pour proposer l'identification de ses collections iconographiques à travers plusieurs modules de jeu : « ships tag », pour les collections navales, « book tag » pour les livres numérisés et « portrait tag » pour les portraits de personnalités politiques du 16^{ème} et 18^{ème} siècle. Enfin, le site présente un moteur de recherche multicritère pour permettre de retrouver les documents taggés par les usagers. Jusqu'à présent, grâce au site, 36 073 tags ont été proposés en 26 802 parties.

¹¹¹ *op. cit.*

La gamification

Les plateformes de crowdsourcing peuvent également adopter une stratégie de *gamification*. Ce néologisme issu de l'anglais désigne le fait de reprendre les mécanismes du jeu dans un domaine non ludique, comme les archives. Un exemple emblématique souvent cité est Digitalkoot¹¹², programme de correction collaborative d'OCR mis en place par la Bibliothèque nationale de Finlande. En français, on traduit le terme de « digitalkoot » par « volontaires numériques ». Le but du projet est de vérifier l'océrisation des archives de journaux de la bibliothèque. Ce projet a été réalisé par la société *Microtasks*, qui comme son nom l'indique transforme le travail de correction en micro-tâches ludiques. Dans le cas de Digitalkoot, la mascotte est une taupe. Pour chacun des jeux, un tutoriel est proposé sous forme vidéo.

Dans le jeu « Chasse aux taupes »¹¹³ (*Myyräjahti* en finnois, *Mole hunt* en anglais), deux taupes montrent chacune au joueur un mot inscrit sur une pancarte – il s'agit du mot scanné et de sa transcription- afin qu'il vérifie la correspondance entre les deux graphies : il clique alors sur la croix rouge si le mot est non conforme et sur une icône de validation si le mot est conforme. Il s'agit d'un exercice de comparaison visant à repérer les erreurs éventuelles.

Dans le jeu « Pont des taupes »¹¹⁴ (*Myyräsilta* en finnois, *Mole bridge* en anglais), le joueur doit saisir la transcription du mot scanné. Si le mot est reconnu comme conforme, une planche est ajoutée au pont qui permettra à la taupe de traverser. Si une erreur est détectée, la planche est détruite. Si le mot scanné est illisible il a la possibilité de sélectionner un encart portant la mention « impossible ». Il s'agit d'un exercice de transcription pure.

Le jeu est donc pensé pour détecter les réponses fausses en se basant sur l'ensemble des réponses fournies par les précédents joueurs. Ce projet atypique a été couronné de succès : près de 110 000 correcteurs ont réalisé plus de 8 millions de tâches.

¹¹² « DigiTalkoot ». Consulté le 10 juin 2015. http://www.digitalkoot.fi/index_en.html.

¹¹³ « How to play Mole Hunt - YouTube ». Consulté le 10 juin 2015. <https://www.youtube.com/watch?v=G7gXkdSPXWQ>.

¹¹⁴ « How to play Mole Bridge - YouTube ». Consulté le 10 juin 2015. <https://www.youtube.com/watch?v=9-W9cf9u9Qw>.

Mission détective

Au-delà des activités de *tagging* et de transcription de documents d'archives, les internautes peuvent aussi être amenés à mener de véritables enquêtes à la manière de détectives. C'est la stratégie employée par Museum Games¹¹⁵. Ce projet ludique a été mis en place par Mia Ridge, chercheuse dans le domaine du patrimoine digital. Ces jeux ont été développés à destination des musées en 2011-2012 dans le cadre de son mémoire sur la place du jeu dans le crowdsourcing¹¹⁶.

Les deux jeux proposés en exemple s'appuient sur des objets du musée et des illustrations provenant de livres numérisés. Les photographies proposées proviennent des collections de la British Library et sont en ligne sur *Flickr The Commons*. Elles comportent déjà les métadonnées minimales comme le titre, l'auteur, et bien souvent la date et le lieu. Malgré le niveau de connaissance demandé, la tonalité des jeux est plutôt enfantine, à en juger par le nom des jeux et leur mise en scène.

Le premier jeu est intitulé « Dora's lost data », c'est à dire les données perdues de Dora. En cliquant, un avatar féminin apparaît et nous confie cette mission :

« Salut, je m'appelle Dora et je suis une jeune conservatrice. C'est mon premier jour et j'ai fait une grosse bêtise : j'ai accidentellement supprimé toutes les informations que nous allions ajouter à nos collections. J'ai besoin de les cataloguer de nouveau, et vite... »

Il s'agit donc d'une opération de *tagging* libre. Lorsque l'image est tournée dans le mauvais sens il est possible de le signaler à l'aide du tag : « rotatec » ou « rotatecounterc ».

Le deuxième jeu, « Donald's detective puzzle » est plus ambitieux. Cette fois-ci la mission est confiée par un avatar masculin :

« Salut, Holmes ! Dieu merci tu es là ! Peux-tu aider à résoudre le Cas des Objets Mystérieux ? L'ignoble Moriarty a laissé derrière lui ces objets, mais nous ne savons pas pourquoi. Peux-tu utiliser les informations présentes

¹¹⁵ « Museum Metadata Games ». Consulté le 10 juin 2015. <http://museumgam.es/>.

¹¹⁶ Ridge, Mia. « Playing with difficult objects: game designs for crowdsourcing museum metadata ». City University London School of Informatics MSc in Human-Centred Systems, 2011.

sur la page pour trouver d'autres renseignements intéressants, ou donner des liens à propos les images ci-dessous ? »

Lorsqu'il accepte le challenge, l'internaute a la possibilité de choisir l'objet sur lequel il souhaite enquêter. La réponse doit être remplie à l'aide d'un formulaire mentionnant la référence de l'objet (*report title*), le fruit de la recherche (*information discovered*) et la source (lien, livre, ou personne ayant fourni cette information). Relever ce dernier défi vaut 250 points de mérite au détective.

Une time-machine

L'expérience du crowdsourcing peut aussi offrir aux contributeurs un voyage dans le temps, qu'il s'agisse d'un retour dans le passé à travers des personnes ou des lieux, et bien d'une projection dans le futur.

Hommage au passé

Par l'intermédiaire des médias sociaux, le crowdsourcing peut faire ressurgir des individus appartenant au passé. Ce phénomène est particulièrement visible à travers l'hommage rendu aux anciens combattants. Sur *Twitter*, il existe un compte consacré aux soldats de la première guerre mondiale : @1J1P qui signifie 1Jour 1Poilu¹¹⁷. Il s'agit d'une initiative du Ministère de la Défense, qui va de pair avec son site collaboratif *Mémoire des Hommes*. Chaque jour, un soldat est mis à l'honneur à travers des *tweet* émouvants ou plus humoristiques comme celui-ci « Il y a 100 ans, au petit matin, un homme était exécuté pour une affaire de pantalon »¹¹⁸. Cette valorisation des archives transcrites par les utilisateurs, permet également d'inciter d'autres internautes à se lancer dans l'indexation collaborative. En l'espace de quelques mois, le compte @1J1P compte déjà plus de 2000 abonnés.

¹¹⁷ « 1 Jour - 1 Poilu (@1J1Poilu) | Twitter ». Consulté le 10 juin 2015. <https://twitter.com/1J1Poilu/>.

¹¹⁸ « 1 Jour - 1 Poilu sur Twitter ». Consulté le 10 juin 2015. <https://twitter.com/1j1poilu/status/566144904242671616>.

Sur le même modèle, le compte @1J1MAM, 1 jour 1 monument aux morts (Des Pierres des Noms)¹¹⁹ est consacré aux monuments érigés en l'honneur des combattants, présents dans quelques 36 000 communes en France et qui tombent peu à peu dans l'oubli. Les internautes sont invités à photographier ces monuments divers : obélisques, colonnes, statues puis à en relever les inscriptions. Cette initiative s'inscrit elle aussi dans un devoir de mémoire, et permet de rendre hommages aux hommes qui il y a 100 ans, se sont battus pour la patrie.

Géolocalisation du passé

Grâce aux données cartographiques du passé, nous pouvons étudier la configuration de certains lieux. Cette plongée dans la cartographie ancienne est proposée par la New York Public Library à travers son site interactif Building Inspector¹²⁰ mis en place dans le cadre du projet NYC Space/Time directory. L'objectif est de transformer les sources historiques et géographiques en un service numérique permettant de voyager dans le temps dans la ville de New York. Un appel au « citizen cartographers » a été lancé avec pour slogan « Kill Time. Make History » = « Tue le Temps. Refait l'Histoire ».

Pour cela des atlas entiers ont été récupérés chez des compagnies d'assurance puis numérisées : elles remontent au 18^{ème} siècle pour les plus anciennes. Du côté du contributeur, plusieurs tâches sont proposées : dessiner les délimitations (*fix footprints*) et indiquer s'il s'agit ou non d'un bâtiment (*check footprints*), saisir les numéros (*enter addresses*), sélectionner la couleur du bâtiment (*classify color*) indiquant leur nature - les bâtiments en rose sont des résidences en briques ou pierre, en jaune les maisons en ossature, en vert des bâtiments au contenu incertain, en gris les hangars- et saisir les noms de lieux (*find placenames*) comme par exemple « pharmacie ». Ces tâches peuvent être effectuées dans l'ordre souhaité, avec la possibilité de passer en mode aléatoire grâce au module *surprise me*. Toutes les informations récoltées serviront à élaborer une carte interactive capable de retracer l'historique de l'aménagement, année après année, couche après couche grâce à une géolocalisation des données.

¹¹⁹ « DesPierresDesNoms (@1J1MAM) | Twitter ». Consulté le 10 juin 2015. <https://twitter.com/1J1MAM>.

¹²⁰ « Building Inspector by NYPL Labs ». Consulté le 10 juin 2015. <http://buildinginspector.nypl.org/>.

Retour vers le futur

Dans une démarche inverse, les projets de crowdsourcing peuvent aussi s'orienter vers le futur, en s'adressant aux générations à venir. C'est le projet des archives nationales d'Estonie, qui ont annoncé leur projet de graver sur pierre des messages adressés aux habitants de 2115¹²¹. Cette inscription apparaîtra sur la pierre angulaire du nouveau bâtiment des archives à Tartu, deuxième ville d'Estonie. Les travaux devaient commencer ce 15 avril 2015 et finir à la fin du printemps 2016. Pour participer à ce projet de grande envergure, chaque internaute est libre d'écrire son message dans la langue de son choix, et de le poster sur Twitter en utilisant le hashtag #archiiv2115. Les premiers messages tournent autour du patrimoine culturel et montrent l'importance la mémoire nationale. Reste à savoir ce qu'en penseront les Estoniens dans 100ans. A noter que pour assurer la pérennité des messages, la pierre a été préférée au numérique !

Les institutions souhaitant développer ce type de projet, se retrouvent donc face à des choix en ce qui concerne les pratiques à développer, les outils à mettre en place, ainsi que les stratégies à déployer. L'objectif étant de susciter la motivation des publics.

¹²¹ « Crowdsourcing für die Grundsteinlegung im estnischen Nationalarchiv ». Consulté le 10 juin 2015. <http://archive20.hypotheses.org/2363>.

LES MOTEURS DE LA PARTICIPATION

DU COTE DES PUBLICS, LES FACTEURS DE MOTIVATION

D'après les modèles théoriques classiques sur les motivations des publics, on distingue les motivations intrinsèques des motivations extrinsèques. Tandis, que la motivation intrinsèque est liée à l'activité en tant que telle, la motivation extrinsèque elle, relève d'une satisfaction extérieure. Dans le premier cas de figure, l'activité de crowdsourcing est donc une fin en soi, tandis que dans le deuxième cas, il ne s'agit que d'un moyen pour atteindre un objectif. Pour cette typologie des motivations, nous nous appuyons sur l'étude de Alam et Campbell qui synthétisent les grandes théories, tout en les appliquant au secteur des GLAM ¹²².

Motivations intrinsèques

Intérêt personnel

Les motivations relevant de l'intérêt personnel sont souvent les plus représentées, dans les études menées auprès des publics.

En effet, le choix d'une activité collaborative est souvent en lien avec les centres d'intérêt du contributeur. En même temps que d'aider la communauté, l'internaute se documente aussi pour lui-même. Bien souvent, dans le cadre de l'indexation des documents nominatifs, on se rend compte que les contributeurs recherchent d'abord leur nom de famille, ou bien leur commune, car ce sont des informations qui les concernent directement. En dehors de leur histoire personnelle, les internautes choisissent bien souvent des sujets qui les intéressent, qu'il s'agisse de sciences, d'histoire ou même d'animaux. Les activités proposées dans le cadre du crowdsourcing sont aussi l'occasion d'emmagasiner des connaissances de manière non scolaire, ce qui peut aussi constituer un facteur de motivation.

En dehors de l'attractivité du sujet, le contributeur peut aussi apprécier les tâches qui lui sont proposées. Par exemple, les tâches de correction de texte sont généralement

¹²² Alam, Sultana Lubna, et John Campbell. « Crowdsourcing motivations in a not-for-profit GLAM context : The Australian Newspapers Digitisation Program ». 23rd Australasian Conference of Information Systems. Geelong, 3 décembre 2012.

perçues comme « addictives », si bien que certains correcteurs peuvent y passer plusieurs heures par jour. Selon l'étude de Holley, les correcteurs les plus acharnés peuvent y consacrer 40 à 60 heures par semaine, l'équivalent d'un temps plein, tandis que les 200 correcteurs les plus actifs y consacrent l'équivalent d'un mi-temps.

Un autre facteur de motivation est le fait de relever des défis. On remarque que plus le volume de documents en ligne est important, plus les participants ont la volonté de faire progresser ce travail collaboratif. D'autre part, les internautes sont reconnaissants des ressources mises à leur disposition, d'où ce sentiment d'être redevable. A noter qu'ils travaillent malgré tout pour leur profit personnel en améliorant l'accès aux ressources.

Enfin, paradoxalement, on se rend compte que plus l'internaute est laissé libre dans l'exercice de son l'activité, plus il se montre consciencieux et fournit un travail de bonne qualité. La liberté est donc un facteur de motivation pour l'internaute.

Intérêt collectif

La motivation des internautes peut aussi reposer sur le sentiment d'appartenance à une communauté. Les contributeurs apprécient ce sentiment de faire partie d'un tout, de participer à quelque chose qui les dépasse. Nous pouvons reprendre l'exemple des généalogistes, ces communautés de passionnés qui partagent un même intérêt et s'entraident entre pairs. Au sein d'un projet collaboratif, les contributeurs recherchent aussi la création d'une communauté dédiée, afin de pouvoir échanger entre eux et de partager leurs expériences. Dans ce but, certains sites prévoient un espace social, permettant l'interaction entre les membres. Un exemple de module social est le forum, qui permet aux internautes de poser leurs questions. La communauté peut également s'étendre en dehors du site web, sur les réseaux sociaux ou bien un blog consacré. Tous ces outils contribuent à souder les membres autour d'un projet. La motivation n'est donc plus centrée sur le bien-être d'un individu, mais sur une entité à laquelle il appartient : on appelle cela le collectivisme. Ce sentiment peut s'apparenter un certain à « patriotisme » lorsqu'il s'agit de reconstituer l'histoire d'un pays, comme c'est le cas de nombreux projets initiés par les Archives nationales. L'objectif est alors de créer une unité nationale autour d'un projet donné.

Plaisir personnel

Enfin, le plaisir que procure l'exercice d'une activité peut être un facteur de motivation. En effet, plus l'activité est agréable, plus le contributeur aura envie de s'y attarder. Ce plaisir peut passer par l'ergonomie des interfaces, des fonctionnalités originales, un espace personnel -par exemple galerie personnelle-, et bien sûr la *gamification*. L'internaute ne doit pas être découragé par des tâches difficiles ou trop restrictives. La motivation passe par la simplicité des tâches proposées et leur dimension agréable, voir ludique. Cette activité peut être vue pas certains comme un passe-temps, ou bien une manière de se détendre par le biais de tâches répétitives. Cette dimension de plaisir n'est donc pas à négliger car elle est véritablement au cœur des motivations des internautes.

Motivations extrinsèques

Les contributeurs peuvent aussi être motivés par autre chose que l'activité elle-même. Ces motivations que l'on qualifie d'extrinsèques peuvent être de plusieurs ordres.

L'appât du gain

Dans de nombreux cas, la rétribution financière peut être un facteur de motivation. En effet, comme « tout travail mérite salaire », le contributeur peut recevoir une compensation pour son travail. Ce modèle économique est surtout valable pour les projets qui à terme vont générer du profit. Dans son étude sur les motivations des foules, Daren C. Brabham prend l'exemple de la plateforme de crowdsourcing *Threadless*¹²³. Dans le cadre d'un concours de design pour des tee-shirts, de nombreux prix ont été mis en jeu, jusqu'à 2000 euros, ce qui n'a pas été sans influence sur les motivations des participants. Dans le cas de l'*Amazon Mechanical Turk*, que nous avons déjà évoqué, les contributeurs sont rétribués à la tâche. Toutefois, dans le cas des projets mis en place par les GLAM, le modèle est différent, dans la mesure où les données ne sont pas recueillies dans un but lucratif. Par conséquent, les institutions culturelles doivent trouver des leviers de motivations, autres que l'argent.

¹²³ Brabham, Daren C. « Moving the crowd at Threadless : motivations for participation in a crowdsourcing application ». *Information, Communication & Society* 13, n° 8 (2010): 1122-45.

La reconnaissance sociale

Lorsqu'ils font preuve d'une activité importante, les contributeurs peuvent être remerciés personnellement, par le biais de messages privés, comme c'est le cas aux Archives départementales du Lot-et-Garonne¹²⁴ pour les donateurs de cartes postales. Les remerciements peuvent aussi passer par la citation du nom du contributeur sur le site, comme c'est le cas des correcteurs dont le nom apparaît sous le texte corrigé. Il est également possible, en cherchant sur le profil d'une personne de voir les travaux auxquels il a participé, mais aussi ses activités sociales sur le site. Ces remerciements sont une forme de reconnaissance sociale, de même que les systèmes de gratification symboliques. En effet, faute de pouvoir être rétribués, les contributeurs peuvent recevoir des prix symboliques, les récompensant pour leur activité. Cette valorisation peut passer par un classement des meilleurs contributeurs (*ranking table*). Par exemple, la plateforme *Correct* de la BnF met en avant les 6 meilleurs correcteurs sur sa page d'accueil. Ce procédé permet donc à la fois de donner une visibilité à ces contributeurs et d'introduire un esprit de compétition entre eux, afin de les encourager à se surpasser. Les Archives Nationales d'Australie ont été encore plus loin en mettant en place un « *Hall of fame* », littéralement le « Temple de la renommée » pour tous les correcteurs ayant transcrit plus de 1500 lignes en un mois dans le cas du projet *Trove*¹²⁵. Quant aux 5 premiers correcteurs, ils ont été invités à Canberra en 2010 pour la fête nationale de l'Australie afin de rencontrer les membres du projet ANDP (Australian Newspaper Digitisation Program)

E-reputation

Enfin, participer à un projet collaboratif peut également s'avérer valorisant pour l'internaute dans son parcours futur. En effet, pour l'*e-reputation* de l'internaute - autrement dit sa réputation sur la toile- cette expérience peut être vue comme un atout et susciter d'autres opportunités. En effet l'investissement dans ce type de projet dénote d'un certain engagement à l'égard de la communauté. En plus de cet investissement positif dans la sphère du web 2.0, l'internaute acquiert des savoir-faire grâce aux outils participatifs, mais aussi des savoirs bruts dans des domaines bien définis. L'ambition de certaines internautes peut être de tirer des fruits de cet apprentissage afin de l'appliquer dans d'autres domaines. En effet, la pratique du crowdsourcing peut donner des idées

¹²⁴ *op. cit.*

¹²⁵ *op. cit.*

d'exploitations collaboratives, en dehors du cadre des documents d'archives. Elle permet de mieux comprendre le potentiel du web et des données générées par les utilisateurs et d'innover dans ce sens.

DU COTE DES INSTITUTIONS, LES FACTEURS DE REUSSITE

Pour susciter la motivation des publics, les institutions doivent pouvoir s'appuyer sur quelques règles de conduite pour garantir le succès du projet de crowdsourcing.

Ingrédients d'un site réussi

Visibilité

Avant tout chose, l'institution doit pouvoir rendre son projet visible des internautes, sans quoi la participation ne sera pas au rendez-vous. Avant le lancement, l'étape préliminaire de la promotion est déterminante car elle permet de cibler le public. Comme nous avons pu le constater, le public des services d'archives publiques est autant constitué d'habitues que de néophytes, d'où l'intérêt de communiquer au sein de l'institution, mais aussi en dehors. Cet appel à la participation peut se faire sur le site institutionnel, au niveau de la page d'accueil ou bien dans l'onglet « actualités ». Il peut être véhiculé par les réseaux sociaux de l'institution, qui permettent un relai rapide de l'information. Lorsque le projet est réalisé à l'échelle départementale, les médias locaux peuvent également constituer un bon relai d'information, comme le site du Conseil Général, ou bien la presse locale. Après le lancement, il est essentiel de veiller à la visibilité du projet sur le long terme. En effet, nous avons pu remarquer au cours de notre étude que sur les sites des services d'archives, les modules d'indexation collaborative ne sont pas directement accessibles. Sur une grande majorité de sites, cette fonctionnalité collaborative n'est pas mise en avant, et il faut parfois visionner les documents en ligne pour voir apparaître le module d'annotation. De surcroît, il faut souvent être identifié pour accéder à ce module. D'autres services d'archives au contraire ont valorisé cette démarche avec un onglet dédié à la participation, proposant différentes manières de collaborer sur les collections.

La visibilité des projets de crowdsourcing passe aussi par la visibilité des collections sur le web. En effet, plus les documents seront faciles d'accès, plus il sera facile de les indexer. Enfin, ne négligeons pas la visibilité du site sur les moteurs de recherche, en particulier Google qui est la plus populaire, d'où l'intérêt d'être bien référencé sur ce géant du web.

Fiabilité

Un gage de la qualité du site est aussi sa fiabilité. Pour garantir cette fiabilité, l'institution doit prévoir certains paliers de sécurité sur son site. En effet, le problème majeur soulevé par les activités collaboratives est le risque d'erreur. Ces erreurs peuvent être commises par inadvertance, ou bien intentionnellement par des usagers parasites. Ce risque d'abus peut être limité de différentes manières. Bien souvent, le fait de devoir créer un compte pour avoir accès aux modules collaboratifs est un élément dissuasif. Au moment de l'inscription, l'utilisateur doit souvent signer une charte de bonnes pratiques l'informant des risques encourus en cas d'abus. Pour se prémunir des erreurs commises par les usagers, une même activité peut être soumise à plusieurs internautes, afin de garantir un résultat plus fiable. Ce mode de fonctionnement repose sur le principe de sagesse collective, selon lequel le résultat d'un grand nombre de personnes est toujours meilleur que la réponse d'un seul individu. Lorsque l'activité requiert une expertise particulière, comme des connaissances en manuscrits anciens, le site peut décider de mettre en place un test de paléographie afin de vérifier et d'évaluer le niveau de l'internaute. Enfin, outre les individus malveillants, le site doit aussi se protéger des robots, qui peuvent venir fausser les résultats. Pour cela, les modules de type *Captcha* peuvent se révéler efficaces.

Popularité

Enfin, un site réussi est un site qui attire de l'audience. Pour gagner en popularité, le site doit être attractif et convivial. En effet, nous avons constaté qu'un facteur de motivation déterminant chez les publics est le principe de plaisir. L'ergonomie du site doit donc être travaillée de manière à rendre la navigation agréable. La présentation sous forme ludique est également un atout. Le fait de gagner des points ou bien d'obtenir des

récompenses pour son activité permet d'entretenir la motivation de l'internaute et de créer chez lui un sentiment de reconnaissance. L'intégration d'espaces réservés à la sociabilité est aussi considérée comme un « plus » pour attirer les internautes. Pour créer du lien entre les contributeurs, le site peut proposer un forum de discussions, ou encore d'entraide. Au-delà de la communauté d'internautes, le rapport avec l'institution doit aussi être privilégié. Pour ce faire, une adresse électronique peut être créée permettant à l'internaute d'adresser librement ses questions, soumettre ses idées. Dans ce cas, les acteurs du projet doivent prévoir un temps de réponse rapide.

Recommandations

Gestion de projet

Pour mettre en place des projets adaptés à leurs besoins, les institutions doivent commencer par se poser les bonnes questions. Avant toute chose, il faut définir des objectifs précis afin d'opter pour une activité collaborative adaptée. Dans un premier temps, il peut être intéressant de vérifier si des sites analogues n'existent pas déjà, sur un projet similaire : une démarche de *benchmarking* des sites existants peut se révéler utile. L'institution peut alors s'inspirer de ces exemples et l'adapter à son propre cas. En parallèle, il s'agit de cibler les médias sociaux qu'il pourrait-être intéressant d'exploiter, comme par exemple *Flickr the Commons*. Ces médias sociaux doivent aussi remplir des objectifs bien précis : disséminer des contenus, servir de plateforme pour indexer, transcrire des archives... Outre l'outil, il s'agit de sélectionner des critères qui permettront de suivre la progression du projet, en vue de son évaluation: ces critères peuvent être quantitatifs mais aussi qualitatifs. Les objectifs doivent également être au service d'une politique qu'il s'agit de déterminer en amont : quelle politique instaurer avec les contributeurs ? Du point de vue de l'équipe en charge du projet, il faut également prévoir un encadrement des administrateurs, voire une formation sur les outils techniques qui seront déployés. Enfin, il s'agit également d'évaluer le temps humain et les ressources nécessaires à la réalisation du projet au sein de l'institution.

Checklist

Dans la revue spécialisée publiée en ligne, *D-Lib magazine*, *The Magazine of Digital Library Research*, Rose Holley, responsable de la bibliothèque nationale d'Australie, publie un état des lieux de la question dans le domaine des bibliothèques. Notons que l'institution australienne a été l'une des premières à lancer une plateforme de correction, pour les journaux australiens : *Australian Newspapers*. Dans sa publication de mars/avril 2010 intitulée « Crowdsourcing : How and why should libraries do it ? »¹²⁶, elle dresse une checklist récapitulant tous les éléments d'un projet réussi. Ces conseils, qui s'adressent en premier lieu aux bibliothécaires sont génériques et valables pour toutes les institutions de type GLAM. Ces éléments sont répartis dans un tableau comportant quatre colonnes sur : le projet en tant que tel (*the Thing*), la plateforme (*the System*), le public (*the People*) et le contenu (*the Content*).

The Thing	The System	The People	The Content
👍 Clear goal	👍 Easy and Fun	👍 Acknowledge	👍 Interesting
👍 Big challenge	👍 Reliable and quick	👍 Reward	👍 New
👍 Progress	👍 Intuitive	👍 Team support	👍 Lots
👍 Results	👍 Options	👍 Trust	👍 History /Science

Figure 8 : La checklist de Rose Holley pour un projet réussi

Concernant le projet, il doit tout d'abord afficher un objectif clair (*clear goal*). Plus cet objectif sera ambitieux, voire difficilement réalisable (*big challenge*), plus la participation sera importante. Par rapport à cet objectif, l'institution doit afficher les progrès accomplis (*progress*) soit de manière visuelle -à travers une barre d'avancement par exemple-, soit à travers des chiffres emblématiques (*results*).

Concernant la plateforme, l'environnement de travail doit être facile d'accès, intuitif (*intuitive*) mais aussi rapide et fiable (*reliable and quick*). D'autre part, plus l'activité sera simple et amusante (*easy and fun*), plus remportera le succès des contributeurs.

¹²⁶ Holley, Rose. « Crowdsourcing : How and Why Should Libraries Do it ? ». *D-Lib Magazine* [en ligne] vol. 16, n° 3/4 (2010).

Concernant les publics, ils doivent être laissés relativement libres de s'identifier et d'alimenter ou non leur profil. En se rendant visibles des autres usagers, ils montrent leur désir de reconnaissance (*aknowledge*). Pour valoriser les contributeurs les plus actifs, un classement peut être mis en ligne en guise de récompense (*reward*). Ce procédé encourage l'esprit de compétition entre les participants. Enfin, un espace dédié à la communication permet de créer un environnement dynamique qui agit comme un comité de soutien (*team support*). L'équipe doit aussi être à l'écoute de ses usagers, en particuliers les volontaires les plus dynamiques, appelés « *super volunteers* » qui sont force de proposition. Enfin, une bonne dose de confiance (*trust*) doit être accordée aux contributeurs, qui sauront apprécier cette liberté et en faire bon usage.

Concernant le contenu, il doit attirer l'attention des internautes en portant sur un sujet attractif (*interesting*). Lorsque le sujet s'y prête, il peut être mis en relation avec l'actualité. Un exemple donné dans l'article est le scandale de 2009 au Royaume-Uni au sujet des dépenses du Parlement, amenant 20 000 volontaires à s'inscrire sur le site pour transcrire des articles relatifs à cette actualité. La plateforme doit aussi être alimentée régulièrement avec de nouveaux contenus (*new*), pour continuer de susciter la motivation. Un autre conseil est de laisser le choix aux contributeurs quant aux contenus qu'ils veulent traiter, mais d'assigner des tâches plus aléatoires pour ceux qui n'auraient pas de préférence en attribuant des lots (*lots*). Enfin, le résultat du travail fourni par les internautes doit être transparent, et communiqué à travers des bilans d'activité. Un blog peut faire état des découvertes réalisées, ce qui valorise le travail des internautes.

Règles d'or

Dans son étude de mars 2012 sur le crowdsourcing, l'OCLC (Online Computer Library Center) dispense des recommandations à l'usage des bibliothèques, des archives et des musées. Ce rapport¹²⁷ est divisé en trois parties : une première partie consacrée à l'étude des sites existants (*Part 1 : Site reviews*), une deuxième partie consacrée à l'analyse de ces résultats (*Part 2 : Survey analysis*) et une troisième partie aux recommandations (*Part 3 : Recommendations and readings*). Dans cette troisième partie sont énoncés 18 principes clés pour mener à bien son projet de crowdsourcing :

1. Etablir des règles claires pour utiliser les médias sociaux

¹²⁷ Smith-Yoshimura, et Cyndi Shein. *Social Metadata for Libraries, Archives, and Museums*. OCLC, 2011-2012.

2. Motiver les usagers et les encourager à participer
3. S'inspirer des sites existants avant de se lancer
4. Inviter les contributeurs sans se soucier des spams et des abus
5. Adapter des règles de confidentialité ou en créer de nouvelles pour l'usage des médias sociaux
6. Préparer son équipe
7. Déterminer des critères d'évaluation
8. Considérer les avantages des sites tiers
9. Privilégier des logiciels open-source
10. Conduire des tests avant et après le lancement
11. Ajouter régulièrement des nouveaux contenus
12. Afficher et indexer les contenus crowdsourcés
13. Réfléchir comment introduire les contenus crowdsourcés dans le catalogue ou les métadonnées descriptives
14. Utiliser les réseaux sociaux pour fonder une communauté
15. Avoir une url pérenne pour le site et les contenus et les rendre visibles
16. Prévoir un plan de migration des données
17. Faire indexer les contenus par Google
18. Avoir un temps de réponse rapide

Retours d'expérience

Si nous nous plaçons à présent du point de vue des institutions qui ont déjà mis en place ce type de projet collaboratif, quels enseignements pouvons-nous tirer de leur expérience ? Nous nous appuyerons pour cela sur les entretiens menés auprès d'un échantillon de six institutions : la BnF pour la plateforme *Correct*¹²⁸, les Archives départementales du Lot-et-Garonne pour la cartothèque¹²⁹, les Archives départementales de l'Ain pour le module *SOS détective*¹³⁰, les Archives départementales d'Alpes-Maritimes pour le partenariat avec *Wikisource*¹³¹, les Archives départementales de

¹²⁸ Entretien avec Isabelle Josse

¹²⁹ Entretien avec Véronique Pouyadou

¹³⁰ Entretien avec Valéry Vesson

¹³¹ Entretien avec Hélène Cavalié

Vendée pour le *L@boratoire des internautes*¹³² et le Ministère de la Défense pour le site *Mémoire des Hommes*¹³³.

Encadrement des usagers

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

Administration du site

Au niveau de l'administration de ces sites, nous avons pu nous rendre compte à travers ces entretiens que le crowdsourcing n'est pas toujours un gain de temps pour le personnel. En effet, il faut systématiquement vérifier les métadonnées saisies par les internautes et les valider « au fil de l'eau ». Pour Véronique Pouyadou, responsable du

¹³² Entretien avec un responsable du service *L@boratoire des internautes*

¹³³ Entretien avec Sandrine Aufray

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

Bilan

Dans l'ensemble, le bilan des institutions est relativement positif, et encourageant pour les projets qui se trouvent dans une phase expérimentale comme la plateforme *Correct* de la BnF. Les données chiffrées fournies par ces institutions confirment les progrès réalisés et l'investissement des publics. Toutefois, ces projets doivent bénéficier d'un suivi sur la longue durée. En effet, le lancement de ce type de projet suscite généralement un enthousiasme au début, dû à l'attrait de la nouveauté, puis connaît parfois un certain essoufflement par la suite. C'est le cas de la cartothèque des AD de Lot-et-Garonne, qui la première année, en 2010, a reçu 300 cartes postales, puis a atteint une moyenne de 30 dépôts par an en 2015. La responsable du projet cherche donc un moyen de relancer l'intérêt du public, à travers des stratégies comme des concours. De même, les responsables de *SOS détective* aux AD de l'Ain nous ont présenté un bilan mitigé. En effet, sur près de 10 000 images mystères à identifier, seules 5 enquêtes ont été résolues à ce jour. Malgré le très bon référencement du site, classé premier du département, le projet n'arrive pas à trouver son public. Alors qu'il aurait à gagner à diffuser ces photographies sur des médias sociaux tels que *Flickr the Commons*, il se heurte aux questions juridiques posées par le Conseil Général.

L'investigation auprès des institutions nous a donc permis de nous rendre compte de la réalité du terrain et des difficultés auxquelles ils sont confrontés : manque de temps, de visibilité du projet, et parfois une certaine frilosité due aux valeurs archivistiques traditionnelles.

LES ENJEUX DU CROWDSOURCING

Enfin, le crowdsourcing n'est pas sans poser un certain nombre d'enjeux d'ordre juridiques et éthiques, mais aussi économiques.

Enjeux juridiques

La question de la propriété intellectuelle

Du point de vue juridique, le statut des données générées par les usagers pose question. En effet, à qui appartiennent les informations produites par le crowdsourcing ? Appartiennent-elles aux contributeurs ou bien aux institutions ? A partir de quand peut-on parler de « droit d'auteur » ? Pour répondre à cette question, nous avons interrogé Lionel Maurel¹³⁴, juriste et bibliothécaire et auteur du blog S.I. Lex¹³⁵.

Au regard de la loi française, deux cas de figure sont possibles : soit les contributions des internautes sont assimilables à des œuvres de l'esprit originales, auquel cas elles sont protégées par le droit d'auteur, soit elles ne le sont pas, auquel cas elles sont considérées comme des données publiques. Une œuvre de l'esprit doit remplir deux critères : l'originalité et la forme. Une œuvre est considérée comme originale lorsqu'elle contient « l'empreinte personnelle de son auteur ». Elle est protégée dès lors qu'elle est mise en forme, formulée selon les termes de l'auteur. Sa protection est alors assurée par le code de la propriété intellectuelle (CPI). La difficulté réside dans la définition de ce « seuil d'originalité ». Dans le cas des productions du crowdsourcing,

¹³⁴ Contacté par mail le 05/05/15

¹³⁵ « :: S.I.Lex :: | Au croisement du droit et des sciences de l'information. » Consulté le 11 juin 2015. <http://scinfolex.com/>.

on peut considérer que les activités telles que l'indexation, la correction collaborative ou encore le *tagging* ne relèvent pas d'une œuvre de l'esprit, car le contributeur n'y laisse pas son empreinte personnelle. En revanche l'activité d'écriture collaborative, comme c'est le cas des wiki, implique plus personnellement l'auteur. En effet, son style personnel transparaît dans le texte, les formulations employées : il ne s'agit pas seulement pour le contributeur de renseigner des informations, mais aussi de mettre en forme ses idées. En théorie, il détient donc un droit exclusif d'autoriser ou d'interdire les réutilisations sur sa production. A contrario, lorsqu'elles ne sont pas protégées par le droit d'auteur, ses données sont soumises à la loi du 17 juillet 1978 qui concerne les données produites ou reçues par les administrations publiques. Par conséquent, la réutilisation de ces données peut être soumise à des licences, gratuites ou payantes.

Les Conditions Générales d'Utilisation

Pour les sites proposant des activités de crowdsourcing, le cadre juridique est énoncé dans les Conditions Générales d'Utilisation (CGU). Ces conditions sont souvent soumises à l'internaute au moment de l'inscription, ou bien sont accessibles en bas de page dans une la rubrique « Mentions légales du site ». L'internaute doit donc prendre connaissance de ces conditions d'utilisation pour connaître ses droits.

Dans la plupart des cas, l'internaute doit renoncer à ses droits d'auteur et céder à l'institution ses droits de publication, de diffusion et de reproduction. C'est le cas du site d'indexation collaboratif *Mémoire des Hommes* qui stipule dans son « Engagement de l'annotateur »¹³⁶ que l'annotateur doit :

« donne[r] [s]on accord pour que le ministère de la défense (direction de la mémoire, du patrimoine et des archives) **jouisse, à titre gracieux, de tous les droits de publication, diffusion publique et de reproduction du travail accompli**, et principalement que le résultat de [s]a contribution soit publié sur le site internet *Mémoire des hommes*. ».

¹³⁶ « Engagement de l'annotateur - Mémoire des Hommes ». Consulté le 11 juin 2015.
<http://www.memoiredeshommes.sga.defense.gouv.fr/fr/article.php?laref=88&titre=engagement-de-l-annotateur>.

De la même manière, la plateforme de correction d'OCR de la BnF précise au sujet de sa plateforme *Correct*¹³⁷ :

« L'Utilisateur comprend et accepte que la suppression d'erreur et la restitution du texte d'origine **ne donnent pas lieu à la naissance de droits d'auteur** ou d'autres droits à son profit. Lesdites corrections ne constituent pas en effet une œuvre de l'esprit dotée d'originalité. Par ailleurs, l'Utilisateur agit de façon purement désintéressée et bénévole et peut cesser sa participation au moment où il le souhaite et sans formalités. »

Ces deux exemples d'indexation et de correction collaborative confirment donc la théorie selon laquelle ces activités ne sont pas considérées comme des œuvres de l'esprit. En revanche, si l'on prend le cas de la *Rosalipédie* de la Bibliothèque numérique de Toulouse, les articles collaboratifs qui composent l'encyclopédie sont protégés par le droit d'auteur¹³⁸ :

« **Les articles mis en ligne sont par défaut sous droit d'auteur.** Par conséquent, la bibliothèque ne modifiera pas le fond d'un article qui lui est soumis sans l'accord de l'auteur. Les contenus peuvent par ailleurs être placés par leurs auteurs sous une autre licence s'ils le souhaitent. Tout article reste la propriété du rédacteur. »

L'encyclopédie collaborative est donc soumise au respect de la propriété intellectuelle.

La question des licences est également à considérer car l'internaute peut fixer lui-même les conditions de réutilisation de son œuvre. Son choix peut être encadré par les licences *Creative Commons*¹³⁹ qui proposent 6 types de contrats, en autorisant ou non la réutilisation commerciale par exemple.

Les données à caractère personnel

Les données produites par les usagers peuvent également être assimilées à des données personnelles. Dans le cas de l'indexation collaborative, les données

¹³⁷ « Conditions générales d'utilisation de CORRECT ». Consulté le 11 juin 2015. <http://www.reseau-correct.fr/?action=terms-of-use>.

¹³⁸ « Contribuer - Rosalis ». Consulté le 11 juin 2015.

<http://rosalis.bibliotheque.toulouse.fr/index.php?contribute#.VXIQD0YwGbR>.

¹³⁹ « A propos des licences - Creative Commons ». Consulté le 11 juin 2015.

<https://creativecommons.org/licenses/>.

personnelles sont protégées par la délibération du 12 avril 2012 de la CNIL¹⁴⁰, qui fixe des délais pour l'indexation. L'article 4.3 portant sur « L'indexation des documents d'archives » distingue deux cas : l'indexation par les outils de recherche internes au service des archives et l'indexation par des moteurs de recherche externes. Dans le premier cas, le délai est échelonné entre 75 et 120 ans suivant leur type (cf figure 3). Dans le second cas :

« L'indexation par des moteurs de recherches externes sur les nom et prénom(s) des personnes concernées est interdite avant l'expiration d'un **délai de 120 ans**, à compter de la clôture des registres des actes en cause ou de la date des documents ».

Pour renforcer cette sécurité des données personnelles,

« des mesures de sécurité doivent être mises en place pour éviter tout téléchargement massif d'archives contenant des données à caractères personnel de personnes toujours vivantes ou dont la divulgation du contenu constituerait une atteinte à la vie privée de leurs ayant droit »

Enfin, le traitement des données doit recevoir l'autorisation de la CNIL. Dans le cas du site privé de généalogie *NotreFamille*, une attention a été portée aux données sensibles touchant à la religion et à la santé des personnes, si bien que ces données ont été systématiquement masquées.

Enjeux éthiques

Les limites éthiques

Ces questions de « données sensibles » relèvent également de l'éthique. En effet, il faut veiller à ce que les données produites par les usagers ne portent pas atteinte à l'intégrité des personnes, physiques ou morales. D'où l'importance de la modération des administrateurs, qui doivent veiller au respect des règles éthiques, mais aussi à la sensibilisation de l'utilisateur, à travers un pacte éthique énonçant des limites. Sur la plateforme *Correct* de la BnF, l'Utilisateur est soumis à un engagement :

« Ainsi l'Utilisateur s'engage notamment à ce qu'aucun Contenu ne contienne:
- D'éléments portant atteinte à des droits de la personnalité (par exemple des

¹⁴⁰ *op. cit.*

- représentations d'une personne sans autorisation consentie au titre de son droit à l'image) ;
- Des termes, propos et/ou des images susceptibles de porter atteinte à l'ordre public, au respect de la personne humaine ou de sa dignité, à la protection des enfants et adolescents (pornographie, pédophilie) ;
 - Des termes, propos et/ou des images encourageant, contenant ou provoquant à la discrimination, l'injure, la haine ou la violence à l'égard d'une personne ou d'un groupe de personnes, notamment en raison de leur origine, ou de leur appartenance ou non à une ethnie, une nation, une race, ou une religion déterminée, de leur handicap, de leurs préférences sexuelles ou de toute autre différence ;
 - Des termes, propos et/ou des images encourageant la commission de crimes et délits tels le commerce et la consommation de substances interdites, la prostitution, le suicide etc. ;
 - Des termes, propos et/ou des images faisant l'apologie ou la négation ou la remise en question des crimes de guerre et/ou contre l'humanité ;
 - Des termes, propos et/ou des images injurieux ou diffamants à l'encontre d'autrui ; ou portant atteinte à l'image ou à la réputation d'une marque ou d'une quelconque personne physique ou morale, de quelque manière que ce soit. »

Ces règles éthiques sont le fondement des bonnes pratiques du crowdsourcing. En effet, le non-respect de ces règles de base peut mettre en péril le projet.

Les cas de mésusages

Malgré l'établissement de règles, les projets de crowdsourcing peuvent être mis à mal par le parasitisme des internautes. Un exemple des dérives possibles peut être illustré par l'expérience de la LOC (Library of Congress) qui depuis janvier 2008 participe au projet de *Flickr the Commons*. Si de manière générale, cette initiative a été couronnée de succès, 1,1 million de vues dans les 24 heures après le lancement, l'institution a aussi « essayé les plâtres ». L'article « LOC, Flickr, and the Limits of Crowd Sourcing »¹⁴¹ datant de 2009, fait état des limites du projet. Il a été écrit par Larry Cebula, professeur d'histoire à l'Eastern Washington University.

Dans son article, l'auteur donne l'exemple d'une photographie qui a fait polémique, prise par le photographe David Bransby. Il s'agit d'une photographie couleur datant des années 40, montrant une jeune femme travaillant dans l'industrie de l'aviation, et visiblement en train de souder des circuits électriques. Cette photographie

¹⁴¹ « Northwest History: "Lick This": LOC, Flickr, and the Limits of Crowd Sourcing ». Consulté le 11 juin 2015. <http://northwesthistory.blogspot.fr/2009/06/lick-this-loc-flickr-and-limits-of.html>.

a donné lieu à de nombreuses réactions des internautes : une trentaine d'annotations, 71 tags et plus de 100 commentaires. Parmi les annotations ajoutées sur la photographie, la plupart sont des commentaires fantaisistes à la limite du politiquement correct. Une annotation indiquant le front de la jeune femme porte l'indication : « lick this ». En ce qui concerne les tags ajoutés, un certain nombre renseignent des données transparentes, comme la date de la prise de vue (1942). Un tag pertinent fait le parallèle avec « Rosie the Riveter » qui renvoie à une icône populaire de la culture américaine symbolisant les six millions de femme qui travaillaient dans l'industrie durant la Seconde Guerre Mondiale. D'autres tags sont plus énigmatiques, comme les initiales « sfv » ou encore « LF ». En ce qui concerne les commentaires, une minorité est consacrée à la composition de la photographie, tandis que la plupart commentent la beauté de la photographie ou bien de son sujet... Enfin on note que cette image a été associée à des albums (appelée « sets ») aux titres provocateurs tels que « Nation Of Domination (We Rule The Universe) ».

Dans le cas de cette photographie, l'intérêt des contenus générés par les utilisateurs est très limité. On constate que le phénomène de « bruit » -informations non pertinentes- empêche la lisibilité du document. Les commentaires apportent très peu de précisions historiques et les seules informations qui auraient pu être valables sont noyées sous une masse de données sans intérêt. Notons que ce parasitisme est probablement dû à la popularité de l'institution Library of Congress, puisque des projets de moindre envergure ne sont pas confrontés à ce travers, comme le confirme Patrick Pecatte, responsable du projet *Flickr PhotosNormandie* dans son blog¹⁴².

Le principe de pédagogie

Malgré toutes les règles juridiques et éthiques qui peuvent être établies, c'est le dialogue avec l'Utilisateur qui est primordial. En effet, le contributeur doit sentir un climat de confiance. La charte éthique doit donc s'exprimer selon des valeurs positives et non pas uniquement des restrictions. C'est de cette manière que *Wikimedia* exprime ses conditions d'utilisation :

¹⁴² *op. cit.*

« Vous êtes libre de :

- Lire nos articles et autres médias, gratuitement.
- Réutiliser nos articles et autres médias sous licences libres.
- Contribuer à et modifier nos différents sites et Projets sous licences libres.

Sous les conditions suivantes :

- Responsabilité
- Courtoisie
- Comportement régulier
- Pas de nuisance
- Conditions d'utilisation et règlement »

La liberté de l'utilisateur est ainsi mise en avant par rapports à ses devoirs. Tout l'enjeu est la transmission de « valeurs collaboratives » dont le projet est porteur.

Enjeux économiques

Au-delà de ces enjeux juridiques et éthiques, le crowdsourcing présente aussi des enjeux économiques, car les données produites par les usagers génèrent de la valeur.

Réutilisation commerciale des données publiques

Dans le cas où les données générées par les usagers ne sont pas protégées par le droit d'auteur, c'est l'institution qui décide de la réutilisation éventuelle de ces données. En tant que données publiques, elles peuvent faire l'objet d'une réutilisation commerciale par des acteurs privés. L'exemple des données généalogiques est souvent invoqué car certains sites proposent un accès payant à leurs bases de données.

Rappelons-nous en 2012 le procès attenté par le Conseil général du Cantal au site généalogique *Notreamille.com*¹⁴³. Le site privé avait demandé à faire une réutilisation commerciale des listes nominatives de recensement, demande rejetée par le Conseil

¹⁴³ « Réutilisation des données publiques : l'exception culturelle balayée en une phrase... |:: S.I.Lex »: Consulté le 11 juin 2015. <http://scinfolex.com/2012/07/07/reutilisation-des-donnees-publiques-lexception-culturelle-balayee-en-une-phrase/>.

Général pour deux raisons : premièrement, en vertu de l'article 11 de la loi 78-753 les archives disposeraient d'un pouvoir discrétionnaire leur donnant le droit de refuser cette demande. Deuxièmement, ce site ne garantirait pas le respect de la vie privée. Or durant le procès, le juge a rejeté ce premier argument de l'exception culturelle, considérant que les données publiques doivent être librement réutilisables. Le Conseil général avait alors fait appel, ce qui n'avait pas modifié le jugement mais apporté une condition : celle de respecter les dispositions de la loi CNIL. C'est sur ce point que le Conseil général a gagné le procès, mais pas l'affaire puisque le site *Notrefamille* a fini par obtenir ladite autorisation.

Avec l'indexation collaborative, le problème de la réutilisation est d'autant plus complexe qu'il s'agit du fruit d'un travail collectif. La solution proposée par Edouard Bouyé dans son rapport¹⁴⁴ est le moissonnage des bases de données généalogiques par le protocole OAI-PMH, ce qui permet à chaque service producteur de rester responsable de sa base, et ce, sans marchander leurs données.

Réutilisation commerciale des données personnelles

Au-delà du risque de marchandisation des données publiques produites par les usagers, les internautes sont aussi exposés à une réutilisation de leurs données personnelles.

Malgré l'engagement des sites, les données personnelles des internautes peuvent servir des intérêts commerciaux. Au moment de l'inscription sur le site des archives départementales de l'Ain, l'institution s'engage à la confidentialité de ces données :

« Les données personnelles collectées sont destinées uniquement à un usage interne aux Archives Départementales de l'Ain, et ne sont en aucun cas cédées ou vendues à des tiers. »

Or si le site ne divulgue pas directement ces données, l'utilisateur génère à son insu des métadonnées exploitables. En effet, l'internaute est à la fois « producteur de contenu » (*content producer*) en tant qu'il participe à la connaissance, et un fournisseur

¹⁴⁴ *op. cit.*

de données (*data provider*) en tant qu'il laisse de précieuses informations sur lui-même. Sa simple appartenance à une communauté en ligne fournit des éléments sur ces centres d'intérêt. Grâce à l'association de son adresse IP et des plateformes consultées, l'internaute fournit donc des métadonnées commercialement pertinentes. Le geste contributif n'est donc pas anodin puisqu'il permet de générer une valeur économique.

L'économie de la contribution

Le crowdsourcing participe donc d'une logique économique. C'est pourquoi le philosophe Bernard Stiegler¹⁴⁵ parle d'une « économie de la contribution » comme d'une troisième phase du développement du capitalisme. D'un point de vue sociohistorique, il identifie un premier modèle du capitalisme industriel, selon lequel les ouvriers représentent une force de travail. Au début du XX^{ème} siècle émerge le modèle de l'organisation tayloriste du travail, qui permet un basculement dans le consumérisme : les ouvriers ne sont plus seulement considérés comme de la main d'œuvre, mais aussi comme des consommateurs. Dans ce troisième modèle industriel du XXI^{ème} siècle, la dichotomie entre producteur et consommateur s'efface, laissant place à la figure hybride de l'utilisateur-contributeur. Bernard Stiegler définit le contributeur comme « un amateur au vieux sens du terme » : sa motivation s'avère économiquement désintéressée car il agit selon ses centres d'intérêt, ce qui permet de développer une expertise plus pointue. Dans ce modèle économique contemporain, le contributeur apporte plus que sa force de travail en mettant à contribution ses connaissances. Enfin cette économie est à la source de nouveaux modèles culturels, porteurs de « valeurs sociétales ». En somme, l'économie de la contribution est caractérisée par la motivation désintéressée des acteurs, orientée vers l'intérêt du plus grand nombre et la puissance de la multitude. Avec le crowdsourcing, nous rentrons donc dans une nouvelle ère industrielle, où le « flux informationnel » est au cœur des enjeux économiques.

¹⁴⁵ Proulx, Serge, José Luis Garcia, et Lorna Heaton. *La contribution en ligne : pratiques participatives à l'ère du capitalisme informationnel*. Presses de l'Université du Québec. Québec, 2014.

CONCLUSION

La notion d'archives participatives, fondée autour de pratiques participatives diversifiées introduit des enjeux juridiques, éthiques et économiques, avec entre autres, l'introduction d'un nouveau rapport au document d'archives.

Grâce aux usages du web social, on observe un renversement d'une logique verticale, de type *top-down* -c'est-à-dire du web vers l'utilisateur- à une logique horizontale de type *business model*, qui place l'utilisateur au centre du processus.

D'après des études récentes, il semblerait que nous tendions vers un renforcement du phénomène de contribution. En effet, jusqu'à présent, nous observions la règle de 90/9/1 : 1 internaute produit du contenu tandis que 9 l'enrichissent et 90 le consomment. Ces dernières années, la proportion serait davantage de l'ordre de 25/45/30, d'après les nouveaux chiffres présentés par Holly Goodier¹⁴⁶. Les trois-quarts des internautes seraient donc actifs que ce soit en diffusant du contenu ou bien en y contribuant, notamment par le biais du crowdsourcing.

Pour autant, va-t-on dans le sens d'une démocratisation culturelle, ou bien d'un renforcement des logiques industrielles et créatives ? Cette problématique fera l'objet d'une étude financée par l'ANR (Agence Nationale de la Recherche), dans le cadre du programme « Le numérique au service des arts, du patrimoine, des industries culturelles et éditoriales », qui donnera ses résultats en 2016¹⁴⁷.

¹⁴⁶ « BBC - The Participation Choice ». Consulté le 10 juin 2015.

http://www.bbc.co.uk/blogs/bbcinternet/2012/05/bbc_online_briefing_spring_201_1.html.

¹⁴⁷ « Projet COLLAB | ANR ». Consulté le 10 juin 2015. http://www.agence-nationale-recherche.fr/projet-anr/?tx_lwmsuivibilan_pi2%5BCODE%5D=ANR-14-CE24-0001.

Sources

Projets internationaux

<http://www.galaxyzoo.org/> Galaxyzoo, Zooniverse

<http://www.oldweather.org/> Old Weather, Zooniverse

<http://www.ancientlives.org/tutorial/transcribe> Ancient Lives, Zooniverse

<http://www.archives.gov/citizen-archivist/> Citizen Archivist, National Archives and Records Administration (Etats-Unis)

<http://menus.nypl.org/> What's on the Menu ? New York Public Library (Etats-Unis)

<http://buildinginspector.nypl.org/> Building Inspector, New York Public Library (Etats-Unis)

<http://www.transcribe-bentham.da.ulcc.ac.uk> Transcribe Bentham, University College of London (Royaume-Uni)

<http://www.whats-the-score.org> What's the score at the Bodleian ? Bodleian Library (Royaume-Uni)

<http://www.waisda.nl> Waisda ? Institut Néerlandais pour le Son et l'Image (Pays-Bas)

http://www.digitalkoot.fi/index_en.html Digitalkoot, Bibliothèque nationale de Finlande (Finlande)

<http://www.trove.nla.gov.au> Trove, National Library of Australia (Australie)

<https://www.destinationaustralia.gov.au> Destination : Australia, National Archives of Australia (Australie)

Projets européens

<http://monasterium.net/mom/home> Monasterium, ICARUS

<http://www.europeana1914-1918.eu> Europeana 1914-1918

Projets des archives départementales

http://www.archives-numerisees.ain.fr/archives/enquete/enquetes_en_cours/n:55 SOS Détective, Archives départementales de l'Ain (01)

http://fr.wikisource.org/wiki/Wikisource:Partenariats/Archives_D%C3%A9partementales_des_Alpes-Maritimes Transcription de manuscrits sur Wikisource, Archives départementales des Alpes-Maritimes (06)

<http://www.archives-aube.fr/jeux/cordeliere/mystere.php> Le mystère de la Cordelière, Archives départementales de l'Aube (10)

<http://www.cartotheque47.fr/> Cartothèque, Archives départementales du Lot-et-Garonne (47)

<http://laboratoire-archives.vendee.fr> Le L@boratoire des Internautes, Archives départementales de la Vendée (85)

<http://www.copainsdebanlieue.com/gene.html> Copains de banlieue, Archives Municipales de Saint-Denis (974)

Projets en marge des archives

<http://rosalis.bibliotheque.toulouse.fr/index.php?pages/rosalipédie#.VWX5c0Y2WtY> Rosalipédie, Bibliothèque numérique de Toulouse

<http://www.reseau-correct.fr/> Correction et enrichissement collaboratif de textes, Bibliothèque nationale de France

<http://www.memoiredeshommes.sga.defense.gouv.fr> Mémoire des Hommes, Ministère de la Défense

<http://museumgam.es/> Museum Metadatas Games, Mia Ridge

<http://www.geneanet.org/community/> Geneanet

Comptes sur les réseaux sociaux numériques

Twitter :

<https://twitter.com/1j1poilu> 1 Jour-1 Poilu @1J1Poilu

<https://twitter.com/1j1mam> DesPierresDesNoms @1J1MAM

Flickr :

<https://www.flickr.com/photos/photosnormandie/> PhotosNormandie

<https://www.flickr.com/groups/historyhappenshere/> History Happens Her

Bibliographie

Monographies

MATTHEWS, Jacob Thomas. *La culture par les foules ?* Paris, MkF, 2014, 63 p.

MOIREZ, Pauline. « Archives participatives », *Bibliothèques 2.0 à l'heure des médias sociaux*, dir. Muriel Amar et Véronique Mesguich, Paris, Editions du Cercle de la Librairie, 2012, p.187-197.

PROULX, Serge, GARCIA José Luis, HEATON Lorna. *La contribution en ligne : pratiques participatives à l'ère du capitalisme informationnel*. Québec, Presses de l'Université du Québec, 2014, 256 p.

RIDGE, Mia. *Crowdsourcing our cultural heritage*. Royaume-Uni, Ashgate, 2014, 283 p.

SUROWIECKI, James. *La sagesse des foules*. Paris, Lattès, 2008, 384 p.

Publications

BOUYE, Edouard. « Le web collaboratif dans les services d'archives publics : un pari sur l'intelligence et la motivation des publics », *La Gazette des Archives*, Association des Archivistes Français, n°227, 2012-13 [consulté le 24 novembre 2014].

Disponible sur le Web :

http://archives.cantal.fr/download.cgi?filename=accounts/mnesys_ad15/datas/cms/Bouye_RASAD_Gazette.pdf

BRABHAM, Daren C. « Moving the crowd at Threadless : motivations for participation in a crowdsourcing application », *Information, Communication & Society*, Vol. 13, n°8, 1122-1145, 2010

LE DEUFF, Olivier. « Folksonomies : les usagers indexent le web », *Bulletin des bibliothèques de France* [en ligne]. n°4, 2006 [consulté le 3 novembre 2014].

Disponible sur le Web :

<http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002>. ISSN 1292-8399

HOLLEY, Rose. « Crowdsourcing : How and Why Should Libraries Do it ? ». *D-Lib Magazine* [en ligne]. Vol. 16, n°3/4, 2010. [consulté le 30 octobre 2014].

Disponible sur le Web :

<<http://www.dlib.org/dlib/march10/holley/03holley.html>>

JOSSE, Isabelle. « La BnF engagée dans un projet de R&D pour la conception de la plateforme Correct, *Bulletin des bibliothèques de France* [en ligne], n°5, 2013 [consulté le 22 octobre 2014].

Disponible sur le Web :

<<http://bbf.enssib.fr/consulter/bbf-2013-05-0037-008>> ISSN 1292-8399

MOIREZ, Pauline. « Bibliothèques, crowdsourcing, métadonnées sociales », *Bulletin des bibliothèques de France* [en ligne]. n°5, 2013 [consulté le 22 octobre].

Disponible sur le Web :

<<http://bbf.enssib.fr/consulter/bbf-2013-05-0032-007>> ISSN 1292-8399

Rapports d'étude

ALAM, Sultana Lubna et CAMPBELL, John. *Crowdsourcing motivations in a not-for-profit GLAM context : The Australian Newspapers Digitisation Program*. 23rd Australasian Conference of Information Systems. Geelong, 3-5 décembre 2012 [consulté le 6 décembre 2014].

Disponible sur le Web :

<<http://dro.deakin.edu.au/eserv/DU:30049107/alam-crowdsourcingmotivations-2012.pdf>>

BOER (de), Victor. *Nichesourcing : Harnessing the Power of Crowds of Experts*. Conference on Knowledge Engineering and Knowledge Management, 2012.

CIOSI, Laure. *Politique des publics dans les services d'archives*. SIAF, Paris, 2013.

Disponible sur le Web :

<<http://www.archivesdefrance.culture.gouv.fr/static/7087>>

GUIGUENO, Brigitte et JONCHERY, Anne. *Enquête nationale sur les publics dans les services d'archives*. SIAF, Paris, 2014. Disponible sur le Web :

<<http://www.archivesdefrance.culture.gouv.fr/static/7741>>

ERTZSCHEID, Olivier. *Indexation sociale et folksonomies : le monde comme catalogue*. Journées Abes. 20 et 21 mai 2008 [consulté le 3 novembre 2014].

Disponible sur le Web :

<www.abes.fr/content/download/1139/5306/.../session1_OErtzscheid.pdf>

HOLLEY, Rose. *Many Hands Make Light Work : Public Collaborative OCR Text Correction in Australian Historic Newspapers*. National Library of Australia. Mars 2009. [consulté le 14 novembre 2014].

Disponible sur le Web :

<http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf>

HOLLEY, Rose. *Crowdsourcing and social engagement : potential, power and freedom for libraries and users*. PRDLA Annual Meeting and Conference, Nouvelle Zélande, 18-20 novembre 2009. [consulté le 3 novembre 2014].

Disponible sur le Web :

<<http://eprints.rclis.org/13969/>>

MOIREZ, Pauline, MOREUX, Jean-Philippe et JOSSE, Isabelle. *État de l'art en matière de crowdsourcing dans les bibliothèques numériques*, Bibliothèque nationale de France, 2013. [consulté le 22 octobre 2014].

Disponible sur le Web :

<http://www.bnf.fr/documents/crowdsourcing_rapport.pdf>

OOMEN, Johan et AROYO Lora. *Crowdsourcing in the Cultural Heritage Domain : Opportunities and Challenges*. 5th International Conference on Communities and Technologies, New York, 2011. [consulté le 18 décembre 2014].

Disponible sur le Web :

<http://www.iisi.de/fileadmin/IISI/upload/2011/p138_oomen.pdf>

SMITH-YOSHIMURA, Karen et SHEIN, Cyndi. *Social Metadata for Libraries, Archives, and Museums*. OCLC. 2011- 2012. [consulté le 22 octobre 2014].

Disponible sur le Web :

<<http://www.oclc.org/research/activities/agggregating.html>>

STORUP, Bérangère. *La recherche participative comme mode de production de savoirs*. FSC. 2012-2013.

<http://sciencescitoyennes.org/wp-content/uploads/2013/04/FSC_final_recherche-participative_FdF.pdf>

Billets de blog

BRUMFIELD, Ben. « Control for Crowdsourced Transcription ». *Collaborative Manuscript Transcription*. Billet publié le 5 mars 2012 [consulté le 24 novembre 2014].

Disponible sur le Web :

<<http://manuscripttranscription.blogspot.fr/2012/03/quality-control-for-crowdsourced.html>>

CHIGNARD, Simon. « Mais pourquoi contribue-t-on ? » *Données ouvertes*. Billet publié le 22 novembre 2012 [consulté le 28 octobre 2014].

Disponible sur le Web :

<<http://donneesouvertes.info/2012/11/22/mais-pourquoi-contribue-t-on/>>

HOLLEY, Rose. « Crowdsourcing and social engagement in libraries : the state of play ». *Alia Sydney*. Billet publié le 29 Juin 2011. [consulté le 30 octobre 2014].

Disponible sur le Web :

<<http://aliasydney.blogspot.fr/2011/06/crowdsourcing-and-social-engagement-in.html>>

OWENS, Trevor. « Human Computation and Wisdom of Crowds in Cultural Heritage », *Trevor Owens, User Centered Digital History*. Billet publié le 2 juin 2012. [consulté le 22 octobre 2014].

Disponible sur le Web :

<<http://www.trevorowens.org/2012/06/human-computation-and-wisdom-of-crowds-in-cultural-heritage/>>

THEIMER, Kate. « Exploring the participatory archives ». *Archives Next*. Billet publié le 30 août 2011. [consulté le 10 octobre 2014].

Disponible sur le Web :

<<http://www.archivesnext.com/?p=2319>>

Table des annexes

ANNEXE 1 : QUESTIONNAIRE AUX INSTITUTIONS	103
ANNEXE 2 : EXEMPLE D'UNE COLLECTE D'ARCHIVES PRIVEEES : LA CARTOTHEQUE DU LOT-ET-GARONNE	104
ANNEXE 3 : EXEMPLE D'UNE CORRECTION D'OCR : CORRECT	104
ANNEXE 4 : EXEMPLE D'UNE INDEXATION COLLABORATIVE : MEMOIRE DES HOMMES	105
ANNEXE 5 : EXEMPLE D'IDENTIFICATION DE PHOTOGRAPHIE : SOS DETECTIVES	105

Questionnaire sur les activités de crowdsourcing appliquées aux documents d'archives

Objectif : mieux cerner les motivations des services d'archives à mettre en place des activités collaboratives

***Obligatoire**

Nom de la structure :

Nom du projet :

Date de début du projet :

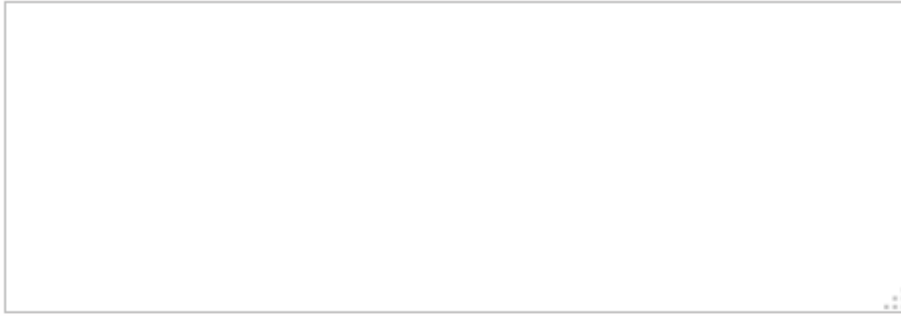
Quel(s) type(s) de pratiques collaboratives avez-vous mis en place ? *

- indexation collaborative
- transcription/correction d'OCR
- écriture collaborative
- collecte de documents
- co-curation
- Autre :

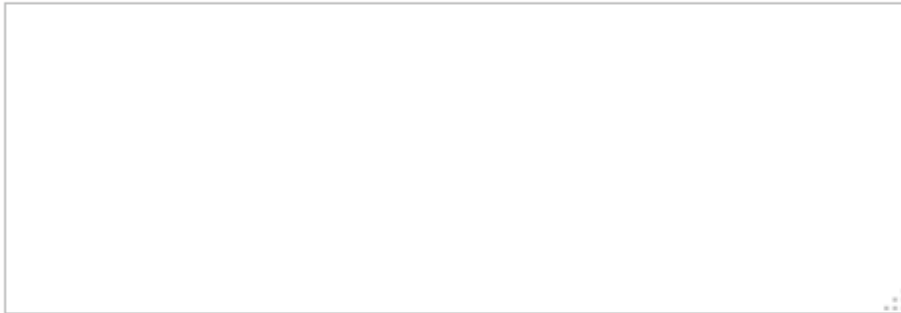
Quelle est la nature de ces documents d'archives ? *

- photographies
- cartes postales
- registres d'état civil
- manuscrits
- Autre :

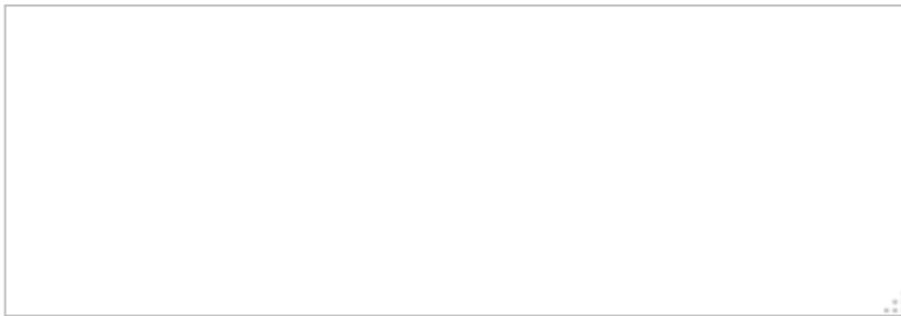
Dans quel cadre ce projet a-t-il été mis en place ? (appel à projet, financements)



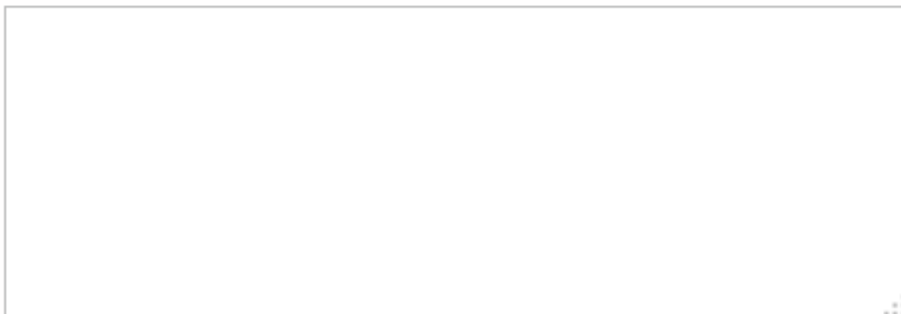
Pourquoi avoir voulu impliquer les internautes ?



De quelle manière avez-vous communiqué sur le projet ?



Quel est le public visé par le projet ?



De quelle manière les internautes sont-ils encadrés ?

- création d'un compte
- test de paléographie
- validation par l'administrateur
- Autre :

Les contributeurs les plus actifs sont-ils encouragés ?

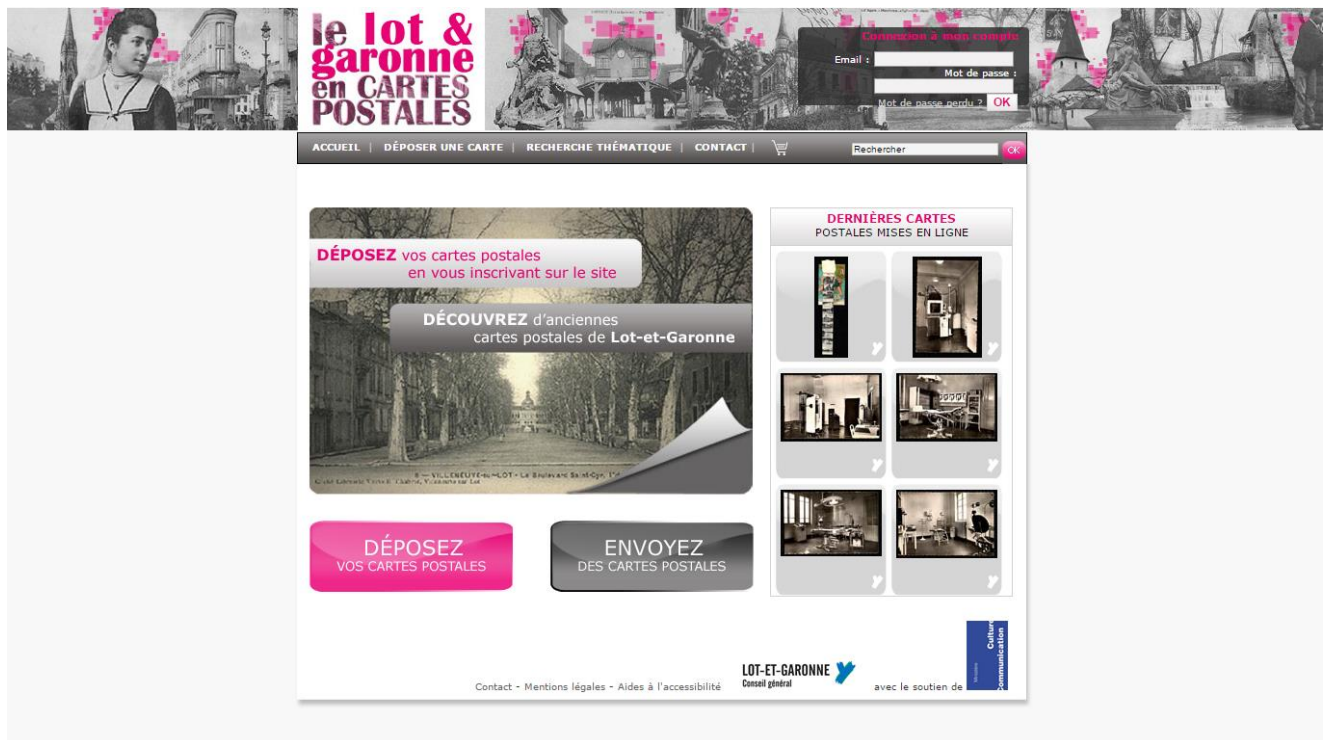
- pas de système de gratification
- top des contributeurs
- obtention de badges
- attribution de points
- Autre :

Ce travail collaboratif fait l'objet d'une valorisation patrimoniale ?

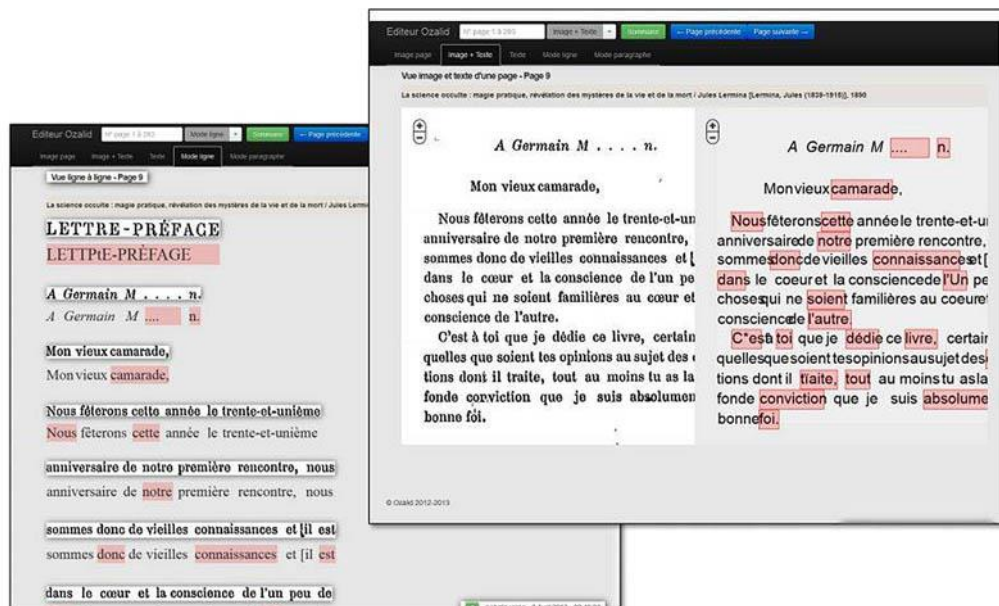
Combien de professionnels ce projet implique-t-il ? (temps de travail, tâches à réaliser)

Dans l'ensemble, le bilan est-t-il positif ? Donner si vous en avez des chiffres, statistiques.

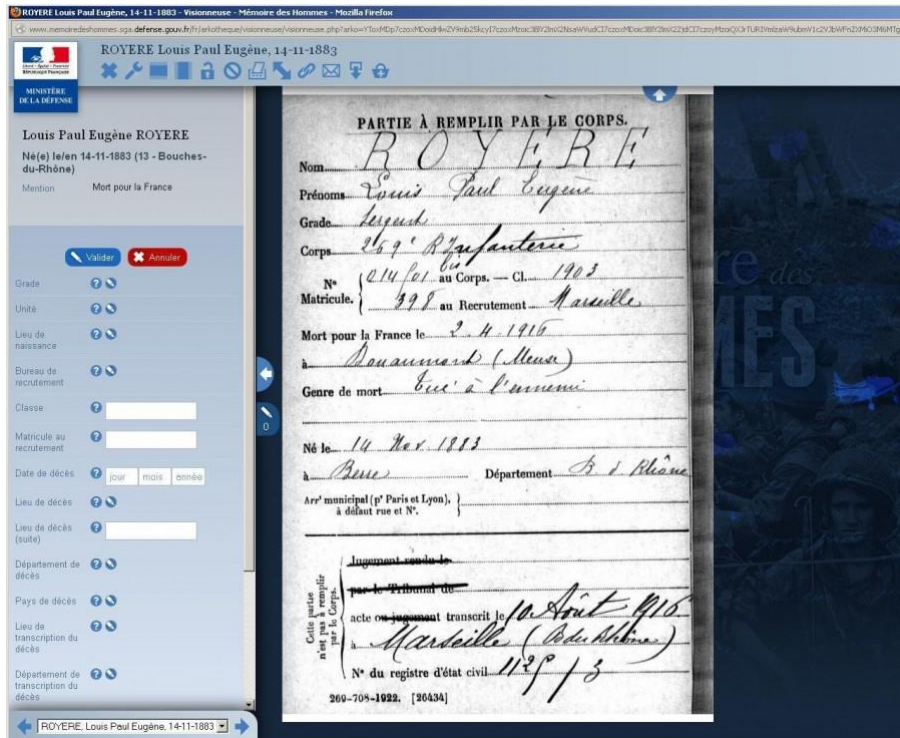
ANNEXE 1 : QUESTIONNAIRE AUX INSTITUTIONS



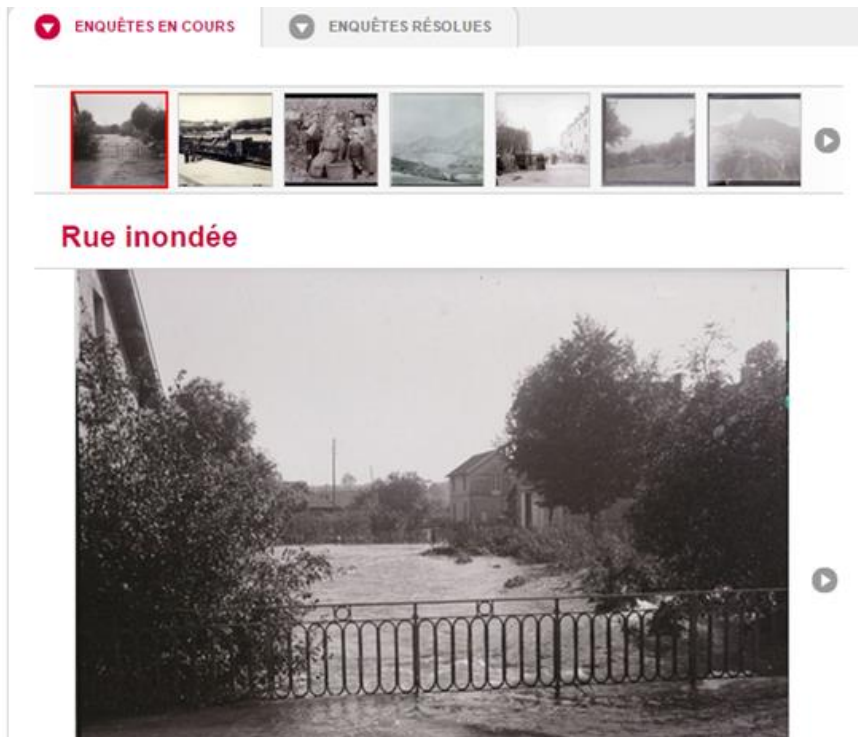
ANNEXE 2 : EXEMPLE D'UNE COLLECTE D'ARCHIVES PRIVÉES : LA CARTOTHEQUE DU LOT-ET-GARONNE



ANNEXE 3 : EXEMPLE D'UNE CORRECTION D'OCR : CORRECT



ANNEXE 4 : EXEMPLE D'UNE INDEXATION COLLABORATIVE : MEMOIRE DES HOMMES



ANNEXE 5 : EXEMPLE D'IDENTIFICATION DE PHOTOGRAPHIE : SOS DETECTIVES

Liste des entretiens

Entretiens téléphoniques

Réalisés le 06/03/2015

Pouyadou, Véronique, Archives départementales de Lot-et-Garonne

Vesson, Valéry, Archives départementales de l'Ain

Réponses au questionnaire

Créé le 06/03/2015

Aufroy, Sandrine, Ministère de la Défense
(09/03/2015)

Cavalié, Hélène, Archives départementales des Alpes-Maritimes *(10/03/2015)*

Josse, Isabelle, Bibliothèque nationale de France
(09/03/2015)

Le L@boratoire des InternauteS, Archives départementales de la Vendée *(18/03/2015)*

Table des illustrations

Figure 1 : Comparatif entre les concepts Human Computation et Wisdom of Crowds par Trevor Owens	14
Figure 2 Cartographie des instruments de recherche en ligne dans les services d'archives publics, réalisée par les Archives de France (avril 2015)	25
Figure 3 : Les délais de diffusion et d'indexation applicables à certains documents mis en ligne par les services publics d'archives.....	27
Figure 4 Cartographie de la numérisation et mise en ligne des registres paroissiaux et état civil dans les services d'archives publics, réalisée par les Archives de France (avril 2015).....	29
Figure 5 : Archives départementales pratiquant l'indexation collaborative début 2015	35
Figure 6 : Fonds d'archives faisant l'objet d'indexation collaborative en janvier 2012 (diagramme réalisé par Edouard Bouyé).....	36
Figure 7 : Systèmes d'indexation collaborative utilisés par les services d'archives en janvier 2012 (diagramme réalisé par Edouard Bouyé)	54
Figure 8 : La checklist de Rose Holley pour un projet réussi	79

Table des matières

SIGLES ET ABREVIATIONS	7
INTRODUCTION.....	8
NOTION D'ARCHIVES PARTICIPATIVES	10
Elements de terminologie	10
<i>Human Computation vs Wisdom of Crowds</i>	<i>10</i>
<i>Web collaboratif vs web participatif.....</i>	<i>15</i>
<i>Du crowdsourcing au nichesourcing</i>	<i>18</i>
Comprendre la démarche des archives	21
<i>Politiques des services d'archives</i>	<i>21</i>
<i>Spécificités des fonds d'archives</i>	<i>27</i>
<i>Valeurs des documents d'archives</i>	<i>31</i>
Un intérêt croissant pour les métadonnées sociales	33
<i>Etat des lieux en France.....</i>	<i>33</i>
<i>Publics participatifs.....</i>	<i>38</i>
<i>Forces et faiblesses de l'indexation sociale</i>	<i>41</i>
TYPOLOGIE DES PRATIQUES PARTICIPATIVES	44
Les activités	44
<i>Correction & transcription</i>	<i>44</i>
<i>Contextualisation.....</i>	<i>47</i>
<i>Enrichissement</i>	<i>49</i>
<i>Classification.....</i>	<i>51</i>
Les plateformes.....	53
<i>Les solutions logicielles</i>	<i>54</i>
<i>Les sites de partage de contenu.....</i>	<i>57</i>
<i>Les initiatives en marge des GLAM</i>	<i>60</i>
Les stratégies	63
<i>Des événements ponctuels</i>	<i>63</i>
<i>Une expérience ludique.....</i>	<i>66</i>
<i>Une time-machine.....</i>	<i>69</i>
LES MOTEURS DE LA PARTICIPATION	72
Du côté des publics, les facteurs de motivation	72
<i>Motivations intrinsèques</i>	<i>72</i>
<i>Motivations extrinsèques.....</i>	<i>74</i>
Du côté des institutions, les facteurs de réussite	76
<i>Ingrédients d'un site réussi</i>	<i>76</i>

<i>Recommandations</i>	78
<i>Retours d'expérience</i>	81
Les enjeux du crowdsourcing	84
<i>Enjeux juridiques</i>	84
<i>Enjeux éthiques</i>	87
<i>Enjeux économiques</i>	90
CONCLUSION	93
SOURCES	94
BIBLIOGRAPHIE	96
TABLE DES ANNEXES	100
LISTE DES ENTRETIENS	106
TABLE DES ILLUSTRATIONS	107
TABLE DES MATIERES	108