

UNIVERSITÉ LIBRE DE BRUXELLES
Faculté de Philosophie et Lettres

Collections iconographiques numérisées et *crowdsourcing* :

Possibilités et limites de la co-crédation de mdtadonnées
par le grand public au travers de trois études de cas

Anne CHARDONNENS

Mémoire présenté sous la direction de
Seth VAN HOOLAND en vue de l'obtention du
diplôme de Master en Gestion Culturelle

Année académique 2014–2015

Résumé

Avec rigueur et patience, les musées, archives et bibliothèques s'emploient à décrire d'immenses volumes de documents iconographiques afin qu'ils puissent être trouvés par les utilisateurs. Le succès de leur entreprise est cependant limité par un manque de temps, de ressources humaines et de moyens financiers. Grâce aux outils numériques, les institutions culturelles peuvent désormais intégrer le grand public dans ce processus de création de métadonnées. Par le biais de dispositifs de *crowdsourcing*, elles invitent les utilisateurs à commenter, identifier et décrire leurs collections iconographiques. Ces nouvelles pratiques collaboratives soulèvent cependant des questions : quelle est la nature, la qualité et la pertinence de ces contributions ?

Ce mémoire porte sur les commentaires produits par le grand public dans ce contexte. Après avoir exposé les enjeux propres au *crowdsourcing* et dressé un état de l'art dans la première partie, ce mémoire propose une approche empirique basée sur trois études de cas. Une analyse à la fois quantitative et qualitative de près de 900 commentaires permet d'observer l'influence que la grandeur du projet exerce sur les contenus produits. L'interprétation des résultats conduit finalement à la formulation de pistes de réflexion et de recommandations à destination des institutions culturelles.

Mots-clés : *crowdsourcing*, co-création ; co-construction ; métadonnée ; commentaire ; contributeur ; patrimoine ; collection iconographique ; intelligence collective ; numérisation ; étude de cas ; typologie.

Remerciements

Au terme d'une riche année d'apprentissage, il me semble important de « conclure » ce mémoire par quelques sincères remerciements :

À Seth van Hooland, qui a accepté de diriger ce mémoire, qui m'a donné l'envie d'élargir le champs de mes connaissances et qui a encadré ma recherche avec la plus grande pédagogie ;

À Nathalie Nyst, qui a accepté d'être ma lectrice ;

À Simon Hengchen, qui a été d'une grande disponibilité et dont les conseils m'ont permis de relever nombre de défis techniques ;

À Ettore Rizza, qui a répondu positivement à ma demande d'aide et s'est chargé de l'extraction de certaines données ;

À Patrick Peccatte, qui a eu l'amabilité de partager son expérience et expertise ;

À Pauline Moirez, qui a inspiré le choix de l'une de mes trois études de cas ;

À Max de Wilde, qui m'a permis de régler quelques questions liées à \LaTeX ;

À Salomé Frémineur, Déborah Schwarzbaum et Michelle Chardonnens, qui ont pris le temps de relire ces pages ;

À tout mon entourage ;

Et enfin, à Evan Fay Earl, que je ne connais pas, mais dont la recherche sur le *crowdsourcing* a inspiré ce mémoire, à une époque où j'ignorais encore ce qu'était une métadonnée.

Table des matières

Remerciements	1
INTRODUCTION	3
I Approche théorique	6
1 Contexte	7
1.1 Brève mise en perspective historique	7
1.2 Du web social à l' <i>Open Data</i>	8
1.3 Intelligence collective et culture participative	12
2 Enjeux	15
2.1 Numérisation et mise en ligne des collections	15
2.2 <i>Crowdsourcing</i>	18
3 État de l'art	24
3.1 Définitions et typologies	24
3.2 Motivations	30
3.3 Types de dispositifs et de plateformes	32
3.4 Initiatives non conventionnelles	35
3.5 Analyse de tags et commentaires	37
II Études de cas	42
4 Trois études de cas	43
4.1 section.4.1	
4.2 <i>The Powerhouse Museum</i> (cas B)	45
4.3 <i>The English Heritage</i> (cas C)	47
4.4 Des approches différentes	48

5	Méthodologie	56
5.1	Objet d'étude	56
5.2	Collecte des données	56
5.3	Constitution des échantillons représentatifs	59
5.4	Traitement des données	60
6	Analyse des données	61
6.1	Comparaison de données quantitatives	61
6.2	Outil d'analyse qualitative	64
6.3	Remarques et observations	72
6.4	Interprétation des résultats	76
7	Implications pour l'institution	88
7.1	Corpus, plateforme et calendrier	88
7.2	Profils des contributeurs	89
7.3	Ressources humaines	90
7.4	Modération et animation	92
7.5	Destin des métadonnées	93
	CONCLUSION	95
	BIBLIOGRAPHIE	106
	ANNEXES	107
A	Exemples de commentaires d'utilisateurs	107
B	Grille d'analyse des commentaires	112

INTRODUCTION

*Crowdsourcing is a story of cooperation, aggregation,
teamwork, consensus, and creativity.*

(Brabham, 2013)

Depuis quelques années, la mise en ligne des collections iconographiques¹ que conservent les institutions culturelles (musées, archives et bibliothèques) donne lieu à des pratiques inédites et voit éclore de nouvelles formes d'interaction avec les utilisateurs. Parmi celles-ci, le *crowdsourcing*. Ce terme, introduit par le magazine *Wired* en juin 2006, désigne un système de co-construction reposant sur la participation de la « foule ». Applicable à de nombreux secteurs, il permet notamment à n'importe quel individu de procéder à l'enrichissement des métadonnées culturelles, pour peu qu'il puisse naviguer sur le web. Différentes plateformes en ligne deviennent ainsi le « terrain de jeu » de citoyens lambda désireux de contribuer à la documentation de ces collections. Ils peuvent, selon les cas, relire, corriger, classer, taguer, décrire ou encore géolocaliser des documents. L'implication peut être de quelques secondes ou se répéter jusqu'à devenir un engagement de plus longue haleine. Cette dynamique collaborative permet de documenter d'importants volumes de données à moindres coûts.

Ce mémoire mettra l'accent sur les commentaires qui permettent aux utilisateurs d'enrichir les métadonnées relatives à des collections iconographiques. Il peut s'agir par exemple de commentaires visant à identifier un moulin vendéen représenté sur une carte postale numérisée. Dans un commentaire posté le 19 août 2012, un utilisateur fait part des informations qu'il possède :

Bonjour, Vous avez tout à fait raison, il s'agit bien du moulin de la Tranche.
De très nombreuses cartes postales le montrant sous toutes ses faces sont actuellement visibles sur le site <http://cartes-postales.delcampe.fr> . Un meilleur tirage de la photo ci-dessus, avec la même référence, précise qu'il est situé à La Tranche sur Mer. On le trouve aussi dénommé Moulin Vieux, Le vieux moulin

1. Les documents iconographiques englobent (entre autres) : fonds ou albums photographiques, cartes postales, dessins, peintures, estampes, affiches, polaroids, plaques photographiques en verre, négatifs, héliogravures, *etc.* (Claerr et Westeel, 2011).

du tènement du bourg... Une carte précise même que ce moulin a été bombardé par les anglais en 1816. Cordialement *Jean-Pierre Bouchet* [E1]²

Si de nombreux commentaires possèdent comme celui-ci une réelle plus-value informationnelle, en revanche d'autres sont constitués d'éléments strictement subjectifs, comme en témoigne par exemple cette contribution très enthousiaste laissée par *Nooura* : « Great Great Great ... shot ... LOVE it » [E2].

Face à ce récent phénomène, il semble primordial de prendre un peu de recul et d'en observer les possibilités et limites. Les institutions peuvent-elles envisager le *crowdsourcing* comme une nouvelle forme de collaboration à long terme ou faut-il plutôt y voir un support à des actions ponctuelles de médiation ou de communication ?

En outre, quelle est la nature des commentaires postés par le commun des utilisateurs ? Peut-on parler de véritable expertise ou s'agit-il davantage d'épanchements nostalgiques face à de vieux clichés ? De plus, d'autres questions annexes émergent : quelles sont les dynamiques qui sous-tendent les interactions entre utilisateurs ? Faut-il animer, modérer, surveiller les échanges ? Les spammeurs sont-ils légion ?

Ce mémoire vise à examiner un phénomène qui est à la jonction entre des institutions conservant des collections iconographiques et les nouvelles potentialités qu'offre le web. À une époque marquée par des mesures d'austérité touchant de plein fouet le secteur culturel, il semble pertinent de s'intéresser aux dynamiques à l'œuvre en lisière des pratiques traditionnelles. En effet, s'agissant d'institutions dont les moyens humains et financiers sont limités, le *crowdsourcing* se présente comme une alternative peu gourmande en ressources humaines et ne nécessitant pas de lourds investissements financiers³. Ce travail a donc pour but de poser un regard critique sur les avantages, mais également les difficultés que représente la co-création de métadonnées par le grand public.

Afin d'avancer des éléments de réponse, il semble important de s'intéresser dans un premier temps au contexte, d'identifier les enjeux et de réaliser un état de l'art permettant de mieux appréhender cette forme inédite d'interaction. Il sera ensuite possible dans une seconde partie d'observer plus en détail la nature des commentaires postés par les utilisateurs. L'étude de trois cas contrastés permettra en effet d'explorer les pratiques des utilisateurs et la façon dont ils s'approprient les dispositifs de *crowdsourcing*. L'étude, basée sur une méthodologie à la fois quantitative et qualitative, permettra d'obtenir de façon empirique des éléments de réponse, d'en tirer des conclusions transposables à d'autres institutions et d'énoncer des recommandations à destination du secteur culturel.

2. Les commentaires cités en exemple, identifiables par un numéro de référence, sont repris en annexe, accompagnés de l'URL correspondante, voir *Annexe A, Figure A.1*.

3. Par exemple, en optant pour une plateforme préexistante comme Flickr, un tel projet peut être lancé très facilement à moindre coût.

Première partie
Approche théorique

Chapitre 1

Contexte

1.1 Brève mise en perspective historique

Le *crowdsourcing* n'est-il qu'un effet de mode ?

Si ce néologisme¹ apparu en 2006 définit une pratique intrinsèquement liée au web, il est toutefois intéressant de prendre du recul et de mettre en perspective ce phénomène. En effet, à plusieurs reprises, depuis le Moyen Âge déjà, le grand public a été sollicité pour prendre part à une création ou construction collective, chaque individu pouvant potentiellement amener sa contribution à un tout dépassant son intérêt personnel. S'intéressant aux folksonomies, Boydens (2010) remarque que ces « classifications représentatives du web 2.0 reposant sous le seul contrôle de l'individu qui les crée » peuvent être associées à certaines pratiques médiévales. Elle fait notamment référence aux générations de moines copistes qui, lorsqu'ils recopiaient des manuscrits, corrigeaient parfois certains éléments, sans quoi ils étaient accusés d'incompétence et de paresse. Il leur arrivait également d'utiliser les marges pour y exprimer des sentiments anticonformistes ou encore pour y dévoiler des blagues paillardes, de la mauvaise poésie ou des images obscènes².

Un autre exemple, légèrement plus récent, concerne l'édition de plans de villes par Braun et Hogenberg. Au tournant du XVII^e siècle, ils publient à Cologne leur collection de gravures représentant des plans et vues de villes. Plus d'une centaine d'artistes et cartographes ont contribué à ce *Civitates Orbis Terrarum* et Füssel (2008) explique, en

1. En français : « externalisation ouverte » (terme introduit par l'Office québécois de la langue française) ou encore « production participative » (expression proposée par la Commission générale de terminologie et de néologie de France), voir http://gdt.oqlf.gouv.qc.ca/fiche0qlf.aspx?Id_Fiche=45436 (consulté le 10.03.15) et <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000029331922> (consulté le 10.03.15))

2. Cet irrespect du texte concerne surtout la littérature profane : « alors que l'on tente de préserver l'intégrité des textes bibliques et des auteurs faisant autorité, les documents considérés comme mineurs pour la société d'alors sont sans cesse affectés par une intervention individualiste qui en fragilise le contenu initial. » (Boydens, 2010)

préface de la réédition des 363 planches couleur, que Georg Braun était encore « disposé à accueillir d'autres villes dans son ouvrage, priant pour ainsi dire ses lecteurs d'agir de manière "interactive" et de lui transmettre des propositions de ville qu'ils souhaitent trouver dans l'atlas »³.

Le phénomène du *crowdsourcing* peut encore être rapproché d'autres pratiques, par exemple la rédaction collective de cahiers de doléances par le tiers état dans la France de l'Ancien Régime, les travaux de généalogie entrepris par les Mormons (utilisation de microfilms pour sauvegarder et exploiter les registres et archives de toute la planète) ou encore la rédaction collaborative de l'Encyclopédie de Diderot et D'Alembert.

Cependant, ce type de rapprochement ne fait pas l'unanimité. Brabham (2013) explique à travers deux exemples qu'il ne s'agit pas véritablement de *crowdsourcing* tel qu'on le comprend aujourd'hui. Il cite la création de l'*Oxford English Dictionary*⁴, ainsi que l'*Alkali Prize*⁵ et met en exergue la différence cruciale qui existe entre ces exemples d'antan et le phénomène plus récent du *crowdsourcing* : Internet. Selon lui, les exemples comportent bien une dimension de production collaborative et de résolution collective de problèmes, mais ne peuvent être comparés aux configurations permises par Internet et les nouveaux médias. Par sa vitesse, sa portée et ses riches fonctionnalités, Internet permet d'abaisser bien des obstacles et rend ainsi possible des expériences de *crowdsourcing* qualitativement différentes des processus d'antan.

Pour Brabham (2013), le web serait donc l'élément-clé, qui, par ses caractéristiques et les possibilités qu'il offre, permet des formes de collaboration tout à fait inédites. La différence établie sur la variable « web » a toutefois ses limites et peut être nuancée : en effet, il n'est pas rare de rencontrer des projets de *crowdsourcing* qui mêlent dans leur stratégie des actions en ligne et *in situ*⁶.

1.2 Du web social à l'*Open Data*

Afin de pouvoir appréhender le phénomène du *crowdsourcing*, il est primordial de s'intéresser dans un premier temps au contexte dans lequel il a pris racine, à savoir, celui du web. L'évolution du web au seuil du XXI^e siècle s'est caractérisée par une dimension sociale, interactive et participative de plus en plus prégnante. Ce web social

3. Braun annonce d'ailleurs en 1576 dans le second volume en langue allemande : « Mais je désirerais inviter aimablement celui qui ne trouve pas la ville de son père ou sa ville natale dans ces deux premiers livres, à la dessiner telle qu'elle est et à me l'envoyer, [...] » (Füssel, 2008)

4. À la fin du XVIII^e siècle, la population était invitée par le biais d'un appel ouvert, à soumettre des mots anglais ainsi que leurs usages pour qu'ils soient indexés dans le dictionnaire.

5. Dans les années 1700, LouisXVI mit sur pied une opération de grande envergure : chacun pouvait proposer sa solution pour optimiser la production de l'Alkali, le vainqueur serait récompensé par le roi.

6. Voir 3.3 *Types de dispositifs et de plateformes*

et participatif, plus connu sous l'appellation « web 2.0 », repose sur des principes tels que le partage, la collaboration et la génération de contenu par les utilisateurs.

La paternité du terme « web 2.0 » est attribuée à Tim O'Reilly. Il en a formalisé la définition un peu plus d'un an après sa première occurrence⁷(O'Reilly, 2005b). Quelques jours après la diffusion de cette définition, il en a proposé une version plus condensée :

Web 2.0 is the network as platform, spanning all connected devices ; web 2.0 applications are those that make the most of the intrinsic advantages of that platform : delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an « architecture of participation », and going beyond the page metaphor of web 1.0 to deliver rich user experiences. (O'Reilly, 2005a)

Cette définition pose les jalons de ce qui caractérise le web 2.0 (décentralisation, amélioration perpétuelle des logiciels, *etc.*), tout en mettant l'accent sur l'évolution du statut des utilisateurs individuels ainsi que sur la notion de « participatif ».

Ces notions propres au web 2.0 ont pris corps autour d'outils et d'applications qui favorisent l'interaction : réseaux sociaux, plateformes de partage de contenu multimédia, blogs, wikis et autres outils de partage de liens et ressources (Pirolli et Créatin-Pirolli, 2011). Bien entendu, des outils d'interaction existaient déjà avant l'émergence du web 2.0, cependant, comme l'explique Rieder (2010), « les applications que l'on regroupe habituellement derrière le terme “web social” se distinguent de leurs prédécesseurs parce qu'elles vont plus loin dans l'organisation et la structuration des interactions entre usagers. »

Cette dynamique nouvelle, favorisée par la gratuité et la facilité des services (Proulx, 2009), conduit à une modification du statut des utilisateurs. Les internautes s'émancipent, leurs pratiques se diversifient (documentation, production, réalisation, publication, *etc*) et les frontières entre les statuts d'auteur, de lecteur et de spectateur sont de moins en moins nettes (Pirolli et Créatin-Pirolli, 2011). Rapidement, les publications se multiplient et des néologismes voient le jour pour désigner ce nouveau visage de l'utilisateur qui échange, propose, commente, s'implique, participe, produit du contenu et le fait circuler ; Pirolli et Créatin-Pirolli (2011) parlent ainsi de « webacteurs ». Ils relèvent d'ailleurs que « l'une des particularités remarquables de cette évolution tient au fait que ces pratiques ne sont à aucun moment imposées aux utilisateurs, mais qu'elles sont induites, directement ou indirectement, par l'utilisation des outils. »

7. « The concept of “web 2.0” began with a conference brainstorming between O'Reilly and MediaLive International », <http://www.oreilly.com/pub/a//web2/archive/what-is-web-20.html> (consulté le 12.03.15)

Pour sa part, Flichy (2010) souligne le rôle que joue Internet dans l'émergence de la figure de l'amateur. Il observe que le web permet à celui-ci d'acquérir de nouvelles compétences et de les mettre en œuvre de diverses manières, ce que ne permettaient pas les médias qui se sont développés au XX^e siècle. Selon lui, cette montée en puissance des amateurs n'est pas un simple effet de mode : elle signifie une nouvelle ère de démocratisation, celle des compétences (Flichy, 2010). Si les projets de *crowdsourcing* vont bénéficier de ce terreau fertile, ils vont également offrir à ces amateurs un nouvel espace d'expérimentation, tout en leur permettant d'obtenir une certaine reconnaissance institutionnelle.

Une autre apparition que l'on doit au Web social, c'est la classification de masse (Rieder, 2010). Grâce à de nouveaux outils disponibles en ligne, les individus peuvent organiser, archiver, gérer et partager tous les types de contenus qu'ils souhaitent retrouver par la suite (Gordon-Murnane, 2006) ; on parle de « folksonomies », de *tagging* et de *social bookmarking*. Ces termes voisins désignent tous le fait qu'en taguant un contenu, l'utilisateur participe à l'indexation des contenus mis en ligne, c'est-à-dire à l'indexation d'une (infime) partie du Web. Si de telles pratiques suscitent un vif enthousiasme, elles souffrent cependant de différents problèmes : l'absence d'une base de mots-clés standards, le manque de précision, l'absence de structure normalisée (singulier ou pluriel, *etc.*), les fautes d'orthographe, le chevauchement de différentes langues, la perte de toute information contextuelle, l'ambiguïté et la confusion dues à l'homonymie, à la synonymie ou à l'usage d'acronymes, l'absence de relations hiérarchiques et la « rivalité » avec des systèmes de métadonnées préexistants (Gordon-Murnane, 2006)). Les projets de *crowdsourcing* invitant les utilisateurs à taguer les contenus souffriront de ces mêmes limites. De plus, Boydens (2010) fait remarquer qu'à l'instar d'autres applications et données issues du web 2.0, les tags sont dit « situationnels », c'est-à-dire qu'ils peuvent rapidement devenir désuets ou renvoyer vers des « liens morts ».

Selon Noiret (2012), ce qui fait la spécificité du web 2.0, au-delà de ses aspects techniques, ce sont les « espaces ouverts qui permettent des formes de participation collective ». Les nouvelles architectures du Web entraînent des effets de réseau entre les internautes connectés les uns aux autres, de telle sorte que « les contributions s'ajoutent, au point de créer un ensemble qui est plus grand que la somme de ses parties » (Pisani et Piotet, 2011). Il y a donc, d'une part, un phénomène d'accumulation et de superpositions et, d'autre part, des processus collaboratifs qui émergent.

Le *crowdsourcing* est un concept qui s'inscrit dans un environnement en constant développement : le web. Ainsi, la notion phare des années 2000, le web 2.0, va être progressivement éclipsée par de nouvelles mouvances. Parmi celles-ci, le mouvement *Open Data* : « Open data is data that can be accessed, distributed, and reused by everyone, even for commercial purposes, without the need to explicitly ask the data

owner for permission » (Baltussen *et al.*, 2013). Sans cette ouverture des données, le recours au *crowdsourcing* serait en réalité difficilement concevable. En effet, pour que l'utilisateur ait quelque chose à décrire, classer ou commenter, encore faut-il que le document soit aisément consultable.

Le mouvement *Open Data* représente donc une approche tout à fait nouvelle des données, en rupture avec la politique traditionnellement menée par les institutions culturelles (accès restreint aux réserves, coût des entrées, interdiction de prendre des photographies des œuvres, *etc.*). Désormais, de plus en plus d'institutions culturelles décident de rendre accessibles, en ligne, leurs collections numérisées. Décuplant les possibilités de sauvegarde et de diffusion des données, le web constitue ainsi un outil puissant pour ces établissements. Si l'institution adhère au mouvement *Open Data*, la mise en ligne des collections représente alors un pas de plus dans l'accomplissement de ses missions, en permettant d'assurer un accès universel à son patrimoine. Comme le souligne Kelly (2013), il s'agit d'une décision qui ne se limite pas à des dimensions technologiques ou légales : « Museums exist to educate and serve their various audiences and that providing access to images of works in the collection is part of the institutional mission. »

Le *Rijksmuseum* est l'un des pionniers en la matière. Le 30 octobre 2012, le musée mettait en ligne 125.000 images, téléchargeables gratuitement en haute résolution. Par cette action, l'institution muséale voulait permettre au public de copier les œuvres et de se les approprier, comme l'explique Tacco Dibbits, directeur des collections : « With the Internet, it's so difficult to control your copyright or use of image that we decided we'd rather people use a very good high-resolution image of "Milkmaid" from the Rijksmuseum rather than using a very bad reproduction »⁸ (Siegal, 2013).

D'ailleurs, il n'a pas fallu attendre longtemps avant que de fervents convaincus se rassemblent pour encourager l'ouverture des données dans le secteur culturel. Aux Pays-Bas par exemple⁹, le réseau *Open Culture Data*¹⁰ a vu le jour à la fin de l'année 2011. Par le biais de *hackatons* et de *masterclass*, le réseau cherche à guider les institutions à travers ce processus, tout en mettant en contact les propriétaires des données avec les développeurs d'applications et les dirigeants. Baltussen *et al.* (2013) relèvent les questionnements de nature légale et technologique auxquels sont confrontées les institutions, mais surtout, ils observent quels sont les plus grands risques perçus par les institutions. Ils en évoquent cinq : la perte d'attribution, la perte de contrôle, la

8. Cette anecdote a d'ailleurs inspiré la création d'un Tumblr qui met à l'honneur ces multiples versions d'une même œuvre <http://yellowmilkmaidsyndrome.tumblr.com/> (consulté le 02.11.14)

9. Voir également le réseau financé par la Commission Européenne, *OpenGLAM* : <http://openglam.org/> (consulté le 10.12.14)

10. <http://www.opencultuurdata.nl/> (consulté le 10.12.14)

perte de revenu potentiel, la perte de valeur de la marque et enfin la question de la vie privée¹¹.

Avec le développement du web sémantique, qui sous-entend la transition d'un web de documents vers un web de données structurées (compréhensibles par les machines)(Courtin, 2014a), ce sont de nouvelles potentialités qui sont discutées au sein de réseaux informels comme *LODLAM (Linked Open Data in Libraries, Archives, and Museums)*¹².

S'inspirant du fonctionnement de *Zotero*¹³, Owens (2014) esquisse la façon dont les descriptions issues du *crowdsourcing* pourraient être liées. Son espoir serait d'avoir des espaces sur le web qui centraliseraient la production des descriptions, transcriptions et annotations relatives à des œuvres issues de différentes institutions. Les institutions, envisagées comme des espaces de « dépôt » des œuvres, pourraient donc faire appel à des plateformes tierces pour ce qui est du *crowdsourcing*, avant de récupérer dans un second temps les descriptions qui leur semblent les plus pertinentes.

1.3 Intelligence collective et culture participative

Jenkins (2013) a introduit la notion de convergence médiatique¹⁴. Selon lui, le développement du web est une évolution du paysage médiatique plus qu'une révolution. Cette notion de convergence médiatique repose sur deux phénomènes qui seront également à l'œuvre dans le cas du *crowdsourcing* : l'intelligence collective et la culture participative.

L'économiste Surowiecki (2004) a développé ce concept d'intelligence collective dans un essai qui a marqué les esprits : *The Wisdom of Crowds : Why the Many are Smarter than the Few*. En recourant à de multiples exemples, il montre comment, selon lui, un groupe d'individus est plus intelligent qu'un individu isolé, aussi intelligent soit-il. Il base son hypothèse sur le fait que l'agrégation d'informations variées favorise la prise de meilleures décisions.

Si cette intelligence collective permet de résoudre des problèmes, elle permet également la co-création de nouvelles idées. Brabham (2013) explique que celle-ci est encouragée par Internet :

The Internet encourages ongoing cocreation of new ideas. Content on the Internet is generated through a mix of bottom-up (content from people) and top-down (content from policymakers, businesses, and media organizations) processes, as

11. Par exemple lorsqu'il est question de données généalogiques.

12. <http://lodlam.net/> (consulté le 12.02.14)

13. Outil de gestion bibliographique

14. « Par convergence j'entends le flux de contenu passant par de multiples plateformes » (Jenkins, 2013)

opposed to solely a top-down model. (...) Internet users are learning how to broadcast their own ideas, uncover buried informations, and remix previous ideas and content into new, innovative forms. Internet users are potentially creative problem solvers. (Brabham, 2013)

Shirky (2005) évoque également cette intelligence collective dans un propos relatif aux folksonomies. Il explique que seul le groupe peut permettre de catégoriser l'ensemble des données. C'est la multiplication et la superposition de « voix » indépendantes qui permettrait d'augmenter la qualité du résultat et de garantir une certaine objectivité.

Bien qu'elle soit très fréquemment évoquée dans le contexte du web 2.0, cette croyance en une intelligence collective ne fait cependant pas l'unanimité. Lanier (2006) réfute cette idée selon laquelle une force supérieure à la conscience individuelle émergerait du collectif. Il reconnaît les qualités du groupe lorsqu'il est question de résoudre des problèmes qui exigent des résultats évaluables objectivement. En revanche, il affirme que le groupe n'est pas forcément bon lorsqu'il est question de goûts et de jugements. À travers deux exemples, les wikis et les méta-méta agrégateurs de contenu, il montre comment le recours à l'intelligence collective finit par causer l'uniformisation et l'appauvrissement de l'information.

Par ailleurs, le web favorise également la publication en ligne de jugements profanes¹⁵ sur les biens et services. Cette pratique participative, voisine du *crowdsourcing*, mobilise également le concept d'intelligence collective. Elle consiste en la publication de commentaires courts, critiques argumentées ou notes (par exemple, nombre d'étoiles), sur l'un ou l'autre de ces dispositifs : blogs, réseaux sociaux, forums, sites de vente ou d'information sur les produits (Pasquier, 2014). Si l'on retrouve, comme dans certains projets de *crowdsourcing*, un système de commentaires rédigés par des personnes qui ne sont pas expertes, il faut bien évidemment souligner que contextes, motivations et enjeux sont différents.

Dans une récente publication, Pasquier (2014) présente une liste des constats communs auxquels aboutissent différents chercheurs en sciences sociales dans le cadre de leurs travaux de recherche sur ce phénomène. En dépit de la spécificité des secteurs concernés, il relève tout d'abord une pression poussant à la positivité de l'évaluation, une concentration de l'attention sur certains produits, le profil particulier des gros contributeurs et les relations complexes qu'ils entretiennent avec les professionnels de l'univers de référence. Par ailleurs, il distingue deux types de constats : ceux qui sont liés à la participation en ligne en général et ceux qui concernent l'exercice critique en tant que tel. Ainsi, certaines de ses observations, portant sur la dimension participative, permettent également d'éclairer le phénomène du *crowdsourcing* et de l'inscrire au sein

15. L'expression utilisée par Pasquier (2014) permet de distinguer ce type de jugements des critiques émises par des professionnels.

de l'écosystème commun qu'est le web participatif. Il s'intéresse par exemple au profil des contributeurs et explique qu'il est difficile d'en dresser un profil fidèle :

Qui sont les évaluateurs ? C'est une question compliquée, tant les effectifs et les profils des internautes varient d'un secteur d'évaluation à un autre et, à l'intérieur d'un même secteur, d'un site à l'autre : [...] il est important de tenir compte à chaque fois de la spécificité des biens ou des services évalués ou de l'effet de cadrage des dispositifs socio-techniques. Il serait donc vain d'espérer avoir une image d'ensemble d'une population qui se monte sans doute à plusieurs millions d'individus dans le monde, avec un fort *turn-over* d'entrants et de sortants, et des engagements dans la pratique très hétérogènes. C'est d'ailleurs là un problème classique auquel sont confrontés les chercheurs en sciences sociales qui travaillent sur des populations d'internautes contributeurs. (Pasquier, 2014)

Enfin, Pasquier (2014) évoque un cadrage idéologique du web 2.0 « [qui] s'affirme dans l'appel à l'évaluation des internautes : la culture de la participation, la démocratisation de la prise de parole en public pour affirmer ses goûts, l'idée d'une intelligence collective qui met en déroute la suprématie des experts, la glorification de l'activité créative des individus ». Ainsi, bien que d'évidents enjeux commerciaux opposent commentaires d'évaluation et commentaires visant à enrichir les métadonnées de collections culturelles, il semble toutefois important de ne pas oublier que le *crowdsourcing* est une pratique s'inscrivant dans un contexte plus large, qui ne cesse d'évoluer et qui agit donc comme un « cadrage idéologique ».

Chapitre 2

Enjeux

2.1 Numérisation et mise en ligne des collections

Le lancement d'un projet de *crowdsourcing*, lorsqu'il porte sur des documents qui ne sont pas nativement numériques, requiert deux étapes préalables : la numérisation des collections, ainsi que leur mise en ligne. Ces étapes impliquent de nombreux enjeux : il s'agit, pour l'institution concernée, de préserver et de sauvegarder le patrimoine culturel dont elle est responsable, mais également de le rendre accessible, de le valoriser et de favoriser son appropriation par les utilisateurs. D'autre part, la numérisation poursuit la transformation de l'accès physique aux œuvres. L'expérience du spectateur n'est plus la même. Des collections, autrefois accessibles uniquement en un temps et en un lieu donnés, deviennent accessibles à tous, en permanence et gratuitement, sur le web. Cela entraîne une profonde modification du rapport au patrimoine culturel concerné : celui-ci, désormais visible sur écran d'ordinateur et reproductible, perd donc l'aura qui le caractérisait (Benjamin, 2000).

Plus concrètement, la numérisation sous-entend « la conversion d'éléments issus du monde analogique en éléments qui relèvent du monde numérique » (Claerr et Westeel, 2011). En assurant la préservation de l'information contenue dans un document au-delà de la durée de vie du support, elle se présente ainsi comme une nouvelle mesure de conservation préventive¹. En outre, elle multiplie les possibilités de communication autour du document et favorise les actions de valorisation, par exemple par le biais de la mise en ligne des produits de la numérisation.

Récemment, deux représentants de la *Tate* (Fildes et Foster, 2015) ont retracé dans un article les différentes phases d'un vaste projet de valorisation des archives : de leur numérisation à leur intégration au sein de l'interface préexistante² du site officiel de

1. En présence de documents anciens et fragiles, qui sont condamnés à une dégradation progressive, la numérisation « permet d'arrêter temporairement ou définitivement la communication des documents originaux » (Bruleaux et Giovannini, sd)

2. Page *Art and artists* : <http://www.tate.org.uk/art/> (consulté le 10.05.15)

l'institution. La stratégie adoptée visait à utiliser le système de gestion de contenu déjà employé pour les collections, afin que les archives se situent dans une section du site web drainant d'ores et déjà un large public. Fildes et Foster (2015) exposent notamment la façon dont les différents services ont été impliqués, des photographes aux responsables des finances, en passant par les conservateurs. Cet article est éclairant car il évoque de façon très pragmatique les difficultés auxquelles les différentes équipes ont dû faire face, comme par exemple le traitement de documents complexes, constitués de plusieurs éléments³. En outre, il illustre parfaitement à quel point la numérisation s'inscrit dans une stratégie globale nécessitant de multiples réflexions et actions sur la façon dont les documents vont être présentés, à quelles fins, *etc.*

Métadonnées

Comme le soulignent Benhamou et Thesmar (2011), « les enjeux de la numérisation vont en effet au-delà de la mise à disposition des publics ; ils incluent la visibilité de ce patrimoine, la qualité du référencement, des métadonnées et de l'indexation. » La numérisation n'est donc qu'une première étape. Une fois les images numérisées, quelqu'un doit les décrire correctement. En effet, bien que certaines collections soient déjà bien documentées, des millions d'artéfacts ne disposent encore que de très peu ou d'aucune information descriptive les concernant (Owens, 2013). Les métadonnées sont au cœur de ce processus de sauvegarde et de valorisation. *The National Information Standards Organization (NISO)* les définit ainsi : « *structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information* »⁴.

De plus, il faut savoir que si les métadonnées sont une part essentielle du processus de conservation et de sauvegarde des collections, elles le sont d'autant plus lorsqu'il s'agit de collections iconographiques. En effet, ce sont des documents auxquels on ne peut accéder à l'aide d'un moteur de recherches s'ils ne disposent pas de métadonnées descriptives (Mesguich et Amar, 2012). Autrement dit, s'il n'est pas possible de tomber sur ce document par hasard, il n'est pas non plus envisageable de le retrouver facilement vu qu'aucune métadonnée ne permet de le distinguer des autres ! Il est donc indispensable de le décrire lors de l'étape du catalogage. Or, le catalogage représente une charge de travail colossale et nécessite des investissements humains et financiers importants. C'est à ce stade que le *crowdsourcing* entre en jeu : il s'agit de donner la possibilité aux internautes de contribuer eux-mêmes à la description des contenus.

3. Sous quelle forme faire apparaître des carnets de croquis dans les résultats de recherche ? S'ils apparaissent de façon groupée, alors comment procéder en ce qui concerne les métadonnées, vu que chaque page possède ses propres métadonnées ?

4. (<http://www.niso.org/publications/press/UnderstandingMetadata.pdf>) (consulté le 24.04.14)

Concrètement, il existe différents types de métadonnées, que Harvey (2010) classe selon les fonctions qu'elles remplissent : 1. Description des données et de leur emplacement ; 2. Présentation des informations techniques nécessaires pour utiliser les données ; 3. Description du cycle de vie des données du point de vue de la conservation. Dans le cadre d'activités de *crowdsourcing*, les métadonnées fournies par les utilisateurs rentrent généralement dans la première catégorie : il s'agit d'identifier l'auteur du document, de décrire son contenu, *etc.*

Interopérabilité

Par ailleurs, aborder la question des métadonnées renvoie à d'autres problématiques qu'il est nécessaire de garder à l'esprit : les standards et les questions d'interopérabilité. En effet, dans un contexte où les efforts se multiplient pour lier les données, il devient stratégique d'avoir des métadonnées qui puissent circuler d'une institution à l'autre. Ainsi, l'utilisation de standards permet d'assurer que l'information descriptive comporte tous les éléments requis, du point de vue de la structure, de la sémantique, et de la syntaxe (Harvey, 2010).

Cependant, sachant que différents types de standards sont disponibles, les difficultés en terme d'interopérabilité des métadonnées subsistent. En effet, si l'on pense à un projet comme Europeana⁵, cette bibliothèque numérique qui vise à rendre accessible au public l'héritage culturel et scientifique européen issu de multiples institutions, il est évident qu'une certaine cohérence est indispensable. Il faut que les données issues de systèmes différents puissent être échangées et rassemblées dans une même base de données. Dans le cas de ce réseau européen, la solution fut de créer un nouveau modèle : *Europeana Data Model*. Inspiré de différents standards préexistants, l'*EDM* permet d'uniformiser des données qui étaient jusque là très hétérogènes (pratiques divergentes selon les institutions, les pays, *etc.*)⁶. Ces problématiques vont également concerner les cas de *crowdsourcing* comportant un projet de réintégration des métadonnées créées par les utilisateurs sur le site web de l'institution ou au sein d'autres ressources.

Par ailleurs, d'autres enjeux sont liés aux métadonnées, comme la problématique du multilinguisme. Lorsque la communauté contribuant à un projet de *crowdsourcing* est internationale, la question de la langue des commentaires ou des tags ne devrait pas être négligée. Un système de traduction pourra-t-il être introduit ? Quid des commentaires postés en espagnol, en portugais ou en chinois... ? Seront-ils tout simplement ignorés par une équipe ne maîtrisant pas ces langues ? Finalement, si l'institution ambitionne de proposer dans un futur relativement proche des données qui peuvent être « liées »

5. <http://www.europeana.eu/> (consulté le 11.11.14)

6. <http://labs.europeana.eu/api/linked-open-data/data-structure/> (consulté le 11.11.14)

avec celles d'autres institutions, alors le multilinguisme s'inscrit également au sein des problématiques qui nécessitent une véritable stratégie en amont.

Droit d'auteur

Enfin, une notion qui peut sembler anecdotique de prime abord, mais qui soulève toutefois des interrogations : la question du droit d'auteur.

Peuvent-elles [les métadonnées] être des œuvres de l'esprit protégées par le droit d'auteur ? Nous l'avons vu, pour être protégée une œuvre doit répondre aux conditions d'originalité. Peut-on considérer comme originales les références attachées à un document ? Il est difficile de répondre de manière péremptoire par la négative, car le travail d'analyse pourrait, dans certains cas, être considéré comme une création intellectuelle. Le plus souvent, il n'existe pas vraiment d'originalité dans la formulation de ces métadonnées considérées indépendamment du document qu'elles identifient. (Claerr et Westeel, 2011)

Le *crowdsourcing* est une démarche de co-création qui semble s'inscrire dans la lignée de la culture libre, il serait dès lors surprenant que des hordes d'utilisateurs revendiquent avec virulence la paternité de leurs commentaires. Cependant, cela ne dispense pas de rester lucide sur ces questions de propriété intellectuelle. En effet, si les métadonnées ne sont pas protégées par le droit d'auteur dans la majorité des cas, Claerr et Westeel (2011) nuancent le propos : « Toutefois, elles [les métadonnées] peuvent être organisées à l'intérieur d'une base de données qui accompagne le document : même si les métadonnées ne sont pas protégées, la base de données, elle, peut être protégée en tant que telle par le droit d'auteur et le droit *sui generis*. »

2.2 Crowdsourcing

Lorsqu'une institution entreprend un projet de *crowdsourcing*, elle pose un geste qui aura forcément des répercussions, qu'elles soient positives ou négatives. Elle intègre une nouvelle dimension dans son fonctionnement, elle propose une forme inédite d'interaction à ses usagers, elle se positionne par rapport au numérique, *etc.* Si le recul est encore faible pour saisir la nature et l'ampleur des impacts que le *crowdsourcing* produira à plus long terme, il est néanmoins déjà possible de relever un certain nombre d'enjeux.

Figure d'autorité

Premièrement, en donnant la possibilité aux internautes de s'exprimer, en laissant proposer du contenu et en acceptant qu'ils puissent venir mettre en doute

les informations qu'elle délivre, l'institution accepte que la distribution des rôles ne soit plus tout à fait la même. Ceci entraîne une modification de la figure d'autorité qu'incarnent traditionnellement les temples du savoir que sont les musées, bibliothèques et archives :

L'institution, auparavant détentrice exclusive du droit à délivrer l'information savante dans son enceinte, se trouve en situation de partager son autorité avec des usagers profanes. Cette dernière se transforme, puisque partagée, dans une certaine mesure, avec les co-créateurs de contenus. (Vidal et Crenn, 2010)

Allant plus loin encore, Brabham (2013) explique que le succès du *crowdsourcing* repose sur un délicat équilibre entre les forces provenant du haut (l'institution) et du bas (les utilisateurs) :

For mutual benefit to be enjoyed, I argue, the locus of control in the creation of goods and ideas in crowdsourcing must reside between the organization and the community in a shared space that maximizes the benefits of top-down, traditional management with the benefits of bottom-up, open creative production.

L'institution sera confrontée à de multiples questionnements : comment peut-elle aborder ces changements ? Est-ce que l'émergence d'une figure d'« amateur-expert » peut venir l'éclipser ou menacer sa crédibilité ? Faut-il exercer un contrôle, une modération, une censure ? Tout l'enjeu, pour l'institution, est de réussir à lâcher du lest sans toutefois perdre la maîtrise.

Qualité

Deuxièmement, ces nouvelles pratiques soulèvent évidemment l'importante question de la qualité des fruits du *crowdsourcing*. Par le biais des commentaires qu'il poste, l'utilisateur crée lui-même des métadonnées. Mais quelle est la valeur de ces métadonnées ? Sont-elles fiables ? Exactes ? Justifiables ? Cohérentes ? Quant à celui qui poste, quelle est sa légitimité ? A-t-il une expertise en la matière ?

Cependant, que la qualité soit au rendez-vous ou pas, il apparaît, en optant pour un second degré de lecture, que les commentaires possèdent un autre type de valeur informative (Shirky, 2005) : ils renseignent l'institution sur les intérêts du grand public, sur sa façon de s'exprimer, sur le rapport qu'il entretient avec les œuvres, *etc.* Ils peuvent ainsi nourrir une réflexion plus globale sur ce que l'institution désire partager, sur la façon dont la médiation peut être pensée, sur le contenu qui mérite d'être valorisé, *etc.*

Communication

En s'engageant dans un projet de *crowdsourcing*, l'institution instaure un nouvel espace de communication, en ligne. Les interactions avec le public s'en trouvent

grandement modifiées. En ligne, une forme d'expression inédite permet d'échanger dans un rapport de proximité qui tranche radicalement avec ce que proposent les canaux de communication traditionnels, connus pour être beaucoup plus formels. Désormais, la communication n'est plus seulement unidirectionnelle. La distance, symbolique ou réelle, qui séparait les visiteurs de l'institution s'en trouve réduite.

Internal communication is only part of the equation given the fundamental importance of museums interacting with the public. The traditional communication ideology of the museum has been to see the museum as the holder of knowledge and truth with a responsibility to exercise one-way communication to the public. In the reinvented museum, communication *between* museum and public is exemplified by a mutually respectful relationship; the ideology of the two-way communication *with* the public creates a more responsive interchange of ideas and shared understandings supplanting the more traditional, paternal approach. (Anderson, 2012)

Mais si les liens peuvent se renforcer par le biais du web, ils peuvent également s'abîmer. Alors, afin de prendre soin de ce nouveau type d'interaction (satisfaire les attentes des visiteurs, garder une certaine cohérence par rapport aux actions passant par les canaux de communication traditionnels, *etc.*), les institutions accueillent en leur sein de nouveaux postes comme celui de « *community manager* » ou de « chargé de projets numériques ». Si toutes les institutions qui se lancent dans un projet de *crowdsourcing* ne peuvent se permettre d'engager de nouveaux employés pour ce type de missions, elles ne peuvent toutefois pas faire l'impasse sur certaines questions et doivent déterminer comment mettre en place et gérer au quotidien ce nouveau type de communication (Beudon, 2014).

Nouveaux usages

Afin de continuer à remplir leurs missions, il est essentiel que les institutions culturelles restent en phase avec leur époque. Solliciter la participation des utilisateurs à travers un projet de *crowdsourcing* constitue donc une façon de s'adapter à leurs usages.

Les bibliothèques s'inscrivent dans un écosystème du web où l'interaction est la norme : l'internaute s'attend à pouvoir intervenir sur les données et sur les contenus, que ce soit pour les partager ou les enrichir. Même lorsqu'il n'utilise pas ces fonctionnalités, elles lui sont familières dans sa pratique courante du web, sur les réseaux sociaux ou les sites marchands. Elles constituent son cadre de référence, il se sentira enfermé et exclu s'il ne les a pas à sa disposition. (Moirez, 2013)

En investissant le web, les institutions s'ajustent à leurs publics et élargissent le cercle des personnes qu'elles peuvent potentiellement toucher par leurs actions. Ce faisant, l'institution envoie un signal fort. Elle affirme sa présence en ligne, ce qui va entraîner des réactions du public et peut influencer sur la façon dont son identité est perçue. Par exemple, si elle ne maîtrise pas certains « codes », l'implémentation maladroite d'un projet de *crowdsourcing* pourrait finir par lui porter préjudice... L'enjeu étant de se positionner comme une institution dynamique, qui s'adapte à l'évolution des technologies et qui se démarque par son usage éclairé et réfléchi du web.

Médiation

Du côté du grand public : une nouvelle relation aux œuvres se constitue. De la découverte du document, qui se fait sur un écran, jusqu'à l'analyse poussée des détails, réalisée parfois en collaboration avec d'autres utilisateurs, l'expérience est transformée. Le *crowdsourcing* peut donc également être considéré comme une façon de s'approprier les contenus. Évidemment, envisager le *crowdsourcing* comme un nouveau dispositif de médiation (Courtin, 2014b) signifie la prise en compte de nouveaux enjeux, tel le fait d'augmenter la visibilité des œuvres ou encore de favoriser de riches échanges autour des œuvres. En allant plus loin, le *crowdsourcing* peut même être abordé comme un outil pédagogique, à l'instar par exemple de ce projet sur l'histoire « cachée » du Mississippi auquel participent le grand public, mais aussi des classes et des professeurs⁷.

Many museums now accept the « constructivist » view that knowledge is actively produced by a learner, focusing not on what an individual learns but on what the museum contributes to his or her existing knowledge. Museums have an inherent conflict with this view, however, because of their role as authoritative cultural institutions. To accord visitors with the power to construct their own meanings is to undermine curatorial authority. In so doing, a museum not only risks losing its authority, but it also absolves itself from interpretive responsibility for the meanings it produces and circulates in the culture. (Tallon et Walker, 2008)

Souvenirs

Plus concrètement, il apparaît que les plateformes se font le réceptacle de témoignages personnels : anecdotes concernant la vie quotidienne d'un quartier, souvenirs relatifs à une institutrice, description d'un *crash* d'avion, *etc.* Les utilisateurs se font producteurs de contenus. Il faut cependant relativiser la nouveauté de cette tendance :

For years, the space of the museum has been the preserve of curators and educators, who were solely responsible for the museum's content. In recent

7. <http://www.historypin.org/project/411-mississippi-history/> (consulté le 01.05.15)

decades there has been an increasing insistence on « bottom-approaches » that open the museum to other voices and other constituencies. The idea of visitors contributing to the museum space is not new, nor is it depend on digital technologies - after all, as the « Memory » exhibition at the San Fransisco Exploratorium showed over a decade ago, sticky notes and a box of pencils is enough to invite a community to rewrite its own history. Nevertheless, the suggestion that visitors can – and should – contribute to the museum environment is both radical and provocative. (Tallon et Walker, 2008)

Par ailleurs, le *crowdsourcing* permet de préserver une mémoire appelée à disparaître. Si l'intention n'est pas toujours clairement affichée par l'institution à l'origine du projet, le processus semble s'imposer de lui-même dans les espaces de commentaires de documents d'archives montrant des lieux, des événements ou encore des personnes. Les plateformes de *crowdsourcing* représentent donc un espace à la fois intime et public, propice au partage d'informations qui, sans cela, seraient condamnées à disparaître avec le décès de leur auteur.

Pérennité

Enfin, l'un des enjeux essentiels du *crowdsourcing* est la notion de pérennité. La question concerne d'une part la plateforme accueillant le projet. Il s'agit de savoir si sa durée d'existence est fixée, si l'information est clairement annoncée, si l'évolution technologique a déjà été anticipée, *etc.* D'autre part, la notion de pérennité se rapporte également aux métadonnées créées par les utilisateurs. Vont-elles être récupérées et rediffusées dans un autre lieu de destination, par exemple sur le site web de l'institution ?

L'enjeu des bibliothèques est donc, au-delà d'une réponse à cette attente des internautes, de faire le meilleur usage possible des contributions des usagers pour enrichir les métadonnées descriptives de leurs collections et pour améliorer l'expérience de recherche et de navigation des utilisateurs. Il serait en effet dommage de n'utiliser les potentialités du web social que de façon « cosmétique », sans en faire véritablement bénéficier le signalement des collections et l'interface de recherche de la bibliothèque. (Moirez, 2013)

Si ces métadonnées sont destinées à être intégrées sur le catalogue en ligne, il s'agira également de réfléchir à la façon dont elles coexisteront avec les métadonnées préexistantes. Vu le succès grandissant des *Linked Open Data*, il sera également intéressant de réfléchir à la façon dont ces métadonnées pourraient être intégrées à des plateformes qui jouent un rôle d'agrégateur entre les données des différentes institutions culturelles, comme le fait par exemple *Europeana*. Cette intégration ne se

limite cependant pas à des questions de faisabilité technique, elle soulève des questions délicates déjà évoquées plus haut : multilinguisme, interopérabilité, fiabilité, *etc.*

Chapitre 3

État de l'art

3.1 Définitions et typologies

Depuis l'introduction du terme par Howe (2006), de nombreuses interprétations et définitions ont émergé. Estellés-Arolas et de Guevara (2012) expliquent qu'au cours d'une analyse de la littérature académique, ils ont répertorié près de 40 définitions différentes. Dans un premier temps, ils ont identifié les sept caractéristiques-clés permettant d'aborder le *crowdsourcing* : la foule ; la tâche à accomplir ; l'initiateur de l'activité ; ce qui est obtenu à l'issue du processus ; le type de processus ; l'appel à participation ; le médium. Ils aboutissent ainsi à une nouvelle définition, qui se veut plus exhaustive que les définitions préexistantes :

Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or compagny proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge, and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (Estellés-Arolas et de Guevara, 2012)

Outre cette première définition qui a la qualité d'être très englobante, la définition soumise par Brabham (2013) mérite également d'être citée. Elle se caractérise par son caractère synthétique qui permet de rapidement comprendre les enjeux en présence. La définition consiste en un résumé des éléments-clés, constitué de quatre parties :

- An organization that has a task it needs performed.
- A community (crowd) that is willing to perform the task voluntarily.
- An online environment that allows the work to take place and the community to interact with the organization.
- Mutual benefit for the organization and the community.

Cette seconde définition a le mérite de mettre en exergue la notion de communauté, qui joue un rôle crucial en ce qui concerne la motivation des contributeurs (Moirez, 2014), ainsi que la notion de bénéfice mutuel. En revanche, elle restreint le profil du *crowdsourcer* à celui d'une organisation, quand Estellés-Arolas et de Guevara (2012) y incluent également l'individu seul.

Si ces définitions permettent de clarifier ce que signifie le terme *crowdsourcing*, il est toutefois utile d'opérer certaines distinctions. Comme le soulignent notamment Dunn et Hedges (2013), le concept ne doit pas être confondu avec la production d'*user-generated content*, ni avec la collecte de données transactionnelles (informations issues de l'activité en ligne des utilisateurs). De même, le *crowdsourcing* n'est pas identique aux sciences citoyennes, pas plus qu'à l'engagement social. Cependant, Dunn et Hedges (2013) nuancent ce type de distingo en expliquant que ce sont des concepts voisins, qui vont parfois s'enchevêtrer. Ils illustrent ce propos par l'exemple du *crowdfunding* : contribution intellectuelle et contribution financière ne sont pas mutuellement exclusives.

Si ce mémoire traite spécifiquement de l'enrichissement de métadonnées liées au patrimoine culturel, il faut savoir que le recours au *crowdsourcing* est loin de se restreindre à ce secteur. Les mécanismes d'« externalisation vers la foule » (Renault, 2014) séduisent également les milieux des entreprises et de l'innovation. La littérature sur le sujet est donc très riche et il existe de multiples typologies, qui abordent cette activité selon des visions plus ou moins transdisciplinaires et englobantes. Dans un premier temps, il semblait judicieux de survoler ces recherches sans se limiter à un secteur en particulier, dans l'optique de relever les conclusions les plus pertinentes et de voir si elles sont également applicables au secteur culturel.

Les auteurs de ces typologies mettent l'accent sur différentes dimensions du *crowdsourcing* : la tâche qui doit être accomplie¹, la participation de la foule², les

1. (Schenk et Guittard, 2012; Nakatsu *et al.*, 2014)

2. (Renault, 2014; Bonney *et al.*, 2009)

processus adoptés par l'institution³ et enfin les ressources que la foule est appelée à fournir⁴.

Schenk et Guittard (2012) abordent le phénomène du point de vue de l'entreprise et de l'innovation⁵. Ils expliquent que le *crowdsourcing* intéresse le monde entrepreneurial dans le sens où il va lui permettre de répondre à un besoin particulier en tirant partie des ressources cognitives que possède la foule. Leur typologie est basée sur deux dimensions : la nature du processus et le type de tâches. Concernant la nature du processus, ils distinguent deux grandes approches : (a) le *crowdsourcing* intégratif (intégration des contributions de la foule dans une base de données); (b) le *crowdsourcing* sélectif (sollicitation de la participation de la foule suivie de la sélection de certaines contributions). La deuxième dimension concerne le type de tâches. Ils en répertorient trois : les tâches simples; les tâches créatives; la résolution de problèmes complexes. Outre cette typologie (également applicable aux cas de *crowdsourcing* issus du domaine culturel), ce qui retient l'attention c'est le fait de spécifier pour chaque tâche les compétences qui sont recherchées : les tâches simples requièrent des compétences génériques et multiples; les tâches créatives requièrent des compétences spécifiques et multiples; la résolution de problèmes complexes requiert des compétences spécifiques et rares. Identifier le type de tâche et lister les compétences requises avant de se lancer dans la concrétisation du projet sont des réflexes que gagneraient sans doute à adopter les institutions culturelles si elles ne le font pas encore.

Également basée sur les caractéristiques de la tâche à accomplir, la taxonomie proposée par Nakatsu *et al.* (2014) a le mérite d'inclure une dimension supplémentaire : l'interdépendance que nécessite certaines tâches. Elle se construit autour de trois dimensions : la structure de la tâche (elle peut être plus ou moins définie); l'interdépendance liée à la tâche (elle peut être résolue par un seul individu ou peut nécessiter le soutien d'autres membres de la communauté); le niveau d'engagement que requiert la tâche. À partir de ces trois dimensions, Nakatsu *et al.* (2014) identifient différents cas de figure possibles :

1. *Contractual hiring with low commitment – 1'. With high commitment;*
2. *Distributed problem-solving (additive/pooled coordination) with low commitment;*
3. *New idea generation - Solo with low commitment – 3'. With high commitment;*
4. *Collaboration (reciprocal coordination) with low commitment – 4'. With high commitment.*

3. (Geiger *et al.*, 2011)

4. (Howe, 2008)

5. Il s'agit plus précisément d'*open innovation* : « *its central idea is that companies should not only rely on their own research and development, but should also draw on knowledge developed in the environment and brought it to the firm* ». (Nakatsu *et al.*, 2014)

Renault (2014) s'intéresse également à cette dimension d'interdépendance. Elle part d'un constat : la participation peut être individuelle et compétitive ou empreinte de partage et de collaboration. Elle propose une taxonomie constituée de quatre catégories, qu'elle articule autour de deux axes : le niveau de concurrence et le niveau d'interaction entre les utilisateurs⁶. La notion de compétition ne semble pas fréquente dans le contexte culturel, excepté dans d'éventuels cas de gamification. Il en va de même pour le principe de rétribution monétaire, que Renault (2014) prend en compte dans sa recherche. Cependant, il serait peut-être intéressant que le secteur culturel se serve davantage de ce type de dynamiques compétitives pour augmenter le taux de participation et le degré d'engagement des utilisateurs.

Travaillant sur la thématique voisine des sciences citoyennes⁷, Bonney *et al.* (2009) proposent quant à eux une typologie qu'il est possible de transposer aux démarches de *crowdsourcing* dans le domaine du patrimoine culturel. Il s'agit de trois modèles qui visent à englober les diverses formes de participation : *Contributory projects* ; *Collaborative projects* ; *Co-created projects*. Ce qui distingue ces trois types de projets : le degré d'implication du public dans les différentes étapes du processus, son interaction avec l'organisation (qui est à l'origine du projet), ainsi que le niveau de compétences requis. Cette notion d'interaction avec l'organisation n'a pas encore été citée jusque là, mais semble toutefois très à propos dans le contexte culturel. En outre, il serait vraiment intéressant que les musées, bibliothèques et archives puissent davantage travailler sur cette notion de degré d'implication. Par exemple, en élaborant des scénarios avec divers degrés d'engagement dans le projet de *crowdsourcing*, selon le profil d'utilisateurs fictifs.

À travers leur taxonomie, Geiger *et al.* (2011) veulent distinguer les différents types de processus que peuvent adopter les organisations, de l'appel à participation jusqu'à l'agrégation des contributions. Pour constituer leurs catégories⁸, ils s'inspirent de 46 exemples, allant du site *TripAdvisor* (descriptions et évaluations d'hôtels) à *Wikipedia* (encyclopédie collaborative en ligne). Il faut tout d'abord relever que le terme *crowdsourcing* est donc employé ici dans un sens beaucoup plus étendu que dans son acception ordinaire. Ce que l'on peut retenir de leur travail, c'est la façon dont ils décomposent les différentes opérations ; ils relèvent quatre grandes phases. La

6. La taxonomie des « 4 C » : cumul (des participations) ; collaboration (des participants) ; compétition ; coopétition (concept à mi-chemin entre « coopération » et « compétition »).

7. Les sciences citoyennes, également appelées « sciences participatives », impliquent l'engagement de non-professionnels pour résoudre des cas réels de recherche scientifique, que ce soit dans le cadre de projets à long terme et à grande échelle, comme par exemple le relevé de nids d'oiseaux, ou dans le cadre d'études plus ciblées. (Bonney *et al.*, 2009)

8. Ils ont identifié 19 types qui, dans un souci de concision, ne seront pas détaillés ici, voir Geiger *et al.* (2011). Ces 19 types ont pu être regroupés en cinq grands types de processus : *Integrative sourcing without remuneration* ; *Selective sourcing without crowd assessment* ; *Selective sourcing with crowd assessment* ; *Integrative sourcing with success-based remuneration* ; *Integrative sourcing with fixed remuneration*.

première étape est l'éventuelle présélection des contributeurs potentiels, qui peut se faire : sur base de leurs qualifications ; selon un contexte particulier ; en combinant ces deux critères ; il peut n'y avoir aucune présélection. Cette première étape est en contradiction avec l'affirmation de Schenk et Guittard (2012), selon laquelle l'une des particularités constitutives du *crowdsourcing* réside dans le fait qu'aucune sélection des contributeurs n'est opérée *a priori*. La deuxième phase concerne l'accessibilité des contributions : sont-elles visibles (ou non), modifiables, évaluables par les autres utilisateurs ? Le troisième stade porte sur l'utilisation des contributions qui est faite par l'organisation : le processus peut être intégratif ou sélectif. Enfin, la quatrième étape concerne l'éventuelle rétribution dont bénéficient les contributeurs : elle peut être fixe, variable (basée sur le succès) ou inexistante. Si la rétribution financière est rare lorsqu'il s'agit du secteur culturel, en revanche il est pertinent de garder en tête ces quatre dimensions. En effet, il importe de réaliser qu'il n'existe pas une seule déclinaison possible du *crowdsourcing* et que l'institution qui décide d'y recourir possède une certaine marge de manœuvre. En outre, le secteur culturel pourrait également s'inspirer de certains de ces éléments, par exemple le concept de présélection : l'institution pourrait imaginer présélectionner les utilisateurs selon leur niveau de connaissances afin de leur proposer des tâches personnalisées.

Enfin, une typologie qui est fréquemment citée : celle de Howe (2008). Deux ans après avoir introduit la notion de *crowdsourcing*, il a soumis sa propre typologie, basée sur les quatre types de ressources que la foule peut être appelée à donner : 1. Intelligence collective ; 2. Travail créatif ; 3. Avis ; 4. Financement. Des réserves peuvent être émises concernant cette typologie. En effet, elle regroupe des formes qui sont très éloignées du concept de *crowdsourcing* tel qu'il a été défini en début de chapitre. Elle permet toutefois d'illustrer la variété des interprétations qui sont faites du terme, dont les contours restent très flous.

Il convient maintenant d'esquisser une vue d'ensemble des tâches que la foule est appelée à effectuer dans le secteur culturel en particulier. Si le type de documents au centre des projets peut varier (documents manuscrits, audios, audiovisuels ou iconographiques), ce qui retient surtout l'attention des chercheurs, c'est le type de tâches que l'utilisateur est amené à effectuer.

Il existe ainsi plusieurs typologies visant à circonscrire les différentes déclinaisons possibles du *crowdsourcing* ; une sélection non exhaustive sera présentée ici. La *Table 3.1* reprend les types de tâches recensés par cinq auteurs 1. Moirez *et al.* (2013)⁹ ; 2.

9. Dans le cadre du lancement de sa plateforme collaborative de correction et d'enrichissement des documents numérisés (<http://www.reseau-correct.fr/> (consulté le 10.04.15)), la Bibliothèque Nationale de France a travaillé avec huit partenaires afin de dresser un « État de l'art en matière de *crowdsourcing* dans les bibliothèques numériques », dans lequel ils proposent une typologie composée de cinq catégories.

Dunn et Hedges (2013)¹⁰ ; 3. Oomen et Aroyo (2011)¹¹ ; 4. Peccatte (2009)¹² ; 5. Ridge (2012, 2014)¹³.

	1	2	3	4	5
<i>Tagging</i> /folksonomies	x	x			x
Indexation/catalogage collab.	x	x		x	
Identif. documents iconogr.	x			x	x
Transcription collaborative	x	x	x		x
Correction/modification de contenus	x	x	x	x	x
Enregistrement et création de contenus		x	x	x	x
Publication de commentaires		x			x
Catégorisation/classification		x	x		x
Contextualisation		x	x		
Spatialisation de l'information		x			
Géoréférencement		x			x
Traduction		x			
Constitution d'un corpus			x	x	
<i>Crowdfunding</i>			x		
<i>Co-curation</i>			x		x
Redocumentarisation				x	
Synthèse de nouvelles informations					x

TABLE 3.1 – Vue d'ensemble | Typologies des tâches pouvant être effectuées par la foule

Concernant la *Table 3.1*, il faut noter que ces catégories sont non exclusives : il arrive en effet que certains projets combinent plusieurs types de tâches.

Par ailleurs, il arrive que certains auteurs regroupent dans leur typologie des éléments qui apparaissent chez d'autres sous forme de catégories distinctes ; c'est le cas par exemple des tâches de correction et de transcription, classées ici de façon séparée alors que Oomen et Aroyo (2011) ont tendance à les unir dans un seul type. Le cas contraire peut également se présenter, à savoir que Ridge (2014) englobe par exemple dans un même type la description d'artéfacts, alors qu'il s'agit de tâches qui peuvent être distincte : *tagging*, classification, *etc.*

En outre, dans l'optique d'uniformiser les types et de limiter leur nombre, quelques nuances n'ont pas pu être reprises ici, à l'instar de la distinction qu'opère Peccatte (2009) entre l'indexation libre et l'indexation contrôlée. C'est la même chose pour la catégorie que Dunn et Hedges (2013) nomment *linking* et qu'ils considèrent comme une sous-catégorie du *tagging*. D'autre part, la correction ou modification de contenus

10. Ils se focalisent sur les activités de *crowdsourcing* au sein des sciences humaines.

11. Leur typologie porte sur l'ensemble des *GLAM*, c'est-à-dire les galeries, bibliothèques, archives et musées.

12. Sa typologie concerne le *crowdsourcing* « appliqué aux images ».

13. Elle s'intéresse aux projets de *crowdsourcing* liés au patrimoine culturel, que ce soit dans un cadre institutionnel ou académique.

intègre bien sûr la correction d'OCR¹⁴, qui fait l'objet d'un type à part chez certains (Moirez *et al.*, 2013), mais aussi le « relevé d'erreurs dans des descriptions d'images qui appartiennent à un corpus déjà documenté par des professionnels » (Peccatte, 2009). Concernant la catégorie « enregistrement et création de contenus », elle englobe tant la production de contenus créatifs (Ridge, 2014) que l'enregistrement de témoignages oraux.

Enfin, la présence d'une catégorie intitulée *crowdfunding* est quelque peu « polémique ». En effet, des auteurs comme Ridge (2012) considèrent qu'il ne s'agit pas de *crowdsourcing*.

3.2 Motivations

Comme le soulignent Moirez *et al.* (2013), « les projets de *crowdsourcing* n'ont de sens qu'à partir du moment où l'on atteint une masse critique de contributions ». Afin de favoriser la participation des utilisateurs, l'institution peut utiliser différents leviers de motivation : thématique du corpus ; intérêt scientifique ; participation à une cause « citoyenne » ; envie de jouer (gamification) ; sentiment de communauté.

Ce qui est intéressant dans la démarche de Dunn et Hedges (2013), c'est qu'ils envisagent la question de la motivation tant du point de vue des contributeurs que de celui des initiateurs des projets (en l'occurrence les chercheurs en sciences humaines dans le cadre de leur étude). Ils relèvent que ce qui intéresse et motive les universitaires entreprenant de tels projets, c'est de numériser, créer ou traiter du contenu, ou encore de procéder à des interventions éditoriales. En ce qui concerne les contributeurs, ils notent que le facteur de motivation dominant est leur passion pour le sujet. Ils approfondissent également la notion de communauté, qui concerne tant les contributeurs que les initiateurs de tels projets.

Pasquier (2014) s'intéresse aux motivations des contributeurs dans un contexte voisin, celui de la publication de critiques amateurs. Il détaille différentes sources de motivations, également applicables au contexte du *crowdsourcing* : gratification personnelle, plaisir de s'exprimer, dynamique relationnelle, volonté de recevoir de l'attention, de montrer sa compétence ou encore d'affirmer sa supériorité, *etc.*

Schenk et Guittard (2012) traitent de projets impliquant les individus et l'entreprise, mais les stratégies visant à motiver la foule restent finalement similaires. Leur revue de la littérature leur a permis d'aboutir à la conclusion qu'il existe deux types de motivations : la motivation intrinsèque et la motivation extrinsèque. La première résulte de la satisfaction que l'utilisateur retire en accomplissant la tâche. La motivation extrinsèque est, elle, associée à diverses formes de rémunération. S'inspirant des

14. OCR : *Optical Character Recognition*, en français : reconnaissance optique de caractères.

recherches menées en psychologie, Lebraty et Lobre (2010) relèvent quant à eux non pas deux, mais trois types de motivations : intrinsèque et extrinsèque, ainsi qu'extrinsèque internalisée. Cette dernière désigne en fait un *continuum* de motivations mixtes, liées au sentiment de valorisation que procure la participation à un tel projet. Ce qui signifie que « les acteurs du *crowdsourcing* disposent donc d'une palette de motivations sur lesquelles ils peuvent s'appuyer, par la mise en place de stimuli variés tels que l'intérêt de la tâche, la rémunération et/ou la valorisation et/ou les opportunités professionnelles pour le vainqueur du challenge » (Lebraty et Lobre, 2010). En se basant sur le type de motivations et de priorités des individus, ils scindent la foule en deux grandes communautés¹⁵, la communauté « passionnée-compétente » et la communauté « compétente-passionnée ». La première, mue par des motivations intrinsèques, se focalise sur une tâche spécifique, plutôt que sur les fins auxquelles cette tâche permet d'aboutir. Tandis que la seconde, mue par des motivations extrinsèques, regroupe des membres pour qui la tâche est un moyen d'obtenir une rétribution. Avoir conscience de la coexistence de ces différents types de motivation offre à l'institution un nouvel outil pour stimuler la participation, par exemple en travaillant sur la reconnaissance ou les récompenses accordées aux utilisateurs.

Holley (2010) s'est appuyée sur ses expériences professionnelles au sein de la *National Library of Australia* pour dresser une très riche synthèse, dans laquelle elle évoque notamment les motivations qui peuvent animer les institutions ou les contributeurs.

Tout d'abord, elle liste les nombreux bénéfices que le *crowdsourcing* peut apporter aux bibliothèques : atteindre, dans un délai très rapide, des objectifs que ses ressources (humaines, financières, en temps) ne lui permettraient pas d'atteindre seule ; construire de nouvelles communautés virtuelles et groupes d'utilisateurs ; impliquer activement la communauté ; utiliser les connaissances, l'expertise et l'intérêt de la communauté ; améliorer la qualité des données ou y ajouter de la valeur ; rendre les ressources plus accessibles ; mieux connaître les désirs des utilisateurs ; démontrer la valeur et la pertinence de l'institution au sein de la communauté grâce à un niveau élevé de participation du public ; établir ou renforcer la confiance et la loyauté des utilisateurs ; favoriser chez l'utilisateur le développement d'un sentiment d'appropriation et de responsabilité envers les collections.

Concernant la motivation des utilisateurs, Holley (2010) affirme que la première moitié des volontaires actifs est mue par un intérêt personnel pour le sujet, tandis que la seconde le fait car elle souhaite se consacrer à des activités de volontariat¹⁶. Voici les

15. « Ces deux communautés ne sont pas étanches, mais constituent plutôt les extrêmes d'un continuum au centre duquel se situe une communauté hybride » (Lebraty et Lobre, 2010).

16. Voir également le passage (très instructif) consacré aux profils des contributeurs : <http://www.dlib.org/dlib/march10/holley/03holley.html> (consulté le 15.04.15)

facteurs de motivation qu'elle relève : l'activité est agréable/stimulante/intéressante ; la satisfaction de pouvoir se consacrer à une noble cause ; la possibilité d'acquérir de nouvelles connaissances ; le sentiment d'utilité ; la volonté de prouver que l'on est digne de la confiance accordée ; la satisfaction d'apporter sa contribution à un projet collectif de grande ampleur.

Si les sources de motivation ne varient pas fondamentalement d'un utilisateur à l'autre, en revanche il faut être conscient que leur engagement peut être radicalement différent. Holley (2010) a ainsi observé que la majorité du « travail » (jusqu'à 80% selon les cas) est réalisée par environ 10% des utilisateurs. Ce constat est à mettre relation avec le phénomène de la participation inégale en ligne, que Nielsen (2006)¹⁷ résume à l'aide du « principe 90-9-1 » : la plupart du temps, 90% des utilisateurs se contentent de lire ou observer et ne contribuent jamais ; 9% des utilisateurs contribuent de temps à autre ; 1% des utilisateurs participent énormément et sont à l'origine de la majorité des contributions.

3.3 Types de dispositifs et de plateformes

Les avancées technologiques facilitent grandement le processus de collecte lorsqu'une institution culturelle sollicite les contributions de ses utilisateurs. Les contingences matérielles sont moindres et le travail logistique est réduit si les internautes sont invités à proposer du contenu depuis leur ordinateur ou leur téléphone. Pourtant, certaines initiatives vont également avoir des répercussions *in situ*. Plusieurs cas de figure se présentent : collecte hors ligne (1), création d'une exposition inspirée d'un projet de *crowdsourcing* (2), conception d'une exposition à partir de ce que le public a préalablement plébiscité en ligne (3), projet incluant une collecte en ligne et hors ligne (4).

1. Collecte à l'ancienne

Il s'agit d'un appel à contribution adressé à la population et qui ne s'accompagne pas d'un dispositif dédié en ligne. Par exemple, l'équipe du *Mons Memorial Museum*¹⁸ est en quête de témoignages et de documents sur les deux guerres mondiales à Mons. L'annonce est accompagnée d'un formulaire de contact, mais aucune interface ne permet de soumettre directement des documents en ligne.

2. Une exposition en hommage aux contributeurs

L'*Australian National Maritime Museum* est à l'origine d'une initiative originale : ils avaient lancé un projet de *crowdsourcing* sur *The Flickr Commons* qui a eu un tel succès qu'ils ont décidé en février 2014 d'en faire une exposition *in situ*. Nommée

17. Le concept d'inégalité de la participation avait cependant déjà été introduit par Will Hill.

18. <http://www.monsmemorialmuseum.mons.be/> (consulté le 12.01.15)

HoodsHarbour, elle est décrite sur le blog de l'institution¹⁹ (consulté le 13.03.15) comme un hommage aux « super détectives » dont les efforts ont permis de résoudre les mystères qui se cachent derrière les photographies de la collection Samuel J. Hood.

3. Le grand public, auteur et curateur

Durant l'été 2008, inspiré par les idées véhiculées par Surowiecki (2004) dans son ouvrage *The Wisdom of Crowds*, le *Brooklyn Museum* marque les esprits et fait figure de pionnier en organisant une exposition dont le commissaire est le public connecté : *Click! A Crowd-Curated Exhibition*. La première étape : collecte de photographies à travers un appel à contribution ouvert à tous. Dans un second temps, les internautes peuvent indiquer sur un site dédié leur préférence parmi les 389 photographies reçues. Plus de 3000 personnes participent à cette évaluation, à partir de laquelle une exposition *in situ* est conçue. Par ailleurs, le *Brooklyn Museum* a ensuite publié un livre reprenant les 78 oeuvres photographiques, accentuant encore cette corrélation entre actions en ligne et hors ligne.

Johnson (2008) revient sur cette exposition en expliquant que son intérêt réside davantage dans les questions qu'elle soulève que dans la qualité artistique des photographies choisies. En s'appuyant sur le postulat de Surowiecki (2004) concernant l'intelligence collective, il s'interroge : le jugement de la foule dans le cadre de cette exposition surpassait-il celui d'un expert ? Comment l'évaluer sans point de comparaison, et qui serait à même de dire quelle est la meilleure sélection, celle de la foule ou celle de l'expert, sinon un autre expert ? Il insiste sur le fait qu'il faudrait pour ce faire posséder des critères appropriés pour évaluer si des jugements sont pertinents ou vrais²⁰.

4. Collecte en ligne et hors ligne

C'est le cas par exemple de la monumentale installation multimédia du *Museum of Copenhagen : The Wall*²¹. Il s'agit d'un écran plasma tactile de douze mètres sur deux, installé successivement à divers emplacements de la ville depuis avril 2010. L'interface interactive permet de découvrir la ville en 3D et rend accessible plus de 10.000 photographies issues des collections muséales ainsi que des récits permettant de découvrir Copenhague à travers le temps, l'espace ou une thématique particulière. En outre, le mur accueille également les contributions (photographies, films, *etc.*) que peut soumettre le public, soit en ligne, soit sur place, directement « sur » le mur (Sandahl *et al.*, 2011).

19. <http://anmm.wordpress.com/2014/02/26/hoodsharbour-exhibition/>

20. Surowiecki (2008) explique sur le blog accompagnant le projet (<http://www.brooklynmuseum.org/community/blogosphere/bloggers/2008/08/08/reflections-on-click-by-james-surowiecki/>) (consulté le 29.04.15) que c'est justement l'absence de normes objectives pour évaluer les goûts artistiques qui l'a dissuadé de traiter de l'art dans *The Wisdom of Crowds*.

21. http://www.copenhagen.dk/en/whats_on/the_wall/ (consulté le 12.02.15)

Trois types de plateforme

En ce qui concerne les interfaces qui permettent au public de participer, Moirez *et al.* (2013) dénombrent trois cas de figure qu'ils illustrent par des exemples. Il peut tout d'abord s'agir d'une plateforme interne au site de l'institution, comme par exemple *Trove*²², le dispositif de correction collaborative d'OCR de périodiques de la Bibliothèque Nationale d'Australie. Deuxièmement, cela peut également être une plateforme dédiée, créée spécifiquement pour le projet, c'était le cas par exemple de *Digitalkoot*²³ (consulté le 01.03.15), le dispositif de correction collaborative d'OCR de la Bibliothèque Nationale de Finlande, dont l'activité a cessé fin 2012. Enfin, l'institution peut faire appel à un prestataire : il s'agit alors d'une plateforme distincte préexistante, comme par exemple la bibliothèque numérique *Wikisource*²⁴ (consulté le 01.03.15).

Flickr : The Commons

*The Flickr Commons*²⁵, dont il sera question à de multiples reprises dans la seconde partie de ce mémoire, est l'une de ces plateformes distinctes préexistantes auxquelles les institutions peuvent faire appel. Le projet, lancé par Flickr²⁶ le 16 janvier 2008, en collaboration avec *The Library of Congress*, vise à augmenter l'accès aux œuvres détenues par des établissements publics et à favoriser l'apport d'informations par le grand public. Les images que les institutions décident de mettre en ligne sur *The Flickr Commons* seront ensuite librement téléchargeables, il importe qu'elles soient donc libres de droit ou du moins qu'aucune restriction de droit d'auteur ne soit connue. En mars 2015, ce sont 93 institutions qui y participent, dont 48 sont basées aux États-Unis ou au Royaume-Uni (Néroulidis, 2015).

Des réserves peuvent cependant être émises à l'encontre d'une telle option. En effet, la facilité d'implémentation permise par ce type de plateforme préexistante implique une certaine perte de liberté pour l'institution utilisant ce type de service. Hernebring a participé à l'engagement de la Bibliothèque de Toulouse dans le projet *The Flickr Commons*. Il explique que si la mise en œuvre technique du projet était facile, en revanche de multiples appréhensions ont émergé parmi les membres de l'équipe responsable. Elles concernaient les réactions du public, la valeur et la nature des contributions, mais également la légitimité de l'institution au sein d'un site communautaire, sa capacité à susciter l'envie de participer, sa « perte de contrôle »

22. <http://trove.nla.gov.au/> (consulté le 20.01.15)

23. http://www.digitalkoot.fi/index_e.html

24. <http://fr.wikisource.org/wiki/Wikisource:Accueil>

25. <https://www.flickr.com/commons> (consulté le 15.05.15)

26. Flickr est un site web de partage de photographies et de vidéos appartenant à Yahoo! <http://www.flickr.com> (consulté le 15.05.15)

en cas de réutilisation des images par les utilisateurs, le risque de perdre de son identité ou encore le fait que *Yahoo!* soit une entreprise commerciale (Mesguich et Amar, 2012).

3.4 Initiatives non conventionnelles

Outre les actions menées officiellement par les bibliothèques, musées et archives, d'autres types de projets émergent sur diverses plateformes : des initiatives privées et individuelles, couronnées parfois d'un franc succès. Deux grandes tendances semblent se dégager concernant ce type de *crowdsourcing*. Premier cas de figure : en se développant, le projet se structure ; il est pensé, coordonné, encadré et documenté par un ou plusieurs individu(s). Cette tendance sera illustrée par le projet *PhotosNormandie*. Deuxième cas de figure : le partage se fait de façon totalement informelle, voire anarchique ; il n'y a pas de modérateur apparent, ni de travail de documentation ou d'archivage. Le groupe Facebook *Audergem d'autrefois* permettra d'illustrer cette deuxième tendance.

*PhotosNormandie*²⁷ est un projet collaboratif, qui recourt depuis 2007 aux fonctionnalités offertes par la plateforme *Flickr* pour améliorer les légendes de plus de 3600 photographies de la Bataille de Normandie (été 1944)²⁸. Concrètement, ce sont plus de 9130 légendes qui ont été corrigées et mises à jour depuis janvier 2007. Peccatte (2012), co-créateur et administrateur du « compte Pro » *Flickr : PhotosNormandie*, explique que le projet est né en réaction à la médiocrité des légendes qui accompagnaient les photos du site web *Archives Normandie 1939-1945*²⁹. Il s'agit d'un travail à finalité patrimoniale, réalisé par un collectif bénévole : « *PhotosNormandie* correspond en fait à une démarche documentaire et rédactionnelle, à un travail collectif avec un objectif de production. C'est une entreprise qui vise à traiter à nouveau une collection de documents en utilisant les fonctionnalités sociales d'une plateforme de partage » (Peccatte, 2012).

Les commentaires d'utilisateurs et les statistiques en disent long sur le succès de cette initiative. Peccatte (2012) estime que cette réussite repose sur la participation d'un noyau dur de passionnés, « un groupe de recherche informel mais très motivé », ainsi que sur le travail d'animation qui est fait (réponses, sollicitations de nouveaux contributeurs, modération, *etc.*). Le projet est d'ailleurs extrêmement documenté et témoigne d'une véritable réflexion sur les tenants et aboutissants de cette production collaborative : en témoignent la page de profil bilingue du compte Flickr, un tutoriel destiné aux contributeurs potentiels, ainsi que ce bilan très détaillé publié lors des 5 ans

27. <https://www.flickr.com/people/photosnormandie/> (consulté le 12.05.15)

28. Celles-ci proviennent des Archives Nationales des États-Unis, du Canada et de la Bibliothèque de Cherbourg-Octeville

29. Le site web n'est aujourd'hui plus accessible ; il avait été mis en place en 2004 par le Conseil Régional de Basse Normandie.

du projet³⁰. Ce mode de gestion très formalisé s'observe également dans la façon dont les fichiers sont nommés et organisés. Les photographies sont rigoureusement classées et rassemblées en sous-ensembles constitués de 100 photos, tandis que les noms des fichiers obéissent à de strictes règles de cohérence. Enfin, il s'agit d'un projet global s'étendant au-delà de ces seuls albums partagés sur *Flickr*. En effet, *Flickr* est un outil parmi d'autres : requêtes sur des forums et autres sources d'informations, recherche de la présence de ces mêmes photographies à d'autres endroits du web, *etc.*

*Auderghem d'autrefois*³¹ fait partie des nombreuses initiatives privées qui, par le biais d'un groupe *Facebook* par exemple, cherchent à rassembler les habitants nostalgiques de tel ou tel lieu. Le groupe *Facebook* représente alors l'occasion de se remémorer l'ancien temps, de partager des photographies d'archives personnelles, des informations factuelles (arrivée du tram, *etc.*) ou encore des anecdotes. Cette simplicité, non-structurée, se retrouve déjà dans le descriptif du « groupe public » *Auderghem d'autrefois*, qui tient en deux lignes : « Tout ce qui évoque le passé d'Auderghem et de ses environs proches est le bienvenu :-) ». Aucune « publication ancree » ne vient compléter cette brève description, aucune consigne n'est donnée aux personnes désireuses de participer. Le groupe compte 1387 membres, dont 10 « administrateurs ». Le fil d'actualité regroupe : des photographies postées seules, sans autre forme de commentaires ; des photographies comportant une date et/ou l'identification de lieux ou personnes ; des demandes d'informations accompagnées ou non d'une photographie ; des messages de bienvenue noyés dans le flux de nouvelles publications ; des requêtes concernant un document en particulier ; des demandes de confirmation concernant un nom de lieu, une adresse, un événement, *etc.* ; des messages de vœux pour les fêtes ; des devinettes ; *etc.*

Les réponses sont autant variées que les publications : exclamations ; hypothèses ; anecdotes personnelles ; informations factuelles ; appréciations subjectives ; demandes de précisions ; échanges à caractère personnel ; ajout d'une nouvelle photographie ; hyperliens renvoyant à une source externe ; *etc.*

Bien que la pertinence des publications semble très aléatoire, il est certain qu'un grand nombre d'informations et de souvenirs personnels est fréquemment publié sur ce groupe. Ce qui semble « fatal » dans cette gestion non-structurée, c'est que l'information est appelée à être enfouie sous des couches de nouvelles publications... Elle est ainsi amenée à « disparaître », sans qu'aucune institution n'en bénéficie ; alors que celles-ci, paradoxalement, doivent parfois se battre pour recruter de nouveaux contributeurs et stimuler la participation.

30. <http://culturevisuelle.org/dejavu/1097> (consulté le 10.05.15)

31. <https://www.facebook.com/groups/551916178254698/> (consulté le 14.05.15)

3.5 Analyse de tags et commentaires

Plusieurs chercheurs se sont déjà intéressés aux tags et commentaires publiés par les utilisateurs dans le cadre d'activités de *crowdsourcing* menées par des institutions culturelles.

Parmi les recherches qu'il semble intéressant de citer, il y a le projet *steve.museum*, qui porte sur l'usage du *social tagging* dans le contexte muséal (Trant, 2010). Derrière ce projet, il y a une volonté d'identifier les types de termes qui seront utiles à d'autres et ceux qui ne le seront pas. Les chercheurs étudient également les façons dont ces tags peuvent être validés : validation statistique automatique selon la récurrence du terme ou opération effectuée manuellement par des membres de l'institution ? Ils ont également mené des comparaisons intéressantes entre les métadonnées produites par des professionnels et celles produites dans le cadre du *crowdsourcing*. Sans véritable surprise, ils parviennent à la conclusion que la majorité des tags produits par le grand public ne correspondait pas aux tags encodés par le personnel des musées, mais qu'elle offrait une bonne représentation du rapport des individus aux collections³². Concernant ces réponses plus subjectives, Trant (2010) explique qu'elles méritent de s'y intéresser, afin d'évaluer leur utilité et de déterminer la façon dont elles peuvent être traitées. Enfin, ils montrent que seule une minorité d'utilisateurs contribuait à la majorité du contenu produit et qu'une grande majorité des utilisateurs n'avait participé au *tagging* que lors d'une unique session. Les études de cas menées en seconde partie permettront de vérifier cette tendance à l'aide de nouvelles données empiriques.

Quant à lui, van Hooland (2006) aborde la question des métadonnées générées par les utilisateurs en se penchant sur la qualité des commentaires postés. Afin de mesurer cette qualité, il propose une approche originale : observer la pertinence des commentaires par rapport aux besoins de la communauté des utilisateurs. Pour ce faire, il analyse un échantillon de 355 commentaires issus d'un projet mené par *The National Archives of the Netherlands*. Il propose une typologie composée de six catégories non-exclusives : correction de métadonnées préexistantes (46%) ; ajout de détails narratifs relatifs à l'image (31%) ; partage d'anecdotes personnelles (9%) ; mention d'une mauvaise mise à disposition de l'image (3%) ; expression d'une opinion ou d'un jugement (1%) ; interaction avec l'institution ou d'autres utilisateurs par le biais de question (1%). En outre, il observe que, tant les commentaires que les requêtes des utilisateurs sont majoritairement composés de termes spécifiques ; peu de termes génériques sont utilisés et quasiment aucune notion abstraite. Il parvient à la conclusion que le contenu des commentaires coïncide donc avec le type de recherches menées par

32. « Tagging is shown to provide a significantly different vocabulary than museum documentation : 86% of tags were not found in museum documentation. The vast majority of tags - 88.2% - were assessed as *Useful for searching by museum staff* (Trant, 2010) »

la communauté d'utilisateurs, remplissant ainsi le critère de « *fitness for purpose* », c'est-à-dire l'adéquation aux usages.

Actuellement, la pratique qui semble gagner en importance est celle de la gamification. Ce terme renvoie à « l'usage d'éléments de design caractéristiques des *games* dans des contextes non ludiques » (Deterding *et al.*, 2014). Flanagan travaille sur la question depuis de nombreuses années, elle codirige une plateforme de jeu en ligne (*Metadata Games*³³) fonctionnant essentiellement avec des logiciels libres et visant à aider les institutions culturelles dans l'enrichissement des métadonnées de leurs archives numérisées. La plateforme propose différents jeux qui visent à la création de tags. Il y a par exemple *Guess What!*, qui est un jeu collaboratif nécessitant la participation de deux joueurs. Le premier voit apparaître sur son écran une image qu'il doit décrire à l'autre joueur, qui voit lui-même 12 images sur son écran : il doit déterminer de laquelle il s'agit en se basant sur la description qui lui en est faite. Les conclusions auxquelles aboutit Flanagan ouvrent de nouvelles perspectives pour les institutions : une expérience ludique peut permettre à des joueurs qui n'étaient initialement pas attirés par le patrimoine culturel de s'y intéresser davantage et d'approfondir leur lien avec les collections et l'institution concernées (Owens, 2013).

Ridge (2011) s'est également intéressée à cette problématique. Sa démarche, à la fois théorique et empirique, a ceci d'original qu'elle vise à savoir s'il est possible de concevoir des formes de jeux permettant de créer des métadonnées plus complexes que de simples *tags*. Prenant l'exemple des collections des musées d'histoire naturelle, elle montre que certains artefacts sont plus difficiles à identifier que d'autres (trop similaires entre eux, visuellement non distincts, dont la signification est obscure, *etc.*) et imagine des alternatives aux *tags* pour les décrire. Ridge (2011) conçoit ainsi un modèle de jeu expérimental visant à la création de formes plus complexes de contenus. Cette expérience se conclue sur un résultat mitigé : elle constate qu'il n'existe encore aucun modèle de validation qualitative des réponses qui permettrait de gérer des données plus complexes que les *tags*.

Ces deux cas d'application font suite aux recherches qui avaient été préalablement menées par von Anh et Dabbish (2004). Ce sont eux qui avaient introduit le concept de *Game with a purpose (GWAP)* : « a game that is fun and can be used to create valuable output ». Le jeu développé par von Anh et Dabbish (2004), baptisé *The ESP game*, visait à étiqueter non pas uniquement le patrimoine culturel numérisé, mais toutes les images du web. Le but était de pallier l'inexactitude des termes associés aux images apparaissant dans les moteurs de recherche³⁴.

33. <http://www.metadatagames.org> (consulté le 29.10.14)

34. Le succès de ce jeu fut tel qu'il a d'ailleurs plus tard été acheté par Google, devenant *The Google Image Labeler*

Earle (2014) s'est penché sur les comportements des utilisateurs sur *The Flickr Commons*. Il base sa recherche sur un échantillon de 1000 tags et commentaires. Pour constituer cette typologie, il a combiné des caractéristiques qui sont plus ou moins présentes selon les catégories : *searchability; quantity; accuracy; depth; new information; exposure*. Il aboutit à une série composée de 10 catégories d'utilisateurs *hobbyist; experts; specialists; geocoders; proofers; uniformists; promoters; remixers; narrators; oddities*.

Par ailleurs, les comptes-rendus d'institutions ayant mis leurs collections sur *Flickr* constituent des témoignages précieux, renfermant des enseignements pertinents, bien souvent transposables à d'autres cas. Il semble dès lors intéressant de se pencher sur les synthèses émises par le personnel de différentes institutions (*The Library of Congress; The Powerhouse Museum; The National Library of New Zealand*).

En janvier 2008, la *Library of Congress* a participé au lancement du projet *The Flickr Commons* en mettant en ligne environ 3000 photographies d'archives. Dans ce rapport, Springer *et al.* (2008) tirent un bilan positif de ce projet pilote : ils relatent les échos favorables qui ont suivi cette mise en ligne, ainsi que les multiples retombées positives qui suivirent. Concernant les commentaires, ils notent que ceux-ci permettent aux personnes regardant les images de se remémorer des souvenirs, de partager des informations, de faire des liens entre passé et présent, d'identifier certains emplacements de façon très précise ou encore d'augmenter leur appropriation du contenu.

Au terme des premiers trois mois de présence du *Powerhouse Museum* sur *The Flickr Commons*, Chan (2008) fait différents constats. Tout d'abord, le rythme de publication annoncé (50 nouvelles photographies/semaine) n'a pas pu être respecté. Par ailleurs, la popularité (basée sur le nombre de vues ou le nombre de commentaires) des premières images publiées a influencé le choix des suivantes. Il met d'ailleurs en exergue l'implication quotidienne du personnel qui est nécessaire pour assurer le suivi de cette « participation continue ».

À l'issue des quatre premières semaines de vie du projet, les données statistiques sont éloquentes. Les photographies ajoutées sur *The Flickr Commons* ont été davantage vues en quatre semaines qu'en une année sur le site web. Selon Chan (2008), les images placées sur *Flickr* diffèrent de celles placées sur le site web de l'institution par leur caractère plus fluide et malléable. Cependant, concernant *The Flickr Commons*, il précise que la façon dont les individus considèrent les images et interagissent avec elles s'avère encore différente. Ce qui est particulièrement intéressant dans ce cas-ci, c'est que le *Powerhouse Museum* était déjà en possession d'un compte *Flickr* standard, qui sert ainsi de base comparative. Selon lui, les facteurs expliquant ces divergences proviennent de la nature historique des photographies utilisées sur *The Flickr Commons*, ainsi que du fait que l'institution n'est pas l'auteur des images dans ce cas-ci.

Les tags sont considérés par Chan (2008) comme « faciles à traiter » ; ils connaissent le même traitement que les autres métadonnées générées par la communauté, c'est-à-dire qu'ils sont intégrés aux données de la collection en ligne, sur le site web de l'institution, à côté des tags préexistants. Concernant les commentaires, il parle d'engagement communautaire à l'ancienne. Il souligne que l'apport (en informations, anecdotes, *etc.*) permis par ces commentaires constitue l'argument le plus percutant pour convaincre d'autres professionnels du secteur culturel de la valeur ajoutée des *Commons*. Cependant, l'intégration des commentaires est moins aisée, car cela nécessite des ressources pour trier et exploiter les informations inédites. Par ailleurs, il a remarqué qu'il est préférable d'éviter de minimiser ou de se moquer de commentaires pouvant sembler creux (par exemple : « *awesome photo* » ou encore « *cool picture* »), car ce type d'interaction a également son importance. Ces commentaires ont une fonction de liant social : il les assimile au café ou au siège confortable d'un musée.

Johnston (2008) publie un bilan sur le blog de la *National Library of New Zealand*, un an après l'ouverture d'un compte pro Flickr³⁵ par l'institution. Il ne s'agit donc pas, cette fois-ci, d'un compte faisant partie de *The Flickr Commons*. Elle explique quels étaient les objectifs de ce projet pilote : « *Attract users who do not know about our collections, or haven't thought about visiting our subsites. Observe the tagging and commenting behaviours to learn from the experience. Evaluate opportunities and issues for any further posting to Flickr.* » Les interrogations qui ont émergé furent notamment la question des droits d'auteurs, la taille des images, la gestion des demandes d'« amis », mais aussi la difficulté de mesurer le « succès » dans un tel contexte.

Johnston (2008) propose, plutôt que d'observer le nombre de vues ou de commentaires par photographie, de prêter attention aux dynamiques. Elle repère cinq grandes tendances : 1. L'action la plus commune est de mettre en favori une image, suivie par la publication de commentaires et enfin l'ajout de *tags*. 2. Les commentaires les plus fréquents rentrent dans la catégorie « *great photo* » et seule une faible proportion des commentaires comporte réellement des informations inédites. Aucun problème de spam n'a été constaté. 3. L'ajout de nouvelles images conduit à un pic de fréquentation. 4. La plupart des interactions se déroulent durant la nuit, ce qui suggère une large proportion d'utilisateurs issus d'autres zones géographiques. 5. La promotion de nouveaux ajouts par le biais de la *newsletter* se révèle efficace pour augmenter la visibilité des documents iconographiques.

Enfin, après avoir consacré en moyenne deux heures par semaine à ce projet (sélection, mise en ligne, réponse aux commentaires, observation des statistiques, *etc.*), Johnston (2008) résume en quelques lignes ce qu'elle a appris. Elle relève que le contact avec les utilisateurs sur *Flickr* est gratifiant, signale qu'il est nécessaire de se préoccuper

35. <https://www.flickr.com/photos/nationallibrarynz/> (consulté le 17.12.14)

des questions de droits d'auteur et permissions avant de procéder à l'*upload* des images et rend attentif au fait que certaines situations impliquent des questions éthiques. Comme par exemple lorsqu'un utilisateur a fait une demande pour ajouter une des images à un groupe de chasse à la baleine³⁶. En guise de conclusion, elle déclare que c'est un type de présence sur les réseaux sociaux qui nécessite d'investir moins de temps et d'énergie que la tenue d'un blog, par exemple.

Il faut noter que la publication de tags ou de commentaires par le grand public soulève également la question de la qualité des contributions. Dans le cadre de la présentation de leur typologie, Moirez *et al.* (2013) évoquent les limites du *crowdsourcing*. En effet, au-delà des nombreux avantages que ces opérations présentent (simplicité, gratuité, rapidité, adéquation avec le vocabulaire utilisé intuitivement par les utilisateurs dans leurs recherches en langage naturel, conformité aux usages du web, *etc.*), plusieurs difficultés peuvent émerger. Parmi ces limites, ils citent : la qualité de l'indexation qui ne peut rivaliser avec « une indexation professionnelle normalisée et appuyée sur des référentiels contrôlés », les questions de polysémie, d'orthographe, d'absence de hiérarchie, de personnalisation des vocabulaires, la difficulté de susciter la participation et d'obtenir une masse critique, une utilité très réduite si les métadonnées créées ne sont pas réintégrées ailleurs (catalogue par exemple), *etc.* Pour contourner certaines de ces limites Moirez *et al.* (2013) suggèrent de mettre en place des actions visant à favoriser la qualité et la fiabilité des contributions : formation des contributeurs (tutoriel, aide et conseils en ligne) ; évaluation de la compétence des contributeurs et soumission de documents dont le niveau est adapté ; corrections multiples (permettant de croiser les résultats) ; option permettant de signaler des erreurs ; vérification par des professionnels ou des experts ; réutilisation des résultats du *crowdsourcing* (par exemple en les réintégrant au catalogue). En outre, Boydens (2010) rend attentif à une réalité qui concerne également le *crowdsourcing* : l'absence de référentiel permettant de vérifier la correction d'une information.

L'état de l'art réalisé dans ce chapitre montre que le phénomène du *crowdsourcing* a été déjà exploré par de nombreux auteurs, selon des approches qui diffèrent énormément. La question plus spécifique des tags et commentaires a, elle aussi, fait l'objet de plusieurs publications. Ce mémoire vise à compléter ces recherches, en abordant plus en détail le contenu des commentaires que poste le grand public. En effet, si des typologies ont déjà été proposées par van Hooland (2006) et Earle (2014), il semble toutefois intéressant d'approfondir ces analyses, en multipliant les critères d'observations et en utilisant des échantillons de données issus de corpus variés.

36. Dans ce cas-ci, la conclusion a été que ce n'est pas le rôle de l'institution de porter des jugements « moraux » sur la façon dont les individus veulent utiliser des images de leurs collections.

Deuxième partie

Études de cas

Chapitre 4

Trois études de cas

Dans un article publié en 2006, Flyvbjerg (2006) revient sur les croyances relatives à l'usage d'études de cas dans le cadre de recherches en sciences sociales. Il démontre en quoi ces malentendus sont infondés et desservent de telles disciplines scientifiques, insistant sur l'effet positif que peut avoir le recours systématique à un grand nombre de bonnes études de cas. En outre, il propose d'intéressantes stratégies en ce qui concerne le choix d'études de cas et d'échantillons. Parmi celles-ci, *the information-oriented selection*, qui permet notamment de maximiser l'information provenant d'échantillons de petite taille. Détaillant cette stratégie, il mentionne quatre approches possibles : *Extrem/deviant cases* ; *Maximum variation cases* ; *Critical cases* ; *Paradigmatic cases*.

La deuxième approche, *Maximum variation cases* semble pertinente dans le contexte de ce mémoire. Il s'agit de choisir trois ou quatre cas qui se différencient radicalement les uns des autres au niveau d'une dimension donnée. Il peut s'agir de la taille, de la structure de l'organisation, d'un critère géographique, *etc.* La confrontation et l'analyse des données permettent ensuite d'obtenir des informations sur l'importance de ce critère. Cette étude comparative permet d'envisager l'objet de la recherche dans toute sa diversité et de tirer conclusions et enseignements sur une base empirique.

Les cas qui vont être étudiés dans le cadre de ce mémoire se distinguent par l'ampleur du projet de *crowdsourcing* mis en place (le volume d'images concernées varie fortement d'un cas à l'autre). Il s'agit de trois initiatives portées par des institutions culturelles d'Europe (France et Angleterre) et d'Océanie (Australie) : **A. Les Archives départementales de Vendée**, **B. *The English Heritage*** et **C. *The Powerhouse Museum***. Pour chacun de ces trois cas, un échantillon de commentaires d'utilisateurs sera analysé (*voir 5. Méthodologie*).

D'une dimension plus modeste, la première initiative est celle des Archives de la Vendée. Celles-ci proposent au sein d'une plateforme interne à leur site web d'identifier des scènes, personnages ou événements, dans le cadre d'un nombre limité d'« enquêtes » (actuellement, moins de 50) : il s'agit du *L@boratoire des internautes*. Vient ensuite *The*

Powerhouse Museum, qui a mis en ligne sur *The Flickr Commons* un peu plus de 2000 documents iconographiques numérisés. Enfin, *The English Heritage*, qui a développé une plateforme externe regroupant plus de 95.000 photographies aériennes numérisées : *Britain from Above*.

4.1 Les Archives départementales de Vendée (cas A)

Les Archives départementales de Vendée sont un service du Conseil général de la Vendée (Pays de la Loire, France), dont les missions sont de « conserver, collecter, classer et communiquer »¹. Pour l'année 2013, le budget annuel avoisinait les 300.000 euros et le personnel était constitué de 28 agents territoriaux/d'État². Les Archives conservent plus de 23 kilomètres d'archives et comptent plus de 4.2 millions d'images numériques de documents. Le site web, mis en ligne en 2003, rassemble près de 40.000 images, soit un peu plus de 60% de la somme totale des images numérisées par les Archives. En 2013, le nombre de connexions au site web s'élève à 1.200.210, nombre à mettre en perspective avec la fréquentation « hors ligne », le service d'Archives recensant 1012 lecteurs inscrits pour la même année³.

Les Archives départementales de Vendée mettent en place de nombreux projets participatifs qu'il semble intéressant de décrire dans un premier temps. L'institution encourage le partage et la numérisation d'archives personnelles pour enrichir la mémoire collective, notamment à travers sa « Grande Collecte 1914-1918 ». Elle invite également les visiteurs de son site web à signaler des archives inédites ou un événement marquant relatif à l'histoire de la Vendée. Le portail web permet également aux utilisateurs de prendre part à l'élaboration d'un vaste index nominatif (en se basant sur des registres paroissiaux, des recensements de population, *etc.*), afin de permettre des recherches sur base d'un nom ou d'un prénom. Un autre dispositif participatif regroupe trois types de dictionnaires collaboratifs : dictionnaire des Vendéens, dictionnaire historique des communes et enfin dictionnaire toponymique de la Vendée. Les chercheurs, qu'ils soient amateurs ou professionnels, peuvent contribuer à l'enrichissement de ces dictionnaires en complétant des articles ou en créant de nouvelles notices toponymiques, historiques ou biographiques. Enfin, *Le L@boratoire des internautes*⁴ permet, d'une part, d'échanger sur des questions historiques (proposer ou résoudre des énigmes, interroger la communauté, décrire ses méthodes de travail,

1. <http://archives.vendee.fr/Connaitre-les-Archives/Nos-missions/Presentation-des-Archives> (consulté le 10.05.15)

2. 24,3 ETP

3. <http://www.archivesdefrance.culture.gouv.fr/static/7915> (consulté le 10.04.15)

4. Aperçu de l'interface : voir Figure 4.1

etc.) et, d'autre part, d'identifier des documents iconographiques. C'est ce second volet, axé sur l'identification de documents, qui fera l'objet de la première étude de cas.

Les Archives de Vendée ont numérisé des fonds iconographiques riches et variés : cartes postales, photographies, cartes et plans, photographies aériennes, mais aussi dessins, gravures, *etc.* Les documents utilisés dans le cadre des « enquêtes » du *L@boratoire des internautes* sont majoritairement des reproductions de photographies argentiques, mais aussi des reproductions de carnets sténographiés, de peintures, *etc.*

Les billets publiés sur la plateforme présentent des images n'ayant pas encore été documentées, sous forme d'énigmes à résoudre collectivement, à l'aide de la section commentaires. Le nombre de photographies est relativement restreint (moins de 40 en décembre 2014). Ces billets font partie des différentes enquêtes proposées aux visiteurs du *L@boratoire*, il s'agit donc d'un dispositif de *crowdsourcing* qui s'intègre au sein d'un tout. Les billets reprenant les archives à identifier sont répartis entre trois rubriques : scènes/manifestations, personnes, édifices et monuments. On y accède de deux façons, soit en cliquant sur « lire la suite » sur la page d'accueil de la plateforme du *L@boratoire* (huit sujets sont mis en avant, classés par ordre chronologique, mais tous ne font pas partie du dispositif *Identifiez-les*), soit en cliquant sur le nom d'une des trois rubriques (scènes/manifestations, personnes, édifices et monuments) dans la barre de menu. Pour poster un commentaire, il faut obligatoirement saisir un nom, ainsi qu'une adresse email (qui ne sera quant à elle pas affichée).

Visiblement, l'institution n'a pas documenté ou analysé ce projet, qui s'insère au sein d'une action plus globale.

4.2 *The Powerhouse Museum (cas B)*

Le *Powerhouse Museum* est situé à Sydney (Australie). Figurant parmi les plus grands musées d'Australie, il possède des collections d'objets très variés, allant du design aux transports, en passant par la mode, l'histoire sociale, *etc.* Outre ses 22 expositions permanentes, le musée propose également des expositions temporaires ou virtuelles. Le *Powerhouse Museum* forme la branche principale du *Museum of Applied Arts & Sciences* de Sydney, qui inclut également *The Sydney Observatory* et *The Discovery Centre*. Tous départements confondus, environ 250 personnes travaillent au sein du *Powerhouse Museum*, dont le budget annuel s'élève à plus ou moins 45 millions de dollars australiens. Si l'on se réfère aux chiffres de 2012-2013, ce sont plus de 15.000 objets qui sont documentés sur le site web de l'institution, soit environ 60% des collections muséales. Cette collection en ligne a attiré sur une période d'un an plus de 800.000 visiteurs, soit 36.5% du trafic du site web (MAAS, 2013).

En avril 2008, le *Powerhouse Museum* commence à rendre visibles 200 photographies issues de ses collections de négatifs sur plaques de verre⁵, en les mettant en ligne sur *The Flickr Commons*^{6 7}. Ce fonds en ligne, fruit d'un effort collaboratif entre *Flickr* et *The Library of Congress*, héberge depuis 2008 les collections d'images (libres de droit) d'institutions culturelles du monde entier. L'ajout progressif de nouvelles photographies (également issues de la *Philips Collection* et de la *Clyde Engineering Collection, etc.*) accroît la collection qui rassemble aujourd'hui plus de 2500 photographies, toutes accompagnées de la mention « *No known copyright restrictions* ». Parfois, ce sont même des images inédites qui sont partagées, comme par exemple celles de la construction de la Maison d'Opéra de Sidney. Le nombre total de vues est supérieur à 5 millions.

Outre la galerie principale en page d'accueil, 49 albums thématiques ou encore deux sous-collections permettent de naviguer à travers le fonds photographique. Mettre en ligne ces images sur *Flickr : The Commons* permet aux internautes de les voir, de les diffuser, de les télécharger ou encore de les réutiliser (pratique qui est d'ailleurs encouragée par *The Flickr API*). Mais cette mise en ligne permet également aux utilisateurs inscrits sur *Flickr* d'avoir un rôle actif. Outre le tag, *Flickr* donne la possibilité aux internautes d'enrichir les métadonnées relatives à ces documents iconographiques en commentant ou encore en géolocalisant tel ou tel fichier⁸. Enfin, il est intéressant de relever que le musée intègre sur la base de données de ses collections en ligne les tags et données de géolocalisation provenant de *Flickr*, expliquant que les documents sont davantage tagués sur *Flickr* que sur son site web⁹.

Premier musée à rejoindre *The Flickr Commons*, le *Powerhouse Museum* a documenté cette expérience inédite sur différents blogs¹⁰. Chan (2008) explique par exemple comment l'utilisation d'une *API* fut l'occasion de partiellement automatiser la mise en ligne. Elle permet notamment d'envoyer les métadonnées existantes afin qu'elles apparaissent sur *Flickr*, mais aussi de garder un suivi lorsque l'image est commentée, etc. Cela a également conduit à la création de *mashup*, par exemple une application permettant de confronter des photographies historiques d'un lieu donné de Sidney à ce

5. Les photographies mises en ligne sont issues de la *Tyrrell Collection*, qui compte 7903 négatifs sur plaques de verre, provenant des studios de Charles Kerry (1857-1928) et de Henri King (1855-1923). Ces studios qui étaient basés à Sidney représentent une riche source de documentation sur la vie urbaine et rurale entre la fin du XVIII^e siècle et le début du XIX^e. La collection, qui avait préalablement été achetée par James R. Tyrrell, puis vendue à l'*Australian Consolidated Press*, fut donnée au *Powerhouse Museum* en 1985.

6. Le musée proposait déjà de suivre la vie du musée sur un autre compte *Flickr*, datant d'octobre 2007, qui propose des photographies contemporaines et non pas issues de collections (https://www.flickr.com/photos/powerhouse_museum_photography (consulté le 20.01.15))

7. Aperçu de l'interface : voir Figure 4.2

8. http://www.flickr.com/people/powerhouse_museum/ (consulté le 20.01.15)

9. <http://from.ph/28788> (consulté le 20.01.15)

10. <http://www.freshandnew.org/2008/07/commons-on-flickr-a-report-some-concepts-and-an-faq-the-first-3-months-from-the-powerhouse-museum/> (consulté le 20.01.15)
<http://www.powerhousemuseum.com/imageservices/page/1/?s=commons> (consulté le 20.01.15)

qui est donné à voir sur *Google Street View*. Les articles de blogs permettent également de faire connaître diverses initiatives dérivées, comme la publication d'un ouvrage¹¹ ou encore l'organisation d'une rencontre *In Real Life*¹².

4.3 *The English Heritage* (cas C)

La troisième étude de cas portera sur *Britain from Above*, projet lancé en 2011 par *The English Heritage*. *The English Heritage*¹³ est un organe du gouvernement britannique chargé de la gestion du patrimoine historique du pays. L'organisme compte plus de 2500 employés et affiche un budget annuel qui s'élève à plus de 160 millions de livres sterling¹⁴(Heritage (2014)).

The English Heritage poursuit diverses missions, notamment la préservation d'artéfacts historiques : 55% se rapportent à l'archéologie, 30% concernent des livres et fonds d'archives, 9% appartiennent aux arts décoratifs et enfin, un faible pourcentage se rapporte à l'histoire sociale et industrielle, à l'histoire naturelle ainsi qu'aux Beaux-Arts et à l'ethnographie. En outre, l'institution possède de vastes archives (près de 12 millions de documents), constituées de photographies et de cartes postales (79%), de dessins, plans ou matériels graphiques (5%), de dossiers et rapports (2%), de documents numériques (2%) et autres divers documents (12%). D'importantes opérations de numérisation d'archives préexistantes sont menées, dont le coût avoisinait les 550.000 livres sterling durant l'exercice 2013-2014. Ce processus s'étendant sur plusieurs années concernait deux archives acquises et capitalisées relativement récemment par l'institution : *The Architectural Red Boxes* et *The Aerofilms Collection*. C'est cette dernière qui est le point de départ de la troisième étude de cas.

Au coeur de ce troisième projet de *crowdsourcing*, une collection de photographies aériennes dépassant le million de négatifs et comptant plus de 2000 albums : *The Aerofilms Collection*. Un projet visant à documenter les versions numérisées de ces photographies voit le jour en 2011 sous le nom de *Britain from Above*. Le projet est financé par *The English Heritage*, en partenariat avec *The Royal Commission on Ancient and Historical Monuments of Scotland and Wales*. Le projet vise à conserver, numériser et cataloguer l'ensemble de ces négatifs datant de 1919 à 1953, ainsi que

11. Afin de fêter le premier anniversaire de la participation de l'institution et son millier de photographies déjà mises en ligne sur *The Commons*, un livre est édité : « *Then and now : stories from the Commons* » (<http://www.blurb.com/b/630596-then-and-now-stories-from-the-commons> (consulté le 20.01.15)). Dedicacé aux personnes impliquées dans ce projet, il présente des images historiques en intégrant des extraits de commentaires postés sur *Flickr*, mais aussi le nombre de vues, de commentaires, *etc.*

12. Dans le cadre des *Global Flickr Commons Meetup*, une rencontre a été organisée au sein du *Powerhouse Museum* en octobre 2009.

13. Officiellement : *The Historic Buildings and Monuments Commission for England*

14. En 2013-2014, 186.55 millions de revenus et 166.74 de dépenses.

d'autres photographies aériennes prises entre 1919 et 2006. Depuis 2012, une plateforme web dédiée¹⁵ invite chaque utilisateur inscrit à taguer, commenter ou géolocaliser l'une des 95.000 photographies aériennes numérisées (datant de 1919 à 1953).

Il s'agit d'un projet actuel et d'une certaine envergure, étant donné qu'il court sur quatre ans (il s'est officiellement clôturé à la fin de l'année 2014, mais les contributions sont encore possibles)¹⁶, qu'il mobilise un grand nombre d'employés, et que le fonds photographique est conséquent. En parallèle à cette opération de *crowdsourcing* en ligne, d'autres activités ont été planifiées, notamment une série d'expositions et d'ateliers mis en place à travers tout le pays à l'aide d'institutions partenaires et de volontaires.

Britain from Above a fait l'objet d'une présentation intitulée « *Britain from Above : User engagement between the digital and 'real' world* ». Brauer (2014), qui a pris part au projet en tant qu'*activity officer*, illustre comment la notion d'engagement de l'utilisateur a été abordée, tant en ligne que lors d'événements *in situ*, et résume les principaux enseignements qui ont pu être tirés de cette expérience. Elle explique par exemple que seuls 2% des images demeuraient non localisés, mais que cet élément fut primordial pour stimuler la participation et l'interaction des utilisateurs sur la plateforme en ligne. Brauer (2014) revient sur les stratégies en ligne et hors ligne qui ont permis d'atteindre de nouvelles audiences, tout en maintenant le lien avec des publics non connectés. Les « leçons apprises » concernent principalement la nécessité de connaître les utilisateurs et d'effectuer des activités et projets « pilotes » afin de pouvoir procéder à d'éventuels ajustements. Elle développe ensuite les avantages qu'ont permis la cohabitation de l'engagement en ligne et face-à-face (contributions plus nombreuses et de meilleure qualité sur le site web, rétention et motivation des utilisateurs, motivation du personnel et marketing de type « bouche-à-oreille »), souligne les nouveaux défis que pose la question du bénévolat virtuel et conclut en signalant que la motivation est la clé du succès du volontariat en ligne.

4.4 Des approches différentes

Examiner trois cas en particulier (pour une vue d'ensemble voir *Table 4.1*) permet d'avoir un aperçu des différentes façons dont l'institution peut procéder : du ton employé au choix des outils proposés. En ce qui concerne le cas A, l'enjeu de cet « outil participatif » est présenté sur la page « Accueil au public » du site des Archives :

Malgré nos recherches, des photographies ou des films issus des fonds d'archives restent mystérieux, alors que, documentés, ils pourraient retrouver tout leur sens

15. Aperçu de l'interface : voir Figure 4.3

16. <http://www.britainfromabove.org.uk/news> (consulté le 10.04.15)

et se révéler des sources historiques précieuses. [...] L'identification d'image est un moyen divertissant de nous aider à enrichir les descriptions des collections iconographiques et audiovisuelles.

Les gestionnaires du projet semblent considérer ce *L@boratoire* à la fois comme un outil d'enrichissement des archives et un dispositif ludique. En effet le nombre de photographies à identifier est très réduit¹⁷ et le corpus soumis aux commentaires semble faire l'objet d'une sélection répondant à une stratégie. Il n'y a pas de galerie avec des images « en vrac », seulement des billets présentant des énigmes à résoudre collectivement. Les titres se font interrogateurs : « Régiments d'artillerie ou d'infanterie ? », « Un soldat polonais ? », *etc.* Une personne travaillant pour les archives est là pour encourager les utilisateurs, synthétiser les nouvelles informations qui ont pu être trouvées et éventuellement recadrer la discussion. Lorsque des réponses pertinentes ont permis d'identifier les informations qui faisaient défaut, l'équipe des Archives déclare alors que le billet est « clos » (bien que sa consultation reste possible). L'institution est donc très présente, active, et continue finalement à exercer son autorité, en contrôlant et décidant qu'une notice documentaire est aboutie, ce qui contribue à entretenir le clivage opposant experts et amateurs. Il est probable que cette dimension très scolaire puisse susciter de l'incompréhension ou pis encore, s'avérer dissuasive pour certains participants, qui seront par exemple frustrés de ne pouvoir alimenter un billet « clos » avec les nouvelles découvertes qu'ils ont faites. En outre, insister sur la dimension ludique peut sembler réducteur lorsqu'il est question d'un engagement parfois très conséquent¹⁸.

Le caractère plus intimiste de ce premier cas tranche énormément avec le cas B, qui se déroule sur une plateforme préexistante très fréquentée et qui jouit d'une certaine visibilité grâce au « label » *The Flickr Commons*. Il s'agit ici d'une seconde configuration possible pour la mise en place d'activités de *crowdsourcing* par une institution culturelle. Le fait que l'opération de *crowdsourcing* ait lieu sur *Flickr* signifie que certains utilisateurs parcourant ces photographies le font peut-être de façon consciente et volontaire, mais ce n'est pas toujours le cas. En effet, cela peut également être le fruit du hasard, étant donné que ces images voisinent d'autres images n'émanant pas forcément du *Powerhouse Museum*. Cela peut conduire à des situations un peu absurdes voire cocasses, lorsque des utilisateurs suggèrent par exemple que la photographie (prise en 1900) aurait été mieux en couleurs... Au-delà de la pertinence toute relative des commentaires laissés, ce type de dispositif permet donc indéniablement de toucher de nouveaux publics. S'il agit, au contraire, d'un passage volontaire sur ce fonds d'images, il existe également plusieurs cas de figure : d'une

17. Moins de 40 en décembre 2014.

18. Propos inspirés d'un échange avec Patrick Peccatte, administrateur du projet *PhotosNormandie*.

part, il y a la population d'utilisateurs aguerris de *Flickr*, qui peut être vue comme une communauté possédant ses propres « codes », langages et pratiques. D'autre part, il peut s'agir d'un novice s'étant inscrit pour répondre à l'appel lancé par le musée sur sa page de profil *Flickr* : « We need your help - if you know more about the locations or people in these photos, or can help tag them, it would be great ! » Au-delà de cet appel à participation, la présence du musée est peu marquée. Ces premières observations révèlent qu'avoir recours à *Flickr* n'est donc pas un acte anodin : si la facilité de mise en œuvre pèse certainement dans la balance, il faut être conscient que cela draine un public particulier et que le projet prend place au sein d'un écosystème préexistant, qui aura un impact sur la nature des commentaires qui seront postés.

Enfin, troisième déclinaison possible avec le cas C : l'institution crée une plateforme web spécifique pour accueillir le dispositif d'enrichissement des métadonnées. Alors que le cas A propose uniquement de décrire les images en les commentant, *Britain from Above* permet aux visiteurs l'ajout de tag, de commentaire, d'annotation directement sur la photographie, de données de géolocalisation, de photographie ou vidéo personnelle, ou encore de contenu au sein du résumé factuel (wiki). L'objectif de la plateforme est double :

The project's website aims at making the Aerofilms collection available as well as encouraging the wider public to help interpret the collection. By adding layers of interpretation to the original image in the form of personal memories, factual information, photographs and other media, the website users significantly enhance the collection's accessibility for an audience unfamiliar with aerial photographs and thus widen its potential reach.

Cette fois-ci, l'enrichissement des données ne semble pas être considéré comme une fin en soi, mais plutôt comme un moyen de faire découvrir la collection et de permettre au public de s'appropriier ces photographies. La spécificité du corpus n'est sans doute pas étrangère à cette stratégie : il s'agit de plus de 95.000 photographies, qui représentent des portions de territoires vues du ciel. Donc, d'une part l'institution semble consciente qu'attendre de la foule qu'elle catalogue précisément ces 95.000 serait extrêmement ambitieux, pour ne pas dire irréaliste. D'autre part, vu qu'il s'agit de documents dont le contenu peut être assez répétitif (plusieurs vues à la suite d'un même lieu) et ne recelant pas toujours une grande part de « mystère » (s'il s'agit d'une vue de terrains agricoles par exemple, il n'y a pas forcément beaucoup à en dire), il serait absurde d'attendre une participation égale pour chaque document. Combiné au fait que certaines métadonnées sont parfois déjà connues (comme par exemple la date de prise de vue ou des données de localisations), ce caractère répétitif signifie une dimension de « quête informationnelle » parfois très peu stimulante pour les utilisateurs. Finalement,

les caractéristiques de ce corpus laissent donc penser que *The English Heritage* conçoit ce projet de *crowdsourcing* avant tout comme un dispositif de médiation¹⁹.

19. Cela se vérifie en prenant connaissance des multiples activités organisées *in situ*, comme par exemple ces ateliers à destination des publics préexistants mais victimes de la fracture numérique.



VENDÉE
LE DÉPARTEMENT

Retour au portail www.vendee.fr

Connaitre les Archives

Clés pour la recherche

Consulter

Découvrir

Participer

Les Archives de la Vendée 

L@boratoire des internautes

Jugez sur pièce
Débatte
Entraide
Identifiez-les
Forum des administrations

Recherche sur le L@boratoire

Rechercher

Vous êtes ici : Accueil > Identifiez-les > Edifices et monum... > **Un château vendéen ?**

[Retour à la liste](#)

● Tourisme en Charente et en Auvergne vers 1900
 Reconnaissez-vous ces lieux photographiés par un avocat vendéen dans les années 1920-1930 ? ●

Un château vendéen ?



J'aime 0
Twitter 0
4 commentaires

Cette carte postale de l'éditeur Eugène Poupin de Mortagne-sur-Sèvre a pour légende le château de la Bonnière à Mouchamps.
 Après vérification dans différents ouvrages et comparaison avec d' [autres cartes postales du château de Bonnière](#), il semble que la vue présentée ici ne soit pas le château de la Bonnière.

Reconnaissez-vous ce château ? Quel est son nom ? Dans quelle commune se trouve-t-il ? Sachant qu'il est fort probable qu'il ne soit pas non plus à Mouchamps !

Toute information sur l'histoire de ce lieu sera la bienvenue.

1 Num 20/2-153

Date de publication : 26 septembre 2014

Auteur du billet : Archives Vendée ;

Liste des commentaires

Le 26/09/2014 à 17h54,
le00h20 a écrit :

Il faut y voir le logis de La Bonnière sis commune de Saint-Michel-Mont-Mercure.

Le 27/09/2014 à 07h32,
Pénisson Dominique a écrit :

Tout à fait d'accord!
Allez sur <http://www.delcampe.net/page/item/id,151482512,var,85-SAINT-MICHEL-MONT-MERCURE-VIEUX-CASTEL-DE-LA-BONNELIERE.language,F.html> et vous y verrez le même cliché du même photographe avec la bonne légende. C'est un site de vente de cartes postales.

Le 27/09/2014 à 21h21,
Archives Vendée E a écrit :

Grand merci.
A vrai dire, maintenant qu'on sait que c'est la Bonnière, on voit que les Archives avait déjà publié cette photographie, dans une autre édition de carte postale. On y reconnaît bien le même personnage devant le château, debout et le poing sur la hanche.

Mettez "Bonnière" dans le moteur de recherche, choisissez les résultats des "inventaires et catalogues en ligne", puis, sous l'onglet de droite "iconographie numérisée" sélectionnez la ligne "cartes postales isolées".

Le 30/10/2014 à 18h10,
BELAUD Geneviève a écrit :

Il s'agit bien du château de La Bonnière à St-Michel-Mont-Mercure et j'ai un article concernant ce logis. Cet article est paru dans ALOUETTE HEBDO le 25 octobre 1984. Je pense aller fin novembre aux Archives, je peux emporter le document, document écrit par Jean LAGNAU.

Mots clés

- Art sacré (2) Auvergne (1) Bourneau (1) Canon (1) Charente (1) Commequiers (1) Communauté (1) Conseiller général (1) Etat civil (3) Filles de la Sagesse (1) Fontaines (1) Fontenay-le-Comte (3) Georges Clemenceau (2) Guerre 1914-1918 (6) Guerre 1939-1945 (2) Guerre de Vendée (4) Île de Noirmoutier (2) La Châtaigneraie (2) La Mothe-Achard (1) Le Tablier (1) Manifestation publique (2) Missionnaire (1) Moulin (3) Nouvelle-France (1) Officier (2) Onomastique (3) Paléographie (3) Pouzauges (1) Premier Empire (2) Presse (1) Réfugié (1) Saint-Malô-du-Bois (1) Saint-Michel-en-l'Herm (3) Sainte-Cécile (1) Santé (1) Vocabulaire (3)

Ephéméride

C'est arrivé un **18 avril**

1912 La région de La Roche-sur-Yon étant la plus favorable pour observer l'éclipse solaire prévue, les...

[Lire la suite](#)

A propos

Proposer une question

S'inscrire à notre lettre d'information

Noms de Vendée

Dictionnaires

FIGURE 4.1 – A. Archives de Vendée - Aperçu de l'interface

Vous Personnes Groupes Explorer Créer Importer

Powerhouse Museum + suivre

Exchange corner, Pitt St, Sydney

Format: Glass plate negative.

Rights Info: No known restrictions on publication.

Repository: Tyrrell Photographic Collection, Powerhouse Museum
www.powerhousemuseum.com/collection/database/collection=The_Tyrrell_Photographic

Part Of: Powerhouse Museum Collection

General information about the Powerhouse Museum Collection is available at www.powerhousemuseum.com/collection/database

Persistent URL:
www.powerhousemuseum.com/collection/database?irn=28990

Acquisition credit line: Gift of Australian Consolidated Press under the Taxation Incentives for the Arts Scheme, 1985

As the crow flies, calder20002004 et James Douglas ont ajouté cet élément dans les favoris

beachcomber australia 6 a
 Corner of Pitt and Bridge streets, looking south.

Peter 5 a
 Hi, I'm an admin for a group called Sydney's older buildings and structures. , and we'd love to have this added to the group!

beachcomber australia 5 a
 Charles Bayliss - acms.sl.nsw.gov.au/item/itemLarge.aspx?itemID=413535

5,600 vues 3 favoris 3 commentaires Prise environ 1900
 Aucune restriction de droits d'auteur connue

Afficher l'EXIF

Cette photo figure dans 2 albums

- Tyrrell Collection 1005 photos
- Sydney City 137 photos

Cette photo ne figure actuellement dans aucune expo
 Ajouter à l'expo

Tags Ajouter des tags

- Powerhouse Museum
- Pitt St Sydney
- Australia buildings
- man exchange corner
- Glass plate negative
- Tyrrell horse carriage

FIGURE 4.2 – B. Powerhouse Museum - Aperçu de l'interface



Image details
Padgate Camp (No.3 RAF Depot), Padgate, from the south-east, 1946

Love this (0) Add a pin - (hide pins) Add to groups Add to profile Download Buy

Like Follow Tweet

Image details & comments Factual summary Latest activity

Title Padgate Camp (No.3 RAF Depot), Padgate, from the south-east, 1946
Image reference EAW001093
Date 3rd June 1946

Tags Balloon depot, Camp, RAF
Add tags

Place name Padgate
Parish POULTON-WITH-FEARNHEAD
Unitary authority WARRINGTON
Administrative area Warrington
Country England
Easting/Northing 362576, 390245
NGR SJ626902

Please sign in/register to post comments and images.

Balloon sheds

Sign in to reply
report abuse



Graham Yaxley
Tuesday 6th
January
10:15pm

This appears to be Padgate and part of what is now Cinnamon Brow in Warrington

Sign in to reply
report abuse



RingoDave
Thursday 5th
September
11:38pm

Padgate Camp, Warrington ?

Sign in to reply
report abuse



Andrew Bennett
Monday 26th
August
12:32am

Hi Andrew,

Thanks for your suggestion! We'll update the catalogue and post your new data here in due course.

Yours, Katy
Britain from Above Cataloguing Team Leader

report abuse



Katy Whitaker
Tuesday 3rd
September
10:58am

FIGURE 4.3 – C. Britain from Above - Aperçu de l'interface

	A.	B.	C.
Nature du corpus	Archives : photos anciennes de personnes et événements se déroulant en Vendée	Archives : fonds de photos portant sur Sidney et ses alentours	Archives : photos aériennes de la Grande-Bretagne, prises entre 1919 et 1953)
Nbre total de documents iconogr.	36	2527	95 551
Nbre total de commentaires	217	3685	16 216
Type plateforme	Pl. interne	Pl. externe préexistante	Pl. externe dédiée
Durée du projet	Du 21.06.2011 à Aujourd'hui	Du 07.04.2008 à ? (pas de nouvel ajout de photo depuis 1 an)	Officiellement 2011-2014, mais site (lancé le 25.06.2012) encore opérationnel
Public visés	Grand public, en particulier pop. vendéenne	Grand public, notamment les utilisateurs de <i>Flickr</i>	Population de Grande-Bretagne, notamment le public peu connecté
Inscription	Sans inscription, mais nom et adresse email obligatoires pour poster	Inscription (compte <i>Flickr</i>)	Inscription (compte avec page de profil)
Modes de contribution	Commentaires	Commentaires, tags, géolocalisation	Commentaires, tags, wiki, annotations
Vocabulaire libre/contrôlé	Libre	Libre	Libre
Validation manuelle/automatique	Automatique (modération a posteriori)	Automatique	Automatique
Animation et suivi ?	Oui	Non, pas récemment du moins	Oui
Lieu de destination des métadonnées	Site web	Site web, section séparée	-
Dimension sociale	Faible	Très forte (communauté <i>Flickr</i>)	Encouragée (page de profil, groupes,)
Actions <i>In Real Life</i> ?	-	Publication d'un livre, rencontres	Nombreux ateliers et expositions

TABLE 4.1 – Vue d'ensemble des trois études de cas

Chapitre 5

Méthodologie

5.1 Objet d'étude

La méthodologie, basée sur l'étude de trois cas, permet de se pencher sur l'enrichissement des métadonnées par le grand public. Afin de pouvoir approfondir l'exploration du contenu créé par les utilisateurs, seul un des modes de contribution proposés aux utilisateurs a été pris en compte : la publication de commentaires. Ceci exclut donc d'autres types de participation comme l'ajout de *tags* par exemple.

Par ailleurs, de nombreux commentaires étant postés par des membres de l'institution à l'origine du projet, il a fallu déterminer si ce type de commentaires avait sa place dans cette analyse. Finalement, vu qu'il s'agissait de commentaires portant directement sur le contenu, amenant parfois de nouvelles informations, vecteurs d'interactions et pleinement liés au processus de production de métadonnées, ils n'ont pas été exclus.

L'exploration d'un corpus de commentaires permet par exemple d'observer si des dynamiques de création collaborative s'élaborent, si les utilisateurs font preuve de rigueur scientifique et témoignent d'une certaine expertise ou s'il s'agit majoritairement de commentaires à caractère subjectif. Analyser ces commentaires de façon systématique permet d'amener des éléments de réponse objectifs concernant les caractéristiques, forces et faiblesses de ces commentaires. Il sera également possible d'esquisser les facteurs pouvant expliquer des écarts ou convergences entre les données et, enfin, de tirer des enseignements plus généraux sur les possibilités et limites de la construction collective de métadonnées.

5.2 Collecte des données

Si les trois cas proposent aux utilisateurs de commenter le contenu de collections iconographiques numérisées, les interfaces sont cependant différentes. Ce fut l'occasion

d'expérimenter trois types d'extraction de données : l'extraction manuelle, le recours au *data scraping* et enfin, l'utilisation de l'*API Flickr*.

En raison de la taille restreinte du corpus du L@boratoire des Internautes et de l'agencement particulier du dispositif de commentaires, il semblait plus simple de procéder manuellement. La structure est similaire à celle d'un blog : chaque billet renvoie à un contenu dont on ignore certains éléments : date, identité des personnes représentées, localisation, signification, *etc.* Certains billets présentent exceptionnellement une galerie de plusieurs images, par exemple des cartes postales de plusieurs moulins de la région. Les commentaires s'affichent en dessous du billet, par ordre chronologique. Toutes les pages ont été visitées de façon systématique entre le 26 novembre et le 4 décembre 2014 afin de recenser dans un fichier Excel la totalité des commentaires publiés par les utilisateurs entre le 21 juin 2011 et le 7 novembre 2014. Outre le texte des commentaires (langage naturel non contrôlé) récupéré par la technique du « copié-collé », d'autres informations potentiellement intéressantes ont été inventoriées sur la même ligne du fichier : section (Scènes-manifestations / Personnes / Édifices et monuments), titre du billet, questions adressées aux internautes, statut du billet (en cours/clos), date de publication du billet, *URL* du billet, auteur du commentaire, heure et date de publication du commentaire. En outre, une colonne additionnelle intitulée « Intérêt » permettait de spécifier si le commentaire se distinguait des autres par l'un ou l'autre élément (afin de pouvoir le retrouver plus facilement ensuite). À l'issue de cette étape, le corpus comptait 217 commentaires, réponses de personnes travaillant pour les Archives incluses.

Le *Powerhouse Museum* ayant investi *The Flickr Commons*, il est possible de recourir à l'*API (Application Programming Interface)* de la plateforme de partage pour automatiser la collecte de commentaires. La réutilisation d'un script *Python* a permis de lancer des « appels » à cette *API* afin d'extraire dans un fichier XML toutes les données utiles. Après l'obtention d'une clé d'authentification permettant d'utiliser l'*API*, la deuxième étape nécessitait de connaître les *ID* concernés, c'est-à-dire les nombres permettant d'identifier les « collections » mises en ligne par le *Powerhouse Museum*. Ensuite, une fois ces données encodées, une commande a permis d'exécuter le script. Cette étape de nature très technique a été réalisée selon un processus de type itératif (essais-erreurs), qui a nécessité le soutien d'une tierce personne. Au vu du nombre de fichiers et de commentaires concernés, le recours à cette *API* semble judicieux pour réaliser une telle extraction. Les données, publiées par les utilisateurs entre le 8 avril 2008 et le 17 janvier 2015, ont été extraites dans un fichier XML le 18 janvier 2015. Ce fichier a ensuite été traité à l'aide du logiciel *OpenRefine (Interactive*

Data Transformation Tool), une étape de « nettoyage » des données étant nécessaire¹. Cette opération a permis de traiter de façon automatisée les doublons ou cellules vides que comptaient les jeux de données. Dans le cas des données du *Powerhouse Museum*, un important problème de doublons a été constaté². Il a fallu les identifier et les supprimer, en prenant soin de vérifier s'il s'agissait de véritables doublons, ou seulement de contenus similaires rédigés par des personnes différentes, comme par exemple : « Nice! », « :) », « !!! », *etc.*. En ordonnant par « date » les doublons annoncés par le logiciel, il a été possible de supprimer uniquement ceux qui avaient été postés à la même heure et dont l'*URL* de la photographie correspondait. Au terme de cette opération réalisée manuellement, le corpus qui rassemblait 4651 commentaires au départ est passé à 3685.

La plateforme web *Britain from Above* n'a prévu aucune action en matière d'ouverture de ses données. En l'absence d'*API*, il est possible de recourir à une autre technique pour extraire un nombre conséquent de données : le *data scraping*. Cette technique consiste à utiliser un script qui permet de répéter de façon automatisée des mêmes actions sur des centaines ou des milliers de pages web. Par exemple, il est possible d'indiquer au script que la machine doit « visiter » toute une série de pages dont la structure est la même, et sur celles-ci cliquer sur une certaine zone de texte comportant un hyperlien et, après avoir suivi cet hyperlien, copier ces zones de données et les extraire dans un fichier *XML*. Cette opération de *data scraping* qui nécessite des savoir-faire techniques conséquents a été réalisée par une tierce personne. Les données publiées par les utilisateurs entre le 17 mai 2012 et le 4 février 2015 ont été récoltées le 30 décembre 2014 (fiches individuelles), ainsi que du 2 au 5 février 2015 (commentaires), en recourant au logiciel *OutWib Hub Pro*³. La première partie du *scraping* (fiches individuelles) a été effectuée de manière automatique. Le logiciel a reçu pour consigne de partir d'une page du site présentant les résultats lorsque le champ de recherche ne comportait aucun caractère, ce qui correspondait donc à un classement des pages de l'entièreté des photographies hébergées sur le site (96.218)⁴. En partant de cette page, le logiciel devait récolter sous forme de tableau le texte et les liens hypertextes présents dans la page, puis aller à la suivante et ainsi de suite. Les quelques 96000

1. Ce logiciel permet de travailler sur des vastes lots de données, c'est-à-dire qu'il permet de les réorganiser, de traquer les doublons, d'utiliser des filtres, *etc.* *OpenRefine* se présente sous la forme d'une interface intuitive très puissante, dont le graphisme est similaire à celui des tableurs comme Excel.

2. Deux « appels » différents ont dû être lancés afin de couvrir les deux « collections » que propose le musée sur Flickr, ceci dans le but d'obtenir le plus grand nombre possible de commentaires pour ce cas, dans un souci de représentativité. Cependant, il est apparu que certaines photographies (avec les mêmes commentaires) étaient présentes dans les deux collections, ce qui a entraîné la présence de doublons indésirables lorsque les deux fichiers *XML* ont été rassemblés en un seul.

3. <http://www.outwit.com/products/hub/> (consulté le 10.05.15)

4. <http://www.britainfromabove.org.uk/asearch?search&order=4&view=list> (consulté le 10.05.15)

URL récoltées lors de la première phase ont ensuite été explorées une par une, afin d'extraire les commentaires de chaque photo. La difficulté était d'obtenir uniquement le contenu des commentaires et non pas des tags ou des textes relatifs à des annotations sur la carte, afin de préserver une certaine cohérence avec les deux autres études de cas. Les tags HTML qui entourent les commentaires ainsi que des expressions régulières ont permis de contourner ce problème, afin de n'extraire que les commentaires ne contenant ni photographie ni icône indiquant qu'il s'agissait d'une annotation. Après élimination des doublons (même commentaire, du même utilisateur, sur la même photo, à la même minute), il reste 16 216 commentaires. Il semblait intéressant d'avoir également l'année de publication, qui n'était pas explicitement indiquée, ce qui a nécessité une opération supplémentaire⁵.

5.3 Constitution des échantillons représentatifs

Des dispositifs en ligne⁶ permettent de calculer facilement la taille d'un échantillon pour que celui-ci reflète de façon optimale la population cible. Le calcul de la taille de l'échantillon a été réalisé en utilisant un niveau de confiance de 95% et un intervalle de confiance de 5%. Le nombre total de commentaires par corpus peut ensuite être divisé par cette somme. En partant du quotient résultant, il est dès lors possible de sélectionner une entrée à intervalle régulier afin d'obtenir un échantillon représentatif de l'entièreté du corpus.

Pour ce qui est du *L@boratoire des Internautes*, le nombre de commentaires s'élève à 217. En partant de cette somme et en utilisant un niveau de confiance de 95% et un intervalle de confiance de 5%, la taille de l'échantillon requis sera de 136 (pour un total de 217). Cependant, le quotient étant de 1.56 et une ligne ne pouvant être divisée que par des entiers, il a été jugé plus opportun de garder deux commentaires sur trois, ce qui conduit à un échantillon de 145 commentaires (au lieu de 136). Cette action a été réalisée à l'aide du logiciel Excel en masquant un commentaire sur trois.

Concernant la deuxième étude de cas, le corpus de commentaires issus des deux « collections » d'albums mis en ligne par *The Powerhouse Museum* sur *Flickr : The Commons* regroupe 3685 commentaires. En gardant les mêmes paramètres que dans

5. Le site indique pour chaque commentaire l'heure, le jour de la semaine et le mois auxquels il a été posté, mais pas l'année. Sachant que le lancement de la plateforme remonte à 2012 et que tous les commentaires ont donc été rédigés entre 2012 et 2015, l'association jour de la semaine-jour-mois permet de retrouver l'année (ce qui n'aurait pas été le cas si le site avait plus de six ans d'âge). Ce constat a permis de tester avec succès une manœuvre permettant de comparer l'année avec le résultat renvoyé par la formule Excel JOURSEM(). Cette opération a permis de déduire l'année des commentaires, ce qui rend possible des observations telles que nombre de commentaires postés par année.

6. <http://www.surveysystem.com/sscalc.htm> (consulté le 01.05.15)

le premier cas (niveau de confiance : 95%, intervalle de confiance : 5%), la taille de l'échantillon requis est de 348. Ce qui correspond à un quotient de 10.58. Ce quotient est arrondi à l'entier inférieur afin de permettre une division des lignes, ce qui implique de garder un commentaire toutes les 10 lignes (à l'aide de l'option « masquer » sur Excel). Au final, ce sont 370 commentaires qui seront analysés.

Enfin, les commentaires extraits de la plateforme web *Britain from Above* sont au nombre de 16 216. La taille de l'échantillon requis est de 375 si, à l'instar des deux autres études de cas, nous utilisons un niveau de confiance de 95% et un intervalle de confiance de 5%. Le quotient (de 43.24) est à nouveau arrondi, un commentaire est donc gardé toutes les 43 lignes, ce qui conduit à un total de 377 commentaires. Cette opération a été réalisée par une tierce personne à l'aide d'une commande⁷ qui a permis de « couper » le corpus automatiquement.

5.4 Traitement des données

Après avoir collecté les données, la seconde étape a consisté en l'analyse manuelle du contenu des 892 commentaires. Ce processus s'est organisé autour de différentes tâches :

1. Lecture préliminaire de plusieurs centaines de commentaires issus de chacun des cas ;
2. Constitution de variables permettant d'analyser le contenu des commentaires ;
3. Mise à l'épreuve et amélioration des variables prévues sur base d'un échantillon de 80 commentaires ;
4. Séparation de l'analyse en deux parties dans un souci de clarté : (a) Tableau comparatif de données purement quantitatives (b) Analyse qualitative plus approfondie ;
5. Constitution du premier tableau (quantitatif) : passage de nombres entiers à des pourcentages ;
6. Remplissage du second tableau (qualitatif) : visite de la page web concernée ; survol de l'image ; survol des commentaires voisins ; encodage des données ;
7. Constitution d'un tableau récapitulatif reprenant les totaux (nombre de commentaires concernés) des 3 cas ;
8. Création de graphiques permettant d'illustrer les résultats (passage de nombres entiers à des pourcentages).

7. Il s'agit d'un script *PowerShell* sur *Windows*.

Chapitre 6

Analyse des données

6.1 Comparaison de données quantitatives

Étant donné que la taille des corpus et des échantillons varie (145, 370 et 377 commentaires analysés), certains nombres sont exprimés sous forme de pourcentages¹ afin d'être plus aisément comparables.

Ce premier survol des trois cas par le biais de statistiques² (voir *Table 6.1*) permet déjà de faire certains constats. En effet, ces sommes, moyennes et pourcentages viennent mettre en avant certains phénomènes relatifs à l'ampleur du projet ou au type de plateforme choisi par l'institution.

	A.	B.	C.
Taille échantillon	145	370	377
Nbre total de commentaires	217	3685	16 216
Pourcent. d'images commentées (total)	86.1%	20.7%	5.8%
Moyenne de caractères par comm. (total)	562	217	139
Pourcent. de comm. en langues étrangères/échantillon	0%	7%	0%
Nbre de comm. en moyenne par image commentée	8.3	7.1	1.6
Nbre maximal de comm. par image	31	318	77
Nbre total de contributeurs	56	1974	3109
Nbre maximal de comm. du même utilisateur	47	463	3153
Pourcent. de contributeurs uniques/total contributeurs	32.1%	82.8%	56.4%
Pourcent. de comm. postés par des contributeurs uniques	8.2%	44.3%	10.8%

TABLE 6.1 – Survol des données, approche quantitative | Trois études de cas

1. Images commentées/total : A. 31/36 ; B. 522/2527 ; C. 5496/95 551

Langues étrangères : A. 0/145 ; B. 26/370 ; 0/377

Nbre de contributeurs uniques/total : A. 18/56 ; B. 1634/1974 ; C. 1752/3109

Proportion de commentaires postés par des contributeurs uniques : A. 18/217 ; B. 1634/3685 ; C. 1752/16 216

2. Les calculs ont été effectués à l'aide du logiciel *Excel* (en utilisant notamment les tableaux croisés dynamiques) ainsi qu'à l'aide du logiciel *OpenRefine*.

Tout d'abord, il est intéressant de confronter le pourcentage d'images commentées avec la taille du corpus : il apparaît clairement que moins il y a d'images, plus la proportion d'images commentées sera importante. Ainsi, le cas A atteint un pourcentage extrêmement élevé, avec plus de 80% des images qui sont commentées ; pourcentage qui peut presque être divisé par 10 pour le cas C (8.8%). Bien que d'autres facteurs puissent également être à l'origine de ces résultats, comme par exemple la coexistence d'autres modes de contributions (tags, wikis, *etc.*), avoir conscience de cette tendance générale pourrait s'avérer utile pour une institution désireuse de concevoir un projet de *crowdsourcing*.

Deuxièmement, le logiciel *OpenRefine* a permis de calculer la moyenne du nombre de caractères³ par commentaire : 562 (A), 217 (B) et 139 (C). Couplée au pourcentage d'images commentées, cette moyenne est un bon indicateur du type de contributions que produit un utilisateur lambda pour chacun des cas. Ainsi, la taille modeste du premier cas semble favoriser des descriptions plus étoffées (562 caractères en moyenne), tandis que le cas C est le plus en retrait pour ce critère-là (139). Si cette longueur n'est pas forcément gage de qualité, elle fournit cependant un premier indice sur l'investissement du contributeur.

Concernant les commentaires en langue étrangère, ils sont à mettre en relation avec la langue de la plateforme. En effet, il n'est pas très étonnant que les échantillons des cas A et C ne comptent aucun commentaire en langue étrangère, sachant que le site des Archives de Vendée n'est disponible qu'en français et que la plateforme *Britain from Above* propose comme seules options l'anglais et le gallois. En revanche, il n'est pas rare de trouver des commentaires en espagnol, portugais, chinois, *etc.* sur la plateforme multilingue *Flickr*.

Il s'agissait également de calculer combien de commentaires sont publiés en moyenne lorsqu'une image est commentée. Si les moyennes sont relativement similaires pour les cas A et B (8.3 et 7.1), en revanche, la moyenne très basse correspondant au cas C (1.6) retient l'attention. Cette faible moyenne renforce la tendance déjà observée précédemment : alors que seules 5.6% des images du cas C sont commentées s'ajoute le fait que, lorsqu'elles le sont, le nombre de commentaires n'est pas très élevé.

D'autres informations peuvent encore être obtenues en observant la répartition des commentaires. Ainsi, l'écart très marqué entre le nombre maximal de commentaires par image des cas A et C par rapport au cas B traduit un phénomène spécifique à *Flickr*. En effet, il apparaît que certaines photographies particulièrement populaires bénéficient d'une visibilité spécifique (par exemple en étant ajoutées à divers groupes *Flickr*) et suscitent de nombreux commentaires de type « appréciations subjectives ».

3. Espaces compris.

D'autre part, il semblait judicieux de calculer quel était le nombre total de contributeurs différents. Les sommes semblent en cohérence avec la taille des projets concernés (56⁴, 1974, 3109). Il faut toutefois noter que l'écart entre les cas B et C n'est absolument pas proportionnel à l'écart qui sépare la taille de leurs collections d'images destinées à être commentées (2527 pour le cas B ; 95 551 pour le cas C).

Les résultats les plus frappants concernent le nombre maximal de commentaires rédigés par le même utilisateur. Les sommes très élevées attestent la présence de « super contributeurs » qui rédigent parfois jusqu'à plus d'un cinquième de tous les commentaires qui sont publiés ! Pour le cas A, cet utilisateur a posté 47 commentaires, soit plus de 21% des commentaires. Cependant, c'est un cas un peu particulier, vu que l'utilisateur est *Archives Vendée*, c'est-à-dire le ou les membre(s) de l'équipe des Archives qui se charge(nt) de répondre et d'animer la discussion. Pour le cas B, il s'agit de *beachcomber australia*, qui a posté 463 commentaires, soit plus de 12%. Le même type de phénomène touche le cas C : *class31* est à l'origine de 3109 commentaires, soit presque 20% du total des commentaires. Sachant qu'ils sont souvent plusieurs à se partager ce statut de « super contributeur », ces nombres pourraient conduire à relativiser le succès de ces projets. Si l'on supprimait les commentaires postés par les trois plus grands contributeurs de chaque projet, combien en resterait-il ? Il ne s'agit en aucun cas de remettre en question les contributions de ces utilisateurs, dont la participation nourrit et stimule le processus collectif, créant un véritable cercle vertueux. En revanche il convient de se demander s'il est pertinent de les comptabiliser de manière indifférenciée au sein des productions de la « foule ».

Enfin, d'autres statistiques concernent l'antithèse du « super contributeur », c'est-à-dire l'utilisateur de passage, celui qui ne va contribuer qu'une seule et unique fois. Pour le cas A, ils représentent un peu plus d'un tiers du total des contributeurs, pour le cas C il s'agit de plus de la moitié des contributeurs. Pour le cas B, le pourcentage atteint des sommets : plus de 82.8%, soit plus de 4 utilisateurs sur 5 ! Cela révèle la spécificité des plateformes : le cas A favorise une participation plus assidue, le cas B est davantage un lieu de passage et le cas C combine ces deux types de participants de façon plus ou moins équivalente. Ces premières données ont ensuite permis de calculer à combien s'élève le pourcentage de commentaires postés par ces contributeurs uniques : 8.2%, 44.3% et 10%. « Retourner » ces données permet en outre de constater que pour le cas B, près d'un cinquième des contributeurs (18.2%) qui ont posté plus d'une fois sont à l'origine de plus de la moitié des commentaires. Il semble en effet important que les institutions culturelles puissent être au courant de telles tendances, de manière à pouvoir éventuellement adapter leur stratégie : faut-il travailler sur la notion de

4. Il faut noter que les commentaires postés par « anonyme » ont été considérés comme émanant d'auteurs différents (bien que cela n'ait pas pu être vérifié). Sans cela, il n'y aurait pas eu 56 contributeurs différents mais 52.

fidélisation, ou faut-il se réjouir de ce passage et donc de la visibilité croissante dont jouit la collection ?

Ces premières données statistiques permettent de préciser le profil de chacun de ces projets de *crowdsourcing*. Le cas A, avec son plus petit corpus de documents, sa dimension d'« enquêtes » et sa plateforme interne à son site web, favorise la fidélité des contributeurs et la publication de longs commentaires, qui portent sur près de 90% des images. Il s'agit d'un public plus ciblé que celui du cas B qui, lui, se démarque par le très haut taux de contributeurs uniques. Les utilisateurs s'expriment dans différentes langues et, plateforme externe préexistante oblige (*Flickr*), il est plausible qu'ils n'aient pas forcément décidé consciemment de découvrir cette collection en particulier. Enfin, la plateforme dédiée du cas C présente un corpus d'une telle ampleur que seule une faible proportion d'images a fait l'objet de commentaires et ceux-ci se distinguent d'ailleurs par leur brièveté.

6.2 Outil d'analyse qualitative

Afin de mieux comprendre les apports des utilisateurs contribuant à des projets de *crowdsourcing*, différents paramètres peuvent être pris en compte. Outre les aspects purement quantitatifs d'ores et déjà passés en revue, une recherche qualitative permet d'inscrire ce mémoire dans une démarche inductive, visant à comprendre les phénomènes à l'œuvre. Pour ce faire, une grille d'analyse originale se basant sur huit angles d'approche a été élaborée⁵. Ceux-ci étant eux-mêmes subdivisés en diverses variables, ce sont finalement 25 paramètres qui vont permettre une démarche comparative. Cette analyse servira de support à une observation des ressemblances et différences entre les trois cas étudiés.

Les huit angles d'approche sont : les apports de type personnel, la question de la langue, le type d'expertise mobilisée, tout ce qui est relatif aux interactions humaines, l'information, les données « méta » et enfin tout ce qui relève du « spam ». Il s'agit donc de catégories relativement générales, elles-mêmes détaillées par le biais de paramètres plus concrets et objectifs.

Ces paramètres, non exclusifs, devraient permettre, pour chaque commentaire donné, de répondre par l'affirmative ou par la négative à la question : « Ce commentaire présente-t-il cette caractéristique ? ».

5. Le nombre très élevé de données ne permet pas de reprendre toutes ces informations ici. Un aperçu de la grille d'analyse utilisée est toutefois repris en annexe *Voir Annexe B, Figures B.1 et B.2*

Ci-dessous, une description concise et un exemple⁶ illustrant chacun des 25 paramètres :

Premier angle d'approche : **Apports personnels**

— 1. Appréciation subjective

L'utilisateur partage, de façon synthétique ou avec emphase, son ressenti face au document, qu'il soit positif ou négatif.

> « *wow that is awesome!* » *Randy806 [E3]*

— 2. Expression d'avis personnel, de convictions

L'utilisateur exprime son opinion par rapport à des sujets évoqués par l'image ou par d'autres utilisateurs. Il s'agit d'éléments subjectifs qui se distinguent de l'apport de nouvelles informations.

> « *Absolutely amazing! Too bad our society does not look like this anymore. I don't like modern architecture.* » *Tyler Gillihan [E4]*

— 3. Anecdotes personnelles

L'utilisateur relate des anecdotes à caractère personnel ; il peut s'agir du récit de ses démarches pour documenter l'image, de souvenirs personnels, d'histoires familiales, *etc.*

> « *My (maternal) grandparents lived at No 1 Park Lane which was then a ham shop. After the start of WWII hams we [sic] not available so they turned the shop into bar (and as Steve Gregory as already mentioned, the fountain was removed). In 1946 (approx.) the film So Well Remembered was made in and around Park Green, and my grandparents provided tea and cakes to the star John Mills and his wife Mary.* » *Maurice Mottershead [E5]*

Second angle d'approche : **Langage**

— 4. Mot(s) isolé(s)

L'utilisateur recourt à un mot ou groupes de mots isolés pour s'exprimer ; il n'utilise pas de phrases construites (absence de de signes de ponctuation et de majuscules).

> « *Eureka* » *Class31 [E6]*

6. Dans un souci d'authenticité, les textes des commentaires n'ont pas été modifiés. Il subsiste donc quelques coquilles orthographiques et syntaxiques, ainsi que des erreurs typographiques. Les commentaires cités en exemple, identifiables par un numéro de référence, sont repris en annexe, accompagnées de l'URL correspondante, voir *Annexe A*

— 5. Phrase(s) construite(s)

L'utilisateur utilise un langage élaboré. Il s'exprime à l'aide d'une ou plusieurs phrases articulées autour d'un verbe et d'un sujet, comportant un début, une fin, ainsi que les éléments typographiques correspondants.

> « *L'hypothèse du mariage de Pierre Pellerin et d'Emma Bernard est séduisante. Nous disposons, pour la confirmer ou l'infirmier, d'une information fort intéressante. Maurice Coulais, propriétaire au Breuil, qui fut maire de la commune entre 1941 et 1971 et s'éteignit dans sa centième année, était en effet témoin de la mariée. Un(e) Langonnais(e) le reconnaîtrait-il dans le cortège ?* » Philippe Jaunet [E7]

— 6. Jargon

L'utilisateur recourt à un jargon : son discours est parsemé de termes techniques, scientifiques, ou légaux, *etc.*

> « *'Andula' Built at Bartrams/launched 16.12.48 id 5501697.Reg owner. Soc Geral De Comercio Industria @ Transportes. Her Flag.PRT.Lisbon. In collision with, 'Statue of Libert') a Tanker, 8.6.65. Sank 9.6.65 while under tow Lisbon-Casablanca.* » brumor [E8]

— 7. Fonction phatique⁷.

L'utilisateur agrmente son propos de formules de politesse, de divers bavardages ou de mots fluidifiant son discours, le rendant moins sec. Il s'agit d'éléments portant souvent la marque de leur auteur, qui n'amènent pas véritablement de nouvelles informations.

> « *Thanks Dave, great spot and we'll arrange to swap the image around as well as get the correctly identified location information updated. Yours, Katy Britain from Above Cataloguing Team Leader* » Katy Withaker [E9]

Troisième angle d'approche : **Expertise**

— 8. Argument(s) s'appuyant sur des sources externes

L'utilisateur étaye ses propos en se référant à des sources externes, il peut s'agir de témoignages, d'extraits de presse, d'arbres généalogiques, *etc.* Ces références peuvent être accompagnées d'une *URL* mais ne le sont pas systématiquement.

> « *According to David Miller in "Charles Kerry's Federation Australia" ISBN 0908197330. These photographs, taken from a high viewpoint, were made by Willen van der Valden firstly for a commission Kerry had to document the streets of Sydney*

7. Cette formulation est inspirée d'une des six fonctions du langage que répertorie le linguiste Roman Jakobson ; la fonction phatique sous-entend la mise en place et le maintien de la communication, <http://www.signosemio.com/jakobson/fonctions-du-langage.asp> (consulté le 10.04.15)

for the Australian Town and Country Journal. He used a wide angle lens and perched on top of a mobile tower. Ref P. 20 » John Buxton [E10]

— 9. Argument(s) s'appuyant sur le contenu de l'image

L'utilisateur base ses propos sur le contenu iconographique qu'il est en train de commenter, cela peut être les caractéristiques d'une route vue du ciel, l'enseigne d'une boutique, la génération d'un véhicule, une signature, *etc.*

>« Au regard de la posture de ces hommes de troupe, et sachant que le port de la tenue militaire ne tolérerait pas plus, il y a un siècle qu'aujourd'hui, le moindre écart, je crains qu'il faille voir ici quelques camarades de retour d'un corso fleuri ou d'une procession de la Fête-Dieu.. » le00h20 [E11]

— 10. Argument(s) s'appuyant sur la situation actuelle

L'utilisateur justifie ce qu'il avance en faisant référence à la situation actuelle, par exemple en renvoyant à des vues actuelles d'un emplacement pour prouver sa localisation géographique.

>« The building is still visible on google-earth, the words have faded from the roof! » Beachcomber [E12]

— 11. Argument(s) s'appuyant sur un statut ou une fonction particulière

L'utilisateur prouve la légitimité de ses propos en mentionnant son statut ou sa fonction : par exemple le fait qu'il a habité tel lieu ou encore qu'il a exercé tel métier et reconnaît un certain type de technologies qui donne des indications quant à la date de prise de vue d'une photographie, *etc.*

>« My name is Jim Thompson and my great, great grand father was Captain David Thompson who purchased Clifton Gardens in 1871. He purchased what was then, the "Clifton Arms Hotel" which later became the "Marine Hotel". My family auctioned off the area in 1912. Unfortunately, the "Marine Hotel" was pulled down in 1967 due to the bad crowd it drew at that time. Captain David had built a wharf, pleasure grounds and a dance hall. Thompson St and David St are named after him [...] » Jim Thompson1 [E13]

Quatrième angle d'approche : **Interactions humaines**

— 12. Interpellation(s) et/ou échange(s) à caractère personnel

L'utilisateur interpelle un autre utilisateur, le remercie pour sa contribution ou l'invite par exemple à consulter une source extérieure. Parfois un véritable échange à caractère personnel s'engage entre utilisateurs, ils commentent les

trouvailles de l'autre, partagent leurs expériences, impressions, astuces, sources, etc.

>« *Well Done Dave Brunt for getting this one with a twist ie flipped. I had given up on this one after many visits. I am very pleased that we all know where it is as it puts another one to bed. Well done.* » *Class31 [E14]*

— 13. Demande d'avis, précision ou confirmation

L'utilisateur sollicite une tierce personne ou l'ensemble des autres contributeurs afin d'obtenir un renseignement, un avis ou encore une confirmation de l'hypothèse qu'il a suggérée.

>« *@Wardie44B - I had another look, and still think the two photos are taken at the same time (within a few minutes). There are differences caused by strong sunshine and contrast in this picture - a building disappeared (was it the Hotel Australia?). This photo is a full plate glass negative; the other is half plate. In this photo the tram wires have been purposely whited out (see persistent URL). The best proof comes from the two carts on the left and kerb (see notes) - the one in front has the same arrangement of parcels in both pictures - see Persistent URL. The tram wires (post December 1899) and the absence of electric lights (pre 1904) should help date the photos. Or maybe I missed something?!* » *[E15] beachcomber australia*

— 14. Correction(s) ou réponse(s) à d'autres contributeurs

L'utilisateur réagit à un commentaire précédent : il va alternativement reprendre ce qui a déjà été dit en étoffant les données à l'aide de nouveaux éléments, exprimer son désaccord, rectifier une information erronée, répondre à une question, etc.⁸.

>« *Bonjour, Je crois qu'il y a un problème avec l'hypothèse Souzy : avez-vous bien vérifié qu'il était amiral ? (Peu probable : il n'était promo capitaine de vaisseau qu'en 1876). Il ne figure nulle part comme tel... en particulier annuaires maritimes de 1877, 1878 et 1879 (au cas où, pour les nominations tardives de 1878).* » *de Baynast [E16]*

Cinquième angle d'approche : **Information**

— 15. Information descriptive

L'utilisateur décrit des éléments ou l'entièreté de l'image, il fournit les noms ou caractéristiques de ce qui est montré.

8. Ce qui distingue cette catégorie de la catégorie 12 est l'apport de nouvelles informations relatives au document.

> « *Le polonais est une langue slave avec un alphabet latin agrémenté de diacritiques, et des digraphes (ch cz dz ...) qu'on ne voit pas dans le texte. Celui-ci est en cyrillique. Dans ce texte écrit en cursive on remarque que tous les "m" sont surmontés d'un tiret. On retrouve ce tiret dans le cyrillique macédonien et serbe.* » MARQUET [E17]

— 16. Information(s) sur le lieu

L'utilisateur fournit des informations (adresse, nom, coordonnées, caractéristiques, etc.) sur le lieu qui est donné à voir dans le document iconographique.

> « *A bit of an educated(?) guess but I'm thinking that this is an open-cast coal mine site. My reasoning is : (a) Appearance in some of this set of images of a dark seam well below the original ground level and this dark material (coal?) being removed by lorry ; (b) LAING contractors sign on some earthmoving equipment (John Laing Son had a contract for some such work round about 1948. The location is far less certain but one POSSIBLE area is Heanor/Carrington Farm /Shipley Hall/Mapperley/Smalley area, west of Nottingham. There were a few open-cast mines being worked here in the late 1940s. One of them - Carrington - used a very large dragline excavator, on two 'feet' rather than caterpillar tracks. To see these, search BritainfromAbove using "opencast" as the key word.* » ewnmcg [E18]

— 17. Information(s) sur les personnes

L'utilisateur publie des informations (identité, fonction, biographie, liens de parenté, etc.) sur la ou les personnes représentée(s) sur le document.

> « *Les deux photos m'ont fait "tilt" .Sur la première je reconnais parfaitement Louis Victor baron Alquier vice amiral, 2-12-1831/13-11-1906, à la carrière militaire prestigieuse,fut maire de la Flocelliere,dans sa propriété du château de la Bergeliere à la Flocelliere. Sur la seconde photo, de gauche à droite, Laurence, contesse de Froidefond, née de Hillerin,Louis ou Albert coupvent des Graviers, Clothilde Abaudret,Laure Alquier, Soeur de l'amiral,Marie louise valérie marquise de Hillerin, née Alquier,mère de Laurence, soeur de l'amiral,assis, l'amiral Louis Victor Alquier,photo prise aux environs de 1890, au château de la Bergeliere à la Flocelliere,je possède des reproductions de photos prêtées par des descendants de familles de l'ancienne bourgeoisie Flocéenne, Je suis tres intéressé pour visionner le reste du lot,étant administrateur à l,association la Boulite à la Flocelliere. Michel Rambaud.* » RAMBAUD MICHEL [E19]

— 18. Information(s) sur le contexte

L'utilisateur propose des informations sur le contexte (événements liés/représentés

ou contexte juridique, commercial, industriel, *etc.*) de ce qui est donné à voir dans le document iconographique.

> « *Students of Victorian photo art history might find this of interest - this looks suspiciously like a tourist set up studio style photo - they were becoming very popular in the late Victorian era. Clues are : the unruffled hats and tidy bows, neatly turned out fashionable day wear dresses, with lace, women in stays and perhaps even corsets, and not sports clothing. The group shows no signs of exertion, or cold weather sporting. Very similar to other tourist photos using false scene backdrops in studios - EX : families sittings in boats with tropical ocean scene behind them. British Queen Alexandra daughters made one of that style in late 1880s - Also suspect is the difference in perspective and resolution from the men and buildings in the background in respect to the subjects in the foreground. Wonder if anyone else noticed - Pam Eisenberg, photo archivist - OC Register, Santa Ana, CA* » Palma J. Eisenberg [E20]

— 19. Information(s) sur la date

L'utilisateur fournit des informations plus ou moins approximatives (selon le degré d'indices présents dans l'image) sur la date de la prise de vue du document ou sur des dates-clés liées à ce qui est montré (par exemple ouverture ou démolition d'une usine).

> « *Sometime after Decemebr 899 when the trams were introduced to George Street and 1907-08 when the span poles were removed to enable wide-bodied trams to run in the street.* » John Cowper [E21]

Sixième angle d'approche : **Liens externes**

— 20. Renvoi (URL) vers une réutilisation de l'image

L'utilisateur fait mention d'une réutilisation de la même image dans un autre contexte, en accompagnant l'information d'une *URL*.

> « *Tout à fait d'accord! Allez sur [URL]⁹ et vous y verrez le même cliché du même photographe avec la bonne légende. C'est un site de vente de cartes postales.* » Pénisson Dominique [E22]

— 21. Renvoi (URL) vers des images « voisines »

L'utilisateur signale la présence de photographies issues de la même série et propose une *URL* vers les documents concernés. Il s'agit donc de photographies

9. <http://www.delcampe.net/page/item/id,151482512,var,85-SAINT-MICHEL-MONT-MERCURE-VIEUX-CASTEL-DE-LA-BONNELIERE,language,F.html>

prises juste avant ou juste après l'image concernée, qui fournissent ainsi de nouveaux éléments de compréhension.

> « *Much of the area is covered by the image EPW050203 [URL]¹⁰ which is of New housing estates and adjacent countryside » Class31 [E23]*

— 22. Renvoi (URL) vers un ou plusieurs document(s) lié(s)

L'utilisateur évoque une ou plusieurs sources externes (photographie montrant le même lieu par un autre artiste, extrait d'un quotidien numérisé, sources généalogiques, service de géolocalisation, etc.), consultables en ligne (URL).

> « *Here's Fred Leist's Sydney Mail illustration of the Vincennes at Manly (see pp. 1490-1491 - [URL]¹¹; the main Sydney Mail coverage had appeared in the previous week's issue. » quasymody [E24]*

Septième angle d'approche : « **Méta** »-information

— 23. Données « méta » sur le document

L'utilisateur fournit des données sur l'image en tant que document, sur la prise de vue (focale, orientation, auteur, etc.), sur le contexte de sa création ou de sa publication, etc.

> « *Comme il s'agit d'un petit éditeur il a tout naturellement travaillé avec une grosse maison d'édition. Jules Robuchon a ses débuts travaille avec Neurdein également à Nancy et on trouve sur ces cartes ce petit cachet rond. Je serais tenté de dire (si le derrière de la carte est non divisé) que la carte (en fonction de la grandeur de la photo sur le Recto qui laisse un blanc) que nous sommes à la fin de la période précurseur vers 1902. Si la carte est divisée au verso nous sommes au début de la période dite de "l'âge d'or" et Phelippeau ne s'est pas encore mis à la mode (vers 1905) [...] » Maurice Bedon [E25]*

— 24. Données « méta » sur le fonctionnement de la plateforme

L'utilisateur émet des commentaires (remarques, questions, suggestions) sur le fonctionnement de la plateforme. Un autre cas de figure : un administrateur/animateur de l'institution encourage les utilisateurs, répond à d'éventuelles remarques, informe sur la mise à jour des métadonnées, etc.

> « *Dear all, As this image is currently the most commented-upon picture on the site, I thought I'd take a closer look prompted by all your contributions. Clearly we need to change the spelling of Hayes in our caption - so I've that in hand and the change*

10. <http://britainfromabove.org.uk/image/EPW050203>

11. <http://news.google.com/newspapers?id=sYNVAAAAIIBAJ&sjid=zZUDAAAAIIBAJ&pg=7167,7938888>

will appear here in due course (probably as the next data upload scheduled for late October/early November). [...] I wonder if, give the range of detailed knowledge here, anyone would be interested in writing an entry in the Wiki tab for this image; which would help to bring all the information together in a history. yours, Katy Britain from Above Cataloguing Team Leader » Katy Whitaker [E26]

Huitième angle d’approche : **Spam**

— 25. Demande d’ajout à un groupe / pub / hors-sujet

L’utilisateur publie un commentaire de type « spam » ; il s’agit de messages « de masse », impersonnels, postés sur de nombreuses pages, visant à faire la promotion d’un compte ou d’un groupe, ou encore de messages publicitaires hors-sujet qui n’ont aucun rapport avec l’image concernée.

>« *Such an amazing BW shot! Please accept this invitation to [image] Black and white with some creative art work [URL]¹² » NBAH M [E27]*

6.3 Remarques et observations

Ces 25 paramètres ont été mis à l’épreuve au cours de l’analyse de près de 900 commentaires, soulevant au passage quelques ambiguïtés et paradoxes qu’il semble intéressant de brièvement évoquer.

En premier lieu, il faut noter que les contours de certains paramètres sont parfois imprécis, ce qui peut conduire à un « chevauchement » entre deux paramètres. C’est le cas par exemple du paramètre 18 : « Information(s) sur le contexte », qui recoupe régulièrement les paramètres 16 et 19 (informations de lieu et de date). Par exemple, lorsqu’un utilisateur identifie un bâtiment dans le tiers inférieur de l’image, à gauche, et qu’il explique ensuite qu’il s’agissait par exemple de l’usine XY, qui employait un certain nombre de personnes et qui a fermé après tel événement... faut-il répondre par l’affirmative à ces trois paramètres concernant l’apport de nouvelles informations 15 (descriptives), 16 (lieu) et 18 (contexte)? Cependant, comme Boydens et van Hooland (2011) l’ont montré, il serait vain de croire qu’il puisse exister une parfaite correspondance entre une base de données et le réel empirique. En effet, les données portant sur le patrimoine culturel sont empiriques et non déterministes (Boydens et van Hooland, 2011). Cela signifie, d’une part, qu’elles sont par essence sujettes à interprétation dans le temps et l’espace. D’autre part, cela implique qu’il n’existe aucun référentiel « absolu » permettant de tester la validité des catégories choisies. Un certain

12. <https://www.flickr.com/groups/1575512@N25/>

enchevêtrement des paramètres semble donc inévitable et peu préjudiciable au bon déroulement de cette analyse, d'autant plus que les catégories ne sont pas exclusives.

D'autre part, la décision de ne « valider » les paramètres 20, 21 et 22 qu'en présence d'une *URL* ne semble pas toujours pertinente. En effet, certains utilisateurs font référence au nom d'un site web sans donner son *URL*, alors que d'autres, plus nombreux encore, signalent des documents liés qui sont édités sur papier... Ils ne sont dès lors pas comptabilisés, en dehors du très large paramètre 8. Il aurait été intéressant de créer des catégories distinctes, en insérant par exemple une catégorie 22' pour les documents liés dont le lien n'est pas donné.

Concernant la date, il est difficile de savoir que prendre en compte : la date de la prise de vue, mais aussi les dates relatives à ce qui est montré ? Dans le cas de photographies aériennes montrant l'ensemble d'une ville par exemple, lorsqu'un utilisateur mentionne la date de construction ou de démolition d'un bâtiment extrêmement connu, cette date doit-elle avoir la même « valeur » qu'une date de prise de vue déduite au terme de longues observations de divers indices présents dans l'image ? Faut-il toutes deux les « comptabiliser » ? De la même façon, la mention d'une source externe (paramètre 8) ne correspond absolument pas au même investissement s'il s'agit simplement de l'évocation d'un ouvrage ou s'il s'agit d'un utilisateur qui se rend chez les descendants d'une personne photographiée afin d'obtenir des informations inédites... Or, l'information sera prise en compte de la même manière dans la grille d'analyse. De même, le « degré » de richesse de la contribution ne transparait pas dans le cadre d'informations sur le lieu : l'utilisateur donne-t-il simplement un nom de lieu ou une description approfondie se structurant en plusieurs paragraphes ?

Si certaines catégories ont pu être ajoutées au cours de la phase de test (échantillon de 80 commentaires), il subsiste toutefois des lacunes : il aurait par exemple été intéressant d'avoir un paramètre concernant l'auto-correction (qui fut englobée dans le paramètre 14 dans le cadre de cette analyse) ou le simple ajout de précisions à un commentaire posté précédemment (par la même personne). Un autre paramètre aurait pu concerner exclusivement les informations sur l'orientation des photographies (par rapport aux points cardinaux), qui ont été intégrées au paramètre 23 qui comprend les conditions de prise de vue. Il aurait également pu être intéressant de signaler les contributions faisant partie d'un flux très important de commentaires (plus de 50 commentaires/document, par exemple). Une autre dimension non couverte par la grille d'analyse est le caractère hypothétique de certains commentaires¹³. Certains utilisateurs émettent des suppositions et font clairement savoir que ce ne sont pas des informations véridiques, mais il s'agit tout de même de pistes permettant généralement

13. Le paramètre qui se rapproche le plus de tels propos est le 2. (expression d'avis/convictions), mais il ne s'agit pas véritablement de la même chose.

d'en apprendre davantage sur le document. Un peu dans la même idée, quelques utilisateurs zélés réalisent un véritable travail de relecture, de reformulation et de synthèse en publiant des commentaires qui reprennent l'essentiel de l'information déjà publiée ; aucun paramètre n'identifie ce type de contribution.

Enfin, il faut garder à l'esprit que seules les informations précisées par l'utilisateur ont été prises en compte : s'il ne fait pas explicitement mention de ses sources alors qu'il en a peut-être utilisées plusieurs, cette information n'apparaît pas et pourrait biaiser les résultats. De même pour une personne dont le propos est scindé en plusieurs commentaires distincts.... Par souci de cohérence, une information mentionnée précédemment (par le même utilisateur) ne sera pas prise en compte dans l'analyse d'un commentaire ultérieur, ce qui ne reflète pas parfaitement la réalité.

De manière plus générale, il apparaît que l'étude de l'échantillon seul, isolé de son contexte, ne suffit pas. Cela s'explique par le fait que, d'une part, il est appréciable, voire indispensable selon les cas, d'avoir sous les yeux le document commenté afin de bien comprendre le propos. D'autre part, il arrive fréquemment que les commentaires consistent en échanges entre plusieurs utilisateurs. Prend forme alors une dynamique faite de questions, réponses, remarques, rectifications, *etc.* Dès lors, il serait illusoire de croire qu'un commentaire sorti de son environnement puisse être compris et analysé.

En fin de compte, l'exploration des données permet d'affirmer qu'il aurait été plus pertinent d'avoir trois cas se ressemblant davantage, à l'exception du facteur « taille » (du projet et de l'institution). La sélection des trois cas aurait ainsi pu être faite en partant de critères de sélection plus précis. Par exemple : trois archives, possédant des collections semblables (photographies en noir et blanc, avec des sujets similaires) et utilisant le même type de plateforme pour accueillir leur projet. Cela aurait sans doute permis de pouvoir mieux comparer l'impact de l'ampleur du projet sur les résultats, sans que n'apparaisse, par exemple, le « biais *Flickr* ».

GRILLE D'ANALYSE

I. APPORTS PERSONNELS	<ol style="list-style-type: none"> 1. Appréciation subjective 2. Expression de convictions 3. Anecdotes personnelles
II. LA LANGUE	<ol style="list-style-type: none"> 4. Mots isolés 5. Phrases construites 6. Jargon 7. Fonction phatique
III. EXPERTISE	<p>Justification basée sur :</p> <ol style="list-style-type: none"> 8. Des sources externes 9. Le contenu du document 10. La situation actuelle 11. Le statut / la fonction
IV. INTERACTIONS HUMAINES	<ol style="list-style-type: none"> 12. Interpellations / échanges à caractère personnel 13. Demande d'avis / confirmation 14. Corrections / réponses à d'autres contributeurs
V. INFORMATION	<ol style="list-style-type: none"> 15. Informations descriptives 16. Informations sur le lieu 17. Informations sur les personnes 18. Informations sur le contexte 19. Informations sur la date
VI. LIENS EXTERNES	<ol style="list-style-type: none"> 20. Réutilisation de l'image 21. Images voisines (URL) 22. Documents liés (URL)
VII. MÉTA-INFORMATION	<ol style="list-style-type: none"> 23. Données "méta" sur le document 24. Données "méta" sur la plateforme
VIII. SPAM	<ol style="list-style-type: none"> 25. Demandes d'ajout à un groupe, pub, hors-sujet

FIGURE 6.1 – Vue récapitulative de la grille d'analyse.

6.4 Interprétation des résultats

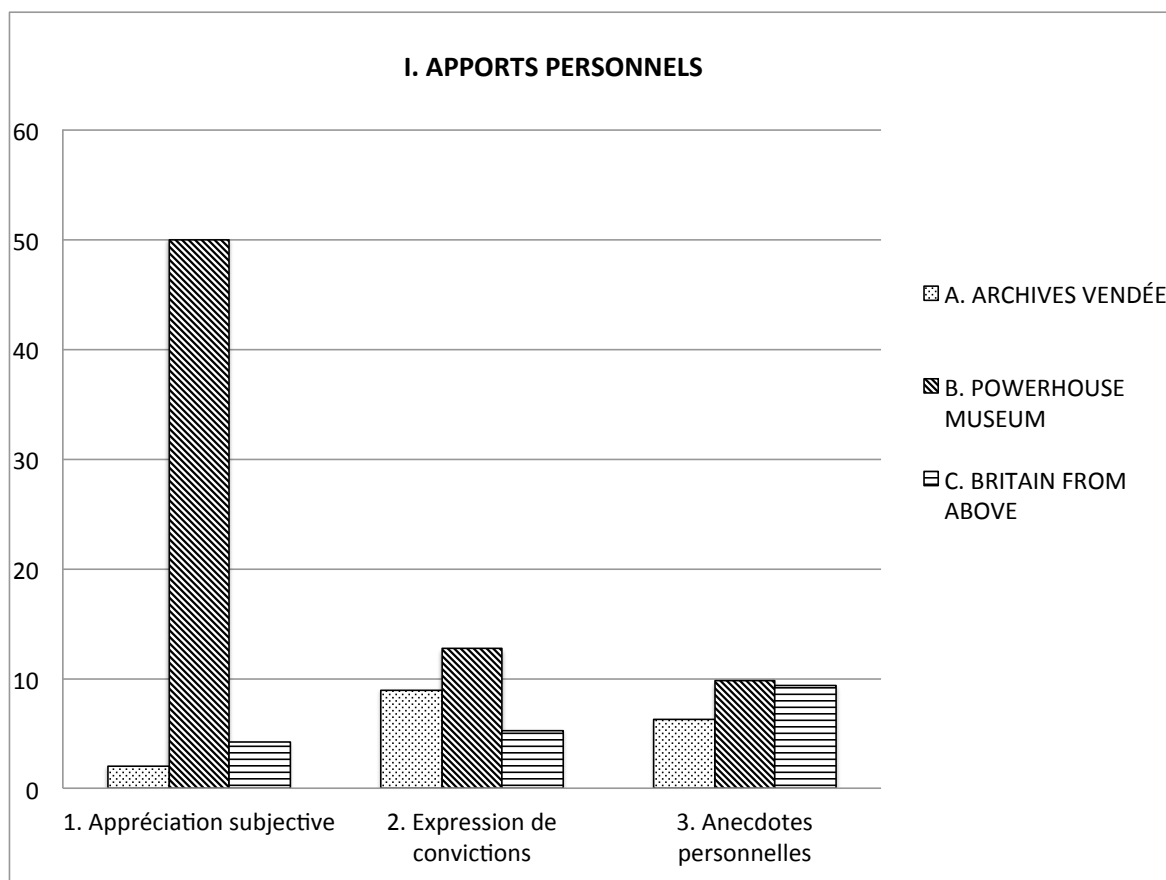


FIGURE 6.2 – I. Apports personnels. Résultats exprimés sous forme de pourcentages.

I. Apports personnels Voir Figure 6.2

1. Appréciations subjectives : l'élément qui retient l'attention de prime abord est le pourcentage très élevé (50%) correspondant aux données du cas B, qui se démarque très nettement des deux autres pourcentages (2% et 4% pour les cas A et C). Cela laisse penser que le facteur le plus influent n'est pas ici la taille de l'ensemble du projet, mais le type de plateforme d'accueil, à savoir *Flickr*. Le « public » n'y est sans doute pas le même et l'atmosphère semble moins formelle. Globalement, *Flickr* s'apparente davantage à un réseau social de partage de contenus qu'à un espace d'« enquêtes ». En effet, c'est un écosystème au sein duquel il est tout à fait courant que l'on manifeste son goût pour une image, ce qui ne semble pas aller de même pour les deux autres cas. D'ailleurs, les utilisateurs ayant laissé ce type de commentaires sont en grande majorité des « visiteurs uniques ». En outre, bien que l'encodage ne détaille pas si les appréciations sont de nature plutôt positives ou négatives, la lecture attentive du corpus a permis de constater que très peu de commentaires comportent

des éléments négatifs. Ce qui amène bien sûr à se demander s'il s'agit d'une tendance naturelle ou si c'est le fruit de la modération.

2. Expression de convictions : les pourcentages sont peu élevés dans les trois cas. Ce type de propos ne semble pas vraiment avoir sa place au sein de tels projets. À l'instar du paramètre précédent, il y a une légère prépondérance du cas B, qui s'explique sans doute par le biais des mêmes raisons.

3. Anecdotes personnelles : dans les trois cas, le pourcentage de commentaires concernés atteint moins de 10%. La tendance y est légèrement moins forte au sein du cas A : l'appellation « laboratoire » n'encourage pas vraiment au partage de récits personnels. En fin de compte, si le web reste indéniablement un espace privilégié pour recueillir des témoignages de « porteurs de mémoire » avant leur décès, ces résultats font toutefois relativiser le poids de cet argument souvent cité par les institutions souhaitant mener des projets de *crowdsourcing*.

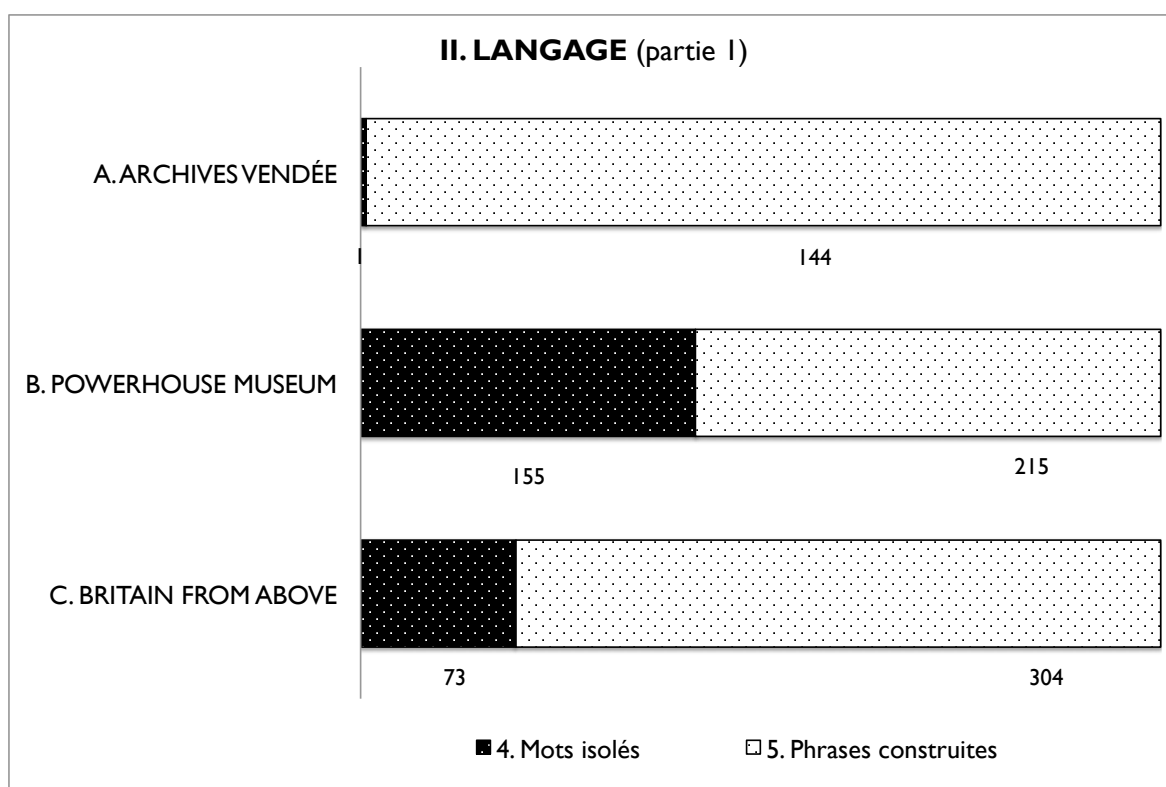


FIGURE 6.3 – II. Langage (partie 1). Mots isolés *versus* phrases construites.

II. Langage Voir Figures 6.3 et 6.4

4. & 5. Mots isolés versus phrases construites : le cas A se démarque radicalement des deux autres, avec 144 « phrases construites » contre un seul commentaire constitué de mots isolés ! Le fait que les utilisateurs soient invités à contribuer à des enquêtes favorise sans doute des formes d'expression écrite plus soignées et élaborées. Au niveau du cas C, les phrases construites prédominent encore largement (304, contre 73 « mots isolés »). Concernant le cas B, les commentaires formés de mots isolés sont déjà beaucoup plus nombreux, atteignant plus de 40% de l'échantillon. Ce résultat peut être rapproché du nombre élevé de commentaires comportant des appréciations subjectives (1.), vu que celles-ci sont très souvent exprimées à l'aide d'un ou de plusieurs mots isolés.

6. Jargon : dans l'ensemble, l'usage d'un jargon dans les commentaires est assez faible, atteignant moins de 10% pour chacun des trois cas. À nouveau, la distribution des pourcentages ne correspond pas à une logique liée à la taille des projets, vu que le cas B n'atteint que 3% contre environ 8% pour les deux autres cas. Cette petite différence pourrait dériver du fait que les deux autres plateformes drainent un public plus ciblé, qui possède peut-être une plus grande expertise.

7. Fonction phatique : plus de 25% des commentaires possèdent cette caractéristique et ce, quel que soit l'échantillon de données : ce pourcentage reflète le ton courtois avec lequel s'exprime une certaine frange des utilisateurs, tandis que les autres se cantonnent à un style beaucoup plus neutre et impersonnel. Un sérieux écart sépare le cas A des deux autres. Au vu de certaines tournures de phrases ou expressions désuètes qu'un bon nombre utilise, il est probable que la majorité des utilisateurs des Archives de Vendée soit plutôt âgée. Par exemple, lors d'échanges entre utilisateurs, c'est le vouvoiement qui prime, alors qu'en général le tutoiement est de rigueur sur le web, peu importe que l'on connaisse déjà son interlocuteur ou non. D'autre part, certains contributeurs ont le réflexe de saluer les autres en début de message et de prendre poliment « congé », en signant de leur nom. Il s'agit d'une « survivance » de réflexes épistolaires, que peu prennent encore la peine d'utiliser sur le web. Le cadre particulier de l'« enquête » déjà suggéré pour les paramètres 4 et 5, ainsi que cette donnée liée à l'âge expliquent sans doute ce haut pourcentage pour le cas A.

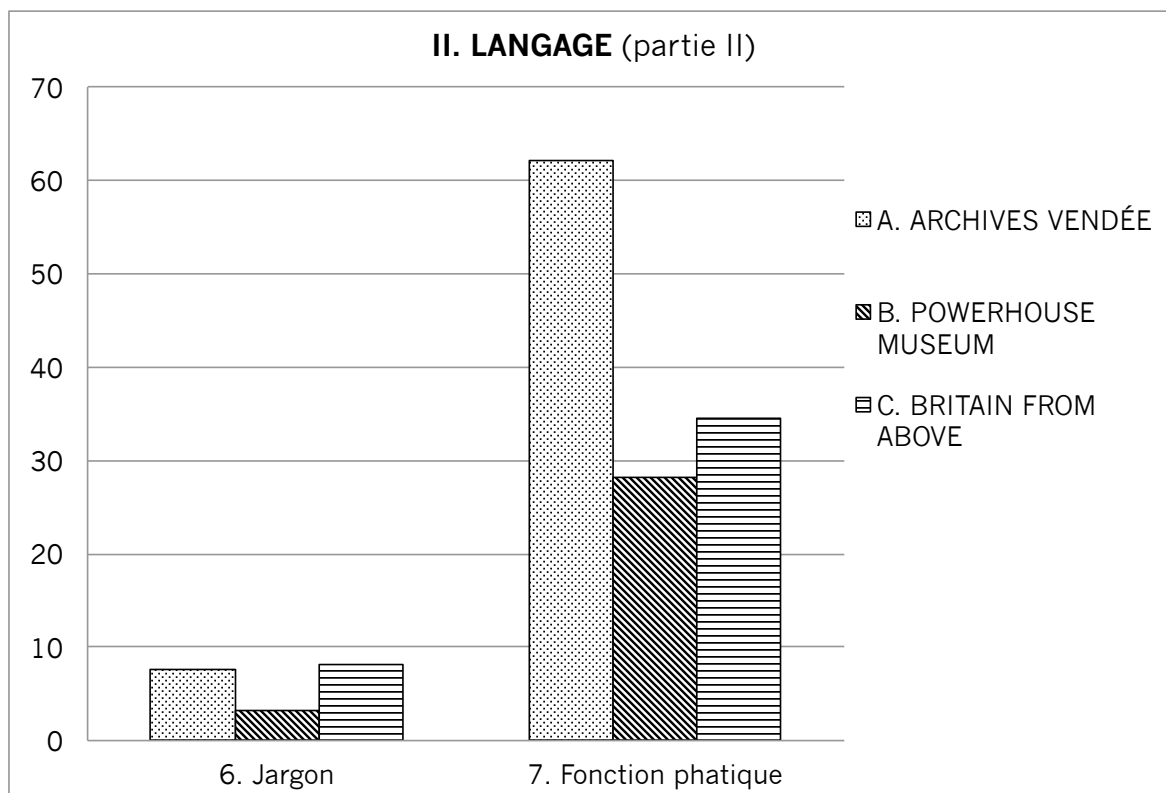


FIGURE 6.4 – II. Langage (partie 2). Jargon et fonction phatique, résultats exprimés sous forme de pourcentages

III. Expertise *Voir Figure 6.5*

8. Justification basée sur des sources externes : à nouveau, le pourcentage très élevé correspondant au cas A retient l'attention (près de 55% alors que l'on se situe sous les 20% pour les cas B et C). Cette fois-ci, la taille du projet est bien l'un des facteurs pouvant expliquer ce résultat. En effet, le nombre restreint d'enquêtes « ouvertes » au même moment (par rapport au 95.000 images mises en ligne du cas C par exemple) permet de concentrer le travail de recherche sur quelques cas et d'approfondir l'investigation. Cela peut par exemple se faire en utilisant des informations issues de sources externes. D'autre part, il est très important de souligner une particularité observée au sein du cas A : si une nouvelle information stratégique n'est pas justifiée par une source externe, alors, l'une des personnes administrant le projet va demander à l'utilisateur de citer ses sources¹⁴.

9. Justification basée sur le contenu du document : comme pour le paramètre 8, le cas A possède le pourcentage le plus élevé. Ces 43% sont sans doute liés à la notion

14. « Le 23 février 1947 semble bien avoir été un dimanche. La réponse ci-dessus paraît définitive. On aimerait toutefois qu'elle s'appuie sur une source : -témoignage oral? -croisement avec d'autres photos disponibles ailleurs? [...] » *Archives Vendée* *£*.

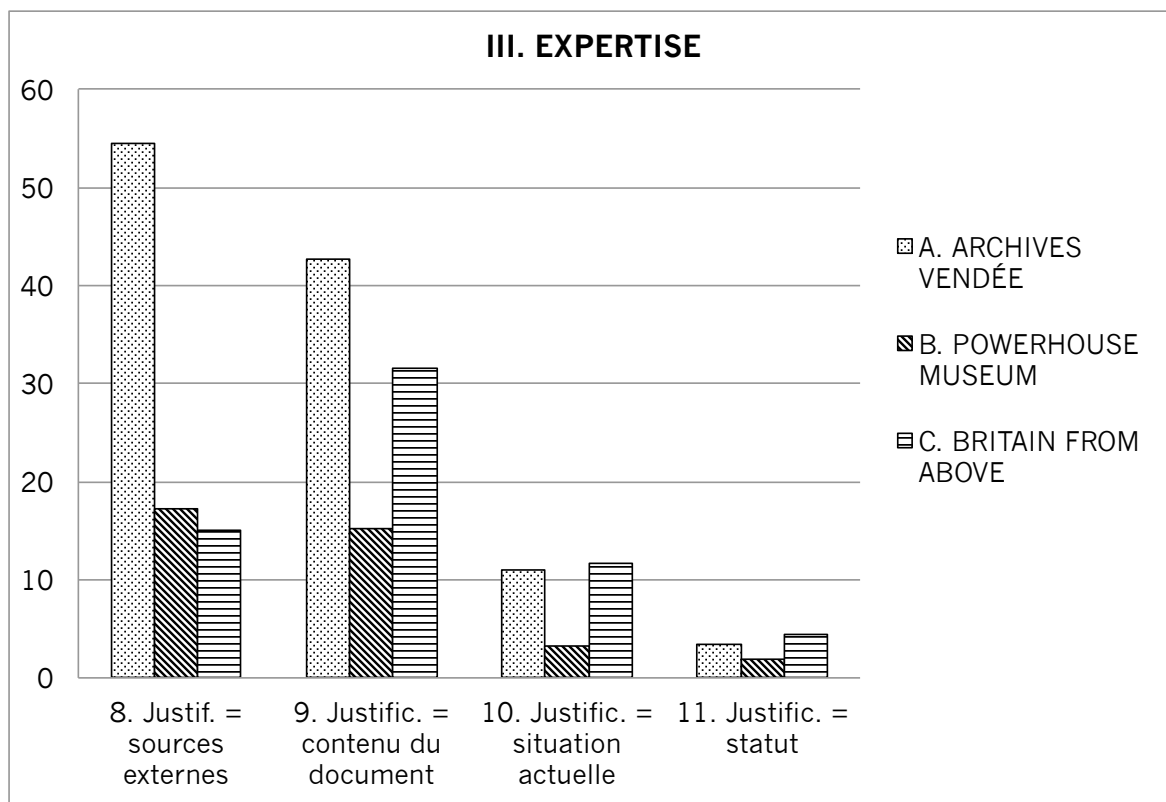


FIGURE 6.5 – III. Expertise. Résultats exprimés sous forme de pourcentages

d'enquêtes, qui favorise une analyse approfondie visant à décortiquer chaque indice. En outre, il ne faut pas oublier que ce sont des Archives qui sont à l'origine de ce projet. Il est donc plausible que les personnes laissant des commentaires soient formées ou du moins habituées à la lecture critique de documents, n'hésitant pas à se référer à des détails de l'image pour légitimer leurs hypothèses. Ce facteur d'explication est sans doute également valable pour le cas C, dont le pourcentage se situe au-dessus de 30, alors que pour le cas B, seuls 15% des commentaires comportent une justification basée sur le contenu de l'image. Cet écart peut donc être dû au fait qu'il s'agit d'un musée et non pas d'archives ou alors peut-être est-il simplement la conséquence de la très forte proportion de commentaires concis qui expriment seulement une appréciation.

10. Justification basée sur la situation actuelle : globalement, la tendance reste la même, avec le cas B qui est en retrait (voir points précédents). Il aurait cependant été intéressant de connaître le lieu de résidence des contributeurs : habiter la région fait d'eux des experts.

11. Justification basée sur un statut ou une fonction particulière : ces résultats font écho au nombre assez faible d'anecdotes personnelles (voir paramètre 3). Finalement, il est assez éclairant de voir que, concrètement, moins de 5% des

commentaires de l'échantillon s'appuient sur le statut ou vécu particulier d'une personne. Ce résultat est toutefois à considérer avec précaution : toutes les personnes qui se sont appuyées sur leur expérience personnelle pour émettre de nouvelles informations ne l'ont pas forcément mentionné.

IV. Interactions humaines *Voir Figure 6.6*

12. Échanges à caractère personnel : le cas A semble être propice aux échanges, avec environ 23% de ses commentaires qui possèdent cette caractéristique, tandis que les proportions pour les cas B et C sont d'environ 8 et 11%. Ce phénomène peut sans doute être mis en corrélation avec la taille plus modeste du projet et avec le fait que seuls 52 contributeurs différents sont à l'origine des 145 commentaires de l'échantillon. Cette dimension plus intimiste encourage sans doute les échanges à caractère personnel entre utilisateurs. Quant aux deux autres interfaces, elles peuvent être davantage assimilées à des lieux de passage, avec moins de fidèles contributeurs (bien qu'il en existe tout de même), ce qui pourrait expliquer les plus faibles pourcentages.

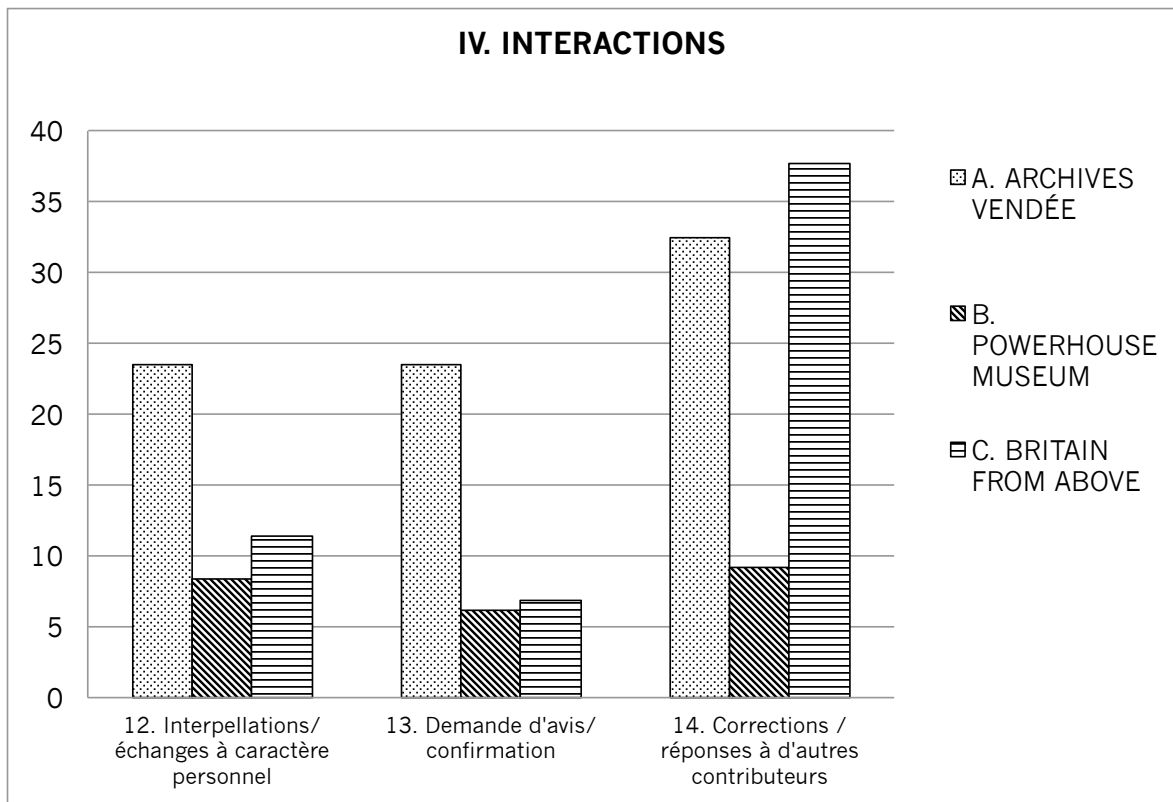


FIGURE 6.6 – IV. Interactions. Résultats exprimés sous forme de pourcentages

13. Demande d'avis ou de confirmation : le contraste est à nouveau marqué entre le cas A et les deux autres. Les raisons sont sans doute sensiblement les mêmes

que pour le paramètre 12.

14. Corrections / réponses à d'autres contributeurs : cette fois-ci, ce qui frappe en premier le regard, c'est l'immense progression du cas C par rapport aux paramètres 12 et 13. Alors que relativement peu d'échanges à caractère personnel (point 11.) ou demandes de confirmations (point 12.) sont émis, il apparaît que plus d'un tiers des commentaires constituent des corrections ou réponses à d'autres contributeurs ! En outre, c'est assez interpellant de voir que les pourcentages sont plus élevés pour chacun des trois cas, si on les met en regard avec les paramètres 12 et 13. Cela témoigne de la façon dont la co-construction des savoirs s'organise : cela se fait par tâtonnements, en partant d'une information donnée par l'autre, qui sera complétée ou corrigée, et ainsi de suite.

V. Information Voir Figure 6.7

15. Informations descriptives : le cas A domine les deux autres. Si les pourcentages des trois cas dépassent les 20%, le cas B est cependant clairement en retrait. Un rapide survol des graphiques suivants vient confirmer ce qui se devinait déjà dans les résultats précédents : si une présence sur *The Flickr Commons* semble propice à une forte participation, celle-ci n'implique pas forcément l'apport de nouvelles informations.

16. Informations sur le lieu : il s'agit du seul paramètre (de l'angle V.) pour lequel le cas C surpasse les deux autres. Ceci s'explique sans doute par la nature particulière du corpus (photographies aériennes d'un territoire). Le cas A le suit avec plus de 30% des commentaires qui répondent à ce critère et enfin le cas B est davantage en retrait. Une piste d'explication concernant ce retrait (mais qui est également applicable aux deux autres cas), réside dans le fait que certains documents ont déjà été partiellement documentés par l'institution et que les données liées concernant le lieu sont donc parfois déjà connues.

17. Informations sur les personnes : le cas A, dont le fonds correspondant comporte plusieurs portraits (d'ailleurs, l'une des trois thématiques parmi lesquelles sont regroupés les billets s'intitule « Personnes »), compte plus de 34% de commentaires qui contiennent des informations sur des individus. En revanche, les pourcentages des cas B et C sont très peu élevés. Une fois de plus, la nature du corpus peut être envisagée comme piste d'explication. Par exemple, le fonds du cas C se caractérise par des photographies au sein desquelles la présence humaine est extrêmement rare ou du

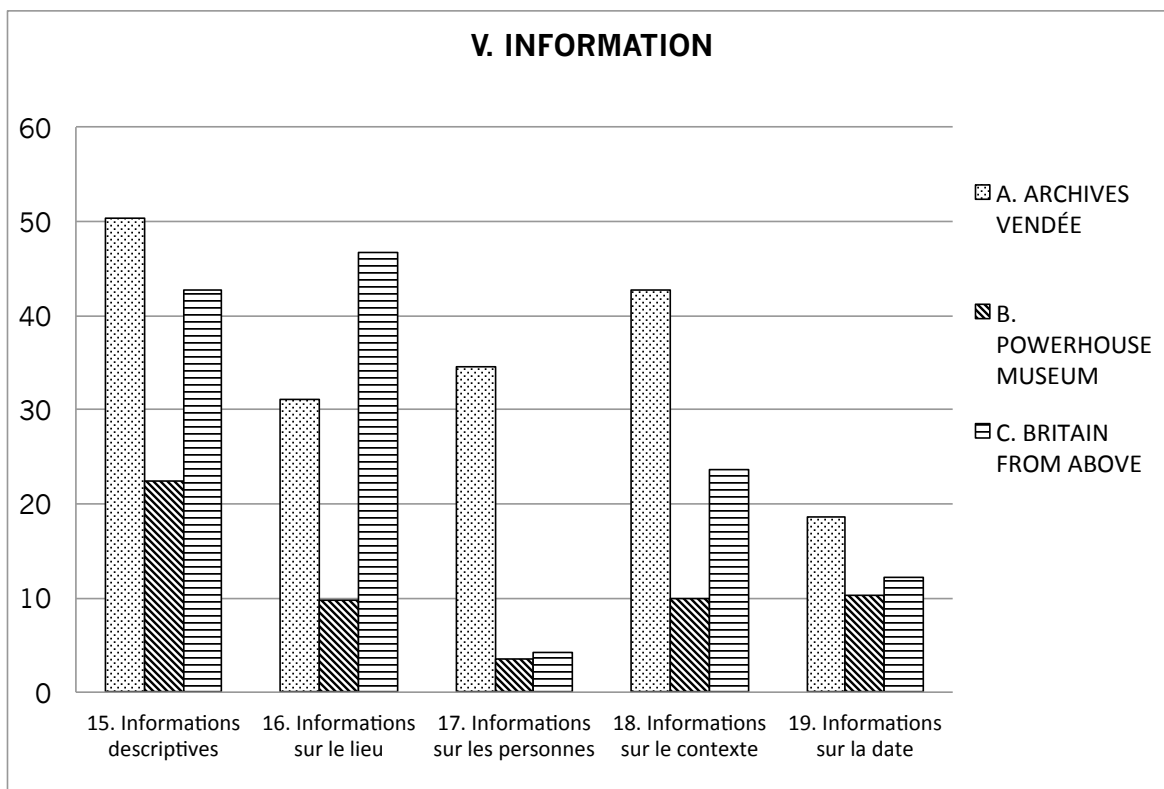


FIGURE 6.7 – V. Information. Résultats exprimés sous forme de pourcentages

moins pas nettement identifiable. Cette rapide comparaison témoigne de l'existence d'une corrélation tout à fait logique entre le type de contenu visible dans le corpus de documents et le type d'information qui figure dans les commentaires des utilisateurs.

18. Informations sur le contexte : la tendance est assez similaire à celle des paramètres 15 et 17 (voir points précédents).

19. Informations sur la date : la tendance reste la même ici encore (voir points précédents), même s'il faut relever que les pourcentages des trois cas dépassent cette fois-ci les 20%. Comme indiqué ci-dessus (paramètre 16), il faut garder à l'esprit que ces résultats peuvent être liés au fait que ces données sont parfois déjà connues.

REMARQUES portant sur l'ensemble du cinquième angle d'approche (V. Information, paramètres 15 à 19) : si l'on considère chaque paramètre de façon isolée, la proportion de commentaires contenant de l'information se révèle en moyenne assez faible. Idéalement, il serait intéressant de pouvoir additionner ces pourcentages afin d'avoir une vision globale : combien de commentaires amènent véritablement de l'information ? Cependant, les résultats seraient faussés par le fait que les catégories ne sont pas exclusives, ce qui signifie que certains commentaires contiennent à la fois des

informations descriptives, contextuelles, *etc.* D'autre part, il faut également prendre en compte l'éventuelle présence de questions que l'institution adresse aux utilisateurs. Ainsi, si le cas A possède dans l'ensemble les plus hauts pourcentages, c'est peut-être lié au fait que chaque billet est introduit par diverses questions très ciblées¹⁵.

VI. Liens externes *Voir Figure 6.8*

20. Réutilisation de l'image : la pratique reste relativement insignifiante, au-dessous des 4% pour les trois cas, avec un ratio particulièrement bas pour le cas C (1 pour 377). Ces proportions sont bien sûr directement liées à la nature et notoriété des documents : s'il s'agissait d'images très populaires qui circulent beaucoup sur le web et/ou font l'objet de détournements, à l'instar de chefs d'œuvres de l'histoire de l'art, les pourcentages seraient sans doute plus élevés. D'autre part, cela dépend également du cadre qui est posé : l'institution met-elle l'accent sur la « redocumentarisation » ou non ?

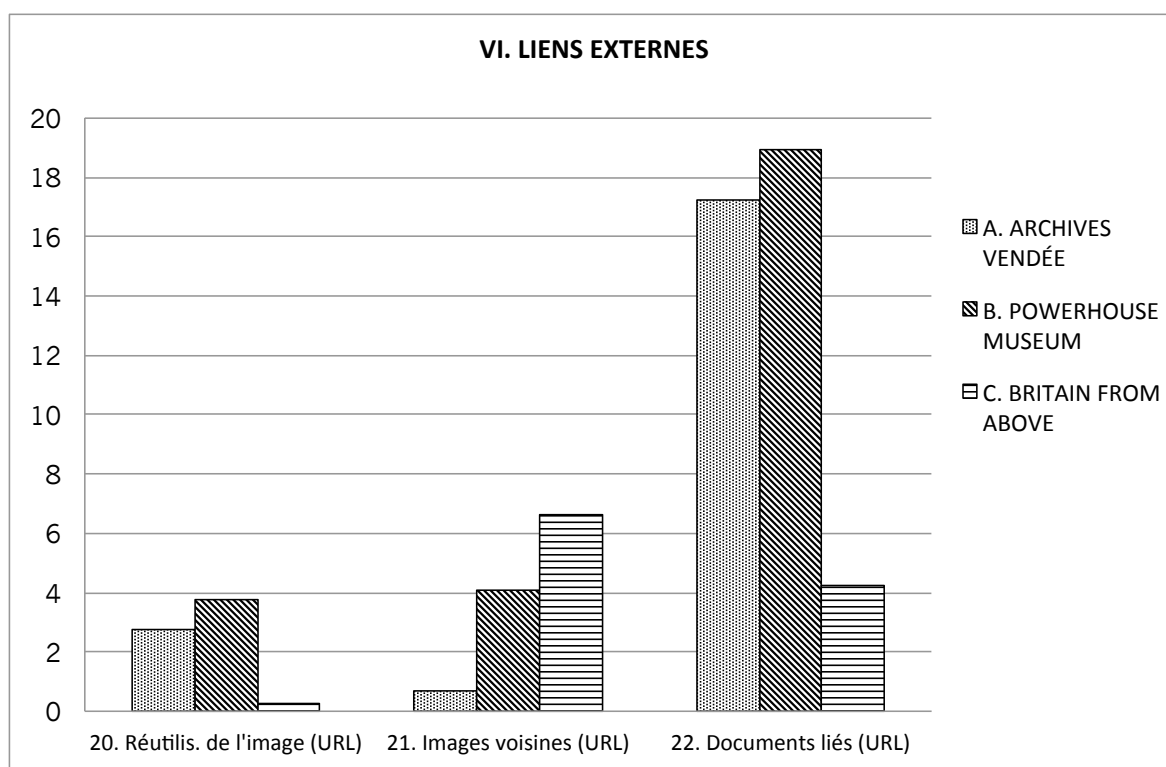


FIGURE 6.8 – VI. Liens externes. Résultats exprimés sous forme de pourcentages

15. Exemple de questions, billet du 02.02.2014 : « Reconnaît-on bien ici une rue du Langon ? Qui sont le violoneux, le hautboïste et le trompettiste ? Sachant que la noce est (de peu ?) antérieure au 6 mai 1911, qui peuvent être les mariés ? »

21. Images voisines : 0.7% (cas A) à 6.6% (cas C) des commentaires comportent des URL renvoyant à des images voisines. Concernant le cas A : lorsqu'il est question d'une série d'images qui se suivent, il arrive qu'elles soient postées au sein du même billet. Cette première cause, combinée au fait que plusieurs utilisateurs ne semblent pas avoir le réflexe d'utiliser des hyperliens lorsqu'ils citent des documents liés, explique que seul 1 commentaire sur 144 en fasse mention. Concernant les cas B (4.1%) et C (6.6%), si la publication de liens renvoyant vers des images voisines existe, l'analyse de l'échantillon montre qu'il ne s'agit pas d'un comportement ordinaire, mais plutôt d'une pratique de niche qui est le fruit d'une minorité d'utilisateurs zélés. Possédant une bonne connaissance du corpus d'images, ils repèrent les images liées et en font part aux autres utilisateurs, sans doute pour montrer leur maîtrise et/ou dans l'idée de faire avancer le processus de documentation.

22. Documents liés : les nombres sont cette fois-ci beaucoup plus élevés, à l'exception du cas C. Pour le cas A et B, plus d'un commentaire sur six comporte au moins un lien vers un document lié. Dans certains cas, le lien vient étayer une recherche très riche et élaborée ; d'autres fois, le lien n'est accompagné d'aucun commentaire, par exemple lorsqu'un utilisateur localise le lieu sur *Google Maps*. Comme indiqué précédemment, cette catégorie concerne des cas si variés qu'il aurait été intéressant de la subdiviser pour pouvoir en tirer des interprétations plus poussées. Concernant le cas C, il est difficile d'identifier pourquoi le pourcentage est si bas : peut-être est-ce lié au contenu des documents, qui fait davantage appel aux connaissances et souvenirs personnels et ne nécessite pas forcément un recours à des sources externes ; il est également possible que le volume de sources potentiellement mobilisables pour ce fonds-là soit plus restreint.

VII. « Méta »-information Voir Figure 6.9

23. Données « méta » sur le document : les trois cas se situent entre 11 et 21%. Il faut noter que les types de commentaires varient d'un cas à l'autre. Pour le cas A, il s'agit plutôt de données sur le support du document, tandis que pour les cas B et C, la majorité des commentaires porte sur la prise de vue (orientation par rapport aux points cardinaux, technique, *etc.*).

24. Données « méta » sur la plateforme : l'écart entre le cas A et le B est très important, tandis que le cas C se situe à mi-chemin. Cette différence s'explique par le fait que l'équipe du *Powerhouse Museum* ne semble pas utiliser cet outil pour communiquer¹⁶. En revanche, un véritable dialogue s'engage entre les membres de

16. Au sein de l'échantillon étudié, aucune intervention n'a été relevée.

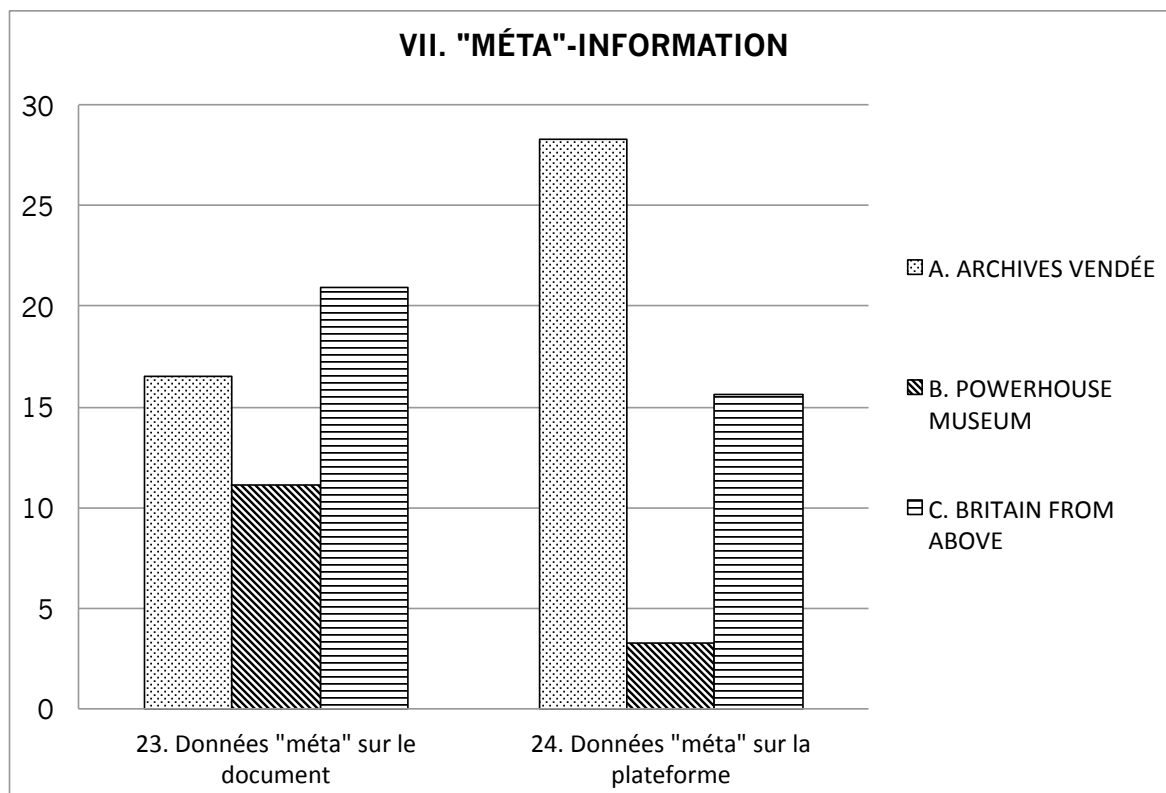


FIGURE 6.9 – VII. « Méta »-information. Résultats exprimés sous forme de pourcentages

l'institution et les utilisateurs pour les cas A et C. Ainsi, l'auteur des Archives de Vendée (cas A) est le contributeur le plus actif et Katy Withaker, l'une des responsables de *Britain from Above* (cas C) a posté plus de 1000 commentaires, soit une part significative des commentaires.

VIII. Spam

25. Spam : les résultats obtenus permettent de déduire si les professionnels qui encadrent ces projets assurent une fonction de modération ou non. Ainsi, pour le cas A, c'est explicite : aucun message hors-sujet ou publicitaire n'a été relevé¹⁷. Concernant le cas C, c'est presque la même configuration. Seuls 0.8% des commentaires rentrent dans cette catégorie et encore, il s'agit de messages légèrement hors-sujet et non pas de véritables spams en tant que tels. Concernant le cas B, la tendance est nettement différente. Aucune modération des commentaires ne semble être faite. Les spams représentent plus de 5% des commentaires publiés. Cependant, ce nombre est à considérer avec précaution. En effet, il se trouve qu'entre l'extraction et le tri

17. Afin de vérifier qu'il s'agissait bien d'un mécanisme de modération, un petit test a d'ailleurs été effectué en postant un message hors-sujet : publié instantanément, il fut effacé dans les 24 heures qui suivirent.

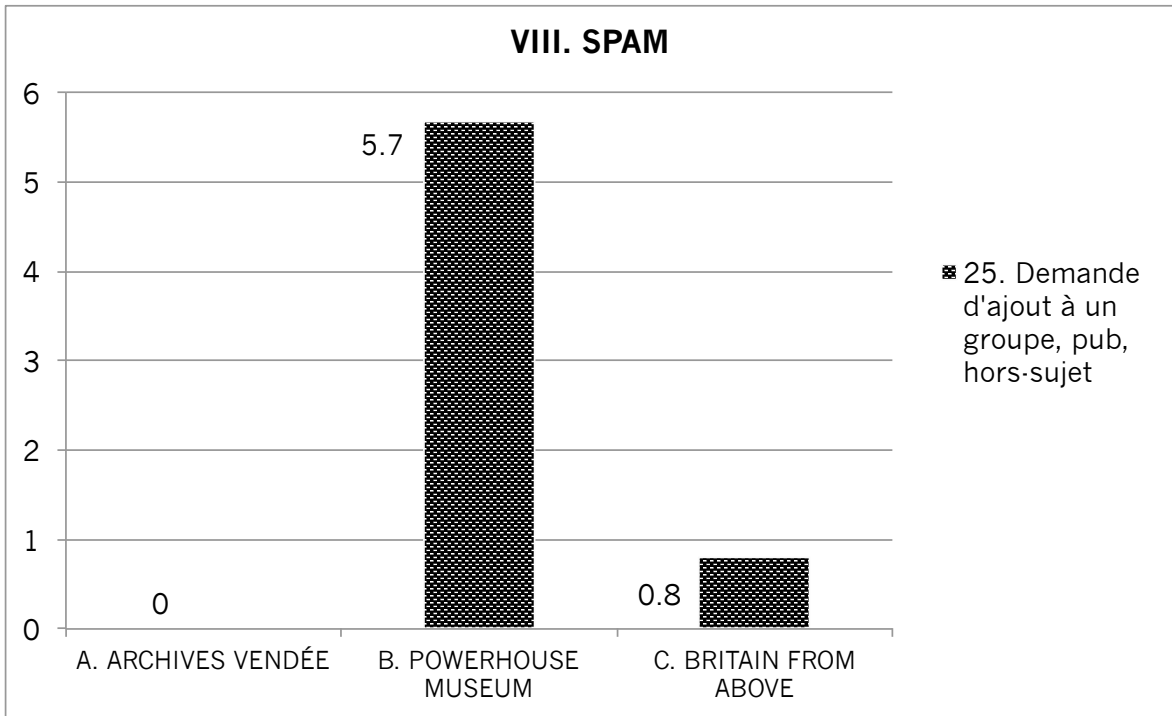


FIGURE 6.10 – VIII. Spam. Résultats exprimés sous forme de pourcentages

des doublons, 264 spams ont été supprimés par mégarde avant que l'échantillon ne soit constitué. Il faut donc impérativement prendre en compte le fait que sans cette suppression accidentelle la proportion de spams aurait vraisemblablement été plus élevée.

Chapitre 7

Implications pour l'institution

7.1 Corpus, plateforme et calendrier

L'interprétation des résultats a mis en lumière les différences qu'implique le choix d'un certain corpus ou d'un certain type de plateforme. Il est donc essentiel que les institutions s'interrogent sur ce qu'elles ont à proposer : quelle est la taille et la nature du corpus d'images ? Quelles sont les connaissances ou compétences qui devront être mobilisées pour pouvoir enrichir les métadonnées : s'agit-il de lieux à reconnaître, d'événements, de personnes ? Est-ce qu'il est indispensable d'avoir habité la région ou faut-il savoir rechercher de nouvelles informations dans des registres paroissiaux en ligne, *etc.* ? Est-ce pertinent de publier l'intégralité des images (*voir cas C*) ou est-ce plus stratégique d'opérer une sélection (*voir cas A*), en vue de concentrer les efforts sur certains documents en particulier ?

Par ailleurs, le choix de la plateforme ne devrait pas être dicté uniquement par des contraintes budgétaires et devrait lui aussi faire l'objet d'une réflexion approfondie. Premièrement, parce que ce choix aura une répercussion sur le type de public touché. Deuxièmement, parce que les analyses effectuées montrent que le choix de l'outil a également une influence sur la nature des commentaires (langues, longueur, fiabilité de l'argumentation, pertinence, *etc.*).

En outre, dans la phase de conception du projet, l'institution devra également clarifier les questions relatives au calendrier. L'institution pourra notamment se demander si elle désire publier l'intégralité des images dès le début, si elle prévoit d'annoncer l'ajout de nouveaux éléments selon une fréquence précise (au risque de provoquer l'insatisfaction des utilisateurs en cas de non-respect de la fréquence annoncée) ou si elle juge préférable que cela soit fait de façon aléatoire, selon la disponibilité du personnel. D'autre part, il est bien sûr essentiel de réfléchir à la pérennité de la plateforme. L'institution peut opter pour un dispositif permanent ou un projet limité dans le temps. En fonction de ces choix, les contraintes en termes de

ressources humaines et de budgets seront différentes. Enfin, si le projet possède une date de fin officielle, il faudra penser à clarifier quel sort connaîtra la plateforme au terme de cette date.

7.2 Profils des contributeurs

Comme suggéré ci-dessus, il est primordial que l'institution réfléchisse à ses attentes concernant le contenu des commentaires publiés par les utilisateurs. Selon ses objectifs, le public cible ne sera pas le même. Concrètement, les profils des participants seront sans doute très hétérogènes vu qu'aucune présélection n'est effectuée, mais l'institution peut toutefois s'employer à mettre en place des mesures favorisant plutôt la participation de tel ou tel profil de contributeurs. Les objectifs des institutions peuvent être répartis en trois grands groupes¹ : le recueil de témoignages ; la production de métadonnées de qualité ; l'offre d'une nouvelle forme de médiation culturelle.

Premièrement, le but du projet peut être avant tout de recueillir des témoignages, par exemple de manière à préserver une mémoire appelée à disparaître. Il peut s'agir de personnes habitant une certaine zone géographique, qui ont fréquenté une certaine institution ou encore qui ont vécu tel événement. Dans ce cas, il est souhaitable de prendre en compte le fait qu'une partie de ce public n'est peut-être pas très connectée. À l'instar de ce qui a été fait par *Britain from Above*, des activités peuvent ainsi être mises en place comme des ateliers d'initiation ou encore la publication de tutoriels en ligne. D'autre part, il est également possible d'intervenir au sein du dispositif même, par exemple en invitant explicitement les visiteurs de passage sur le site à partager leurs souvenirs.

Deuxièmement, le projet peut viser en priorité une communauté plus experte, dans le but de produire une grande quantité de métadonnées de qualité. Les contributeurs visés sont alors plutôt des chercheurs, des étudiants et des amateurs passionnés qui ont développé une certaine expertise. Il s'agit cette fois d'attirer ce public très spécifique et de lui donner envie de contribuer en utilisant un ton adéquat, par exemple en renonçant à donner une dimension trop ludique au projet. Les attentes de l'institution en termes de qualité sont sans doute plus élevées que pour le cas de figure précédent. Dans cette optique, des actions peuvent être envisagées en vue de garantir la pertinence et qualité des métadonnées : création d'un protocole visant à la publication de contributions plus structurées, mise en place d'un processus de validation (par les tiers ou par l'équipe en charge du projet), mise en valeur des contributions les plus riches, système encourageant la justification basée sur des sources externes, *etc.*²

1. Bien entendu, il est possible qu'une institution se fixe plus d'un objectif.

2. Propositions principalement inspirées de Moirez *et al.* (2013).

Troisièmement, l'institution peut envisager le *crowdsourcing* comme un dispositif de médiation plutôt que comme un outil de production. La démarche est alors plus expérimentale, il s'agit de proposer aux visiteurs en ligne une nouvelle forme d'engagement avec les collections en les invitant à faire part de leurs commentaires. Il peut s'agir d'appréciations subjectives, de descriptions, d'une énumération de pensées que leur évoque l'œuvre, des questions qu'ils se posent, de souvenirs, *etc.* La plateforme est alors pensée en termes de communication et de médiation. Elle doit réussir à susciter la curiosité et la participation du plus grand nombre.

7.3 Ressources humaines

Les ressources humaines font partie des paramètres pouvant venir atténuer le succès d'un projet en cours. C'est pourquoi il est nécessaire que l'équipe responsable ait conscience de l'implication que nécessite le suivi d'une telle entreprise : au quotidien, mais également à plus long terme. En effet, si un projet de *crowdsourcing* s'essouffle (déserté par les responsables ou carrément stoppé sans préavis), ce laxisme pourrait avoir des répercussions négatives pour l'institution. Les utilisateurs n'hésitent d'ailleurs pas à utiliser les commentaires pour faire savoir leur mécontentement ou leur déception. Parcourir ce type de commentaires se révèle très instructif. Il y a par exemple un véritable débat qui agite le compte de la *National Library of Ireland* sur *The Flickr Commons*. Cet exemple a tout d'un cas d'école et offre une illustration très concrète des réactions qu'entraîne la mise en veille d'un tel projet. Le point de départ : une photographie postée par la bibliothèque irlandaise en février 2014, accompagnée de ce message :

We wanted to let you know that we're pausing our posting of images on Flickr for the moment, while we integrate the information generated over the last few years into our catalogue, and identify how to resource the project from now on. We're so grateful to all of you who have taken the time to share, comment and research these images over the last few years. All our digitised content - these photographs and many thousands more - is also available on our online catalogue. [...]

La photographie concernée³ comptabilise plus de 150.000 vues et a suscité de vives réactions (plus de 130 commentaires). Les commentaires, plus ou moins virulents, plus ou moins empreints de compréhension, prennent parfois la forme d'un véritable appel au soulèvement, comme en témoigne la première réponse ayant suivi l'annonce de « pause » :

3. <http://www.flickr.com/photos/nlireland/12584715825> (consulté le 10.05.15)

Having read the above I understand where you are and I am sure you took a long time to make this decision. That said - I believe you made a bad decision. Any other Library in the world would give their right arm to have a commons presence half as good as you have. Your operation is world class and a great credit to the NLI [Nation Library of Ireland] and all who work there. There must be a solution that would allow the photostream continue, could the regular commentators help? I would ask the people who regularly look at this photostream to join me in asking you to reconsider this decision. [...] *DannyM8 [E28]*

Les auteurs de commentaires relèvent unanimement la grande érudition dont font preuve les contributeurs et le précieux apprentissage informel que permettent ces espaces de co-construction de savoir. Plusieurs contributeurs confient, dans un véritable cri du cœur, s'être pris de passion pour cette collection : « Please don't stop this project! I am just an ordinary person in the USA, but have become fascinated with your beautiful country and her history. Please reconsider your decision to take a break » *[E29]*.

Nombreux sont ceux qui expriment leur désapprobation (« Monumental mistake ») et proposent ensuite leur aide si cela peut permettre au projet de perdurer, expliquant qu'ils ont tout à fait conscience de l'investissement que cela nécessite (rechercher, choisir, numériser, *uploader*, modérer).

Parmi les autres arguments avancés pour que l'équipe reconsidère sa décision, le rappel de la menace qui plane... la disparition de certains souvenirs : « The greatest enemy of data gathering on old photographs is Time. [...] I can only add that every week that goes by must make it more difficult to gather personal information and reminiscences. » *[E30]* Certains n'hésitent d'ailleurs pas à l'exprimer de façon plus crue : « There is a limited time to gather the information from the generation that have been helping you with this project on Flickr... they will die. At this time their memories go back to the 1950s and provide a link to the changes that have occurred at many of the locations photographed. » *[E31]*

Finalement, le sentiment qui domine est celui d'une amertume généralisée, accompagnée d'une méfiance concernant la réelle motivation de l'institution à poursuivre le projet.

Mise à jour du 16.05.2015 : Depuis la rédaction de cette section, en novembre 2014, la situation a évolué. En janvier 2015, l'institution a repris son activité sur *The Flickr Commons* et depuis, plus de 100 nouvelles photographies ont été ajoutées... grâce au soutien de contributeurs volontaires!

All breaks must come to an end, and thanks to the determination of some marvellously dedicated Flickroonies we are delighted that we'll be resuming posts from today. Different volunteer contributor will be posting images supplied by

ourselves, and responding to your comments, so play nice while they're getting their virtual feet under the online table! Seriously, we're so grateful to the community for making it possible to get going again, and looking forward to working together on this great project⁴.

7.4 Modération et animation

L'exemple précédent met en lumière la nécessité d'une certaine constance. Il renvoie également à la question du rôle des équipes. Plusieurs cas de figure sont envisageables, avec divers degrés d'implication pour les membres de l'équipe. Selon le type de plateforme, il sera question de modérer et/ou d'animer les échanges et découvertes.

Tout d'abord, il arrive que l'interlocuteur qui s'adresse aux utilisateurs soit une personne clairement identifiée. C'est le cas par exemple pour le projet *Britain from Above* : en marge du message que postent les membres de l'équipe, figurent leur avatar, leur nom et prénom, ainsi que leur fonction au sein du projet. D'autre part, ils signent leurs commentaires par leur nom. En ce qui concerne le *L@boratoire des Internautes*, c'est le scénario contraire : les messages sont sobrement signés « Archives Vendée £ »⁵ et le ton est impersonnel : là où le « je » domine du côté de *Britain from Above*, c'est le « on » qui prime du côté du *L@boratoire des Internautes*.

D'autre part, le rôle pour lequel opte l'équipe peut aussi varier. Selon les cas, l'équipe peut se contenter d'ajouter de nouvelles images, elle peut agir en tant que modérateur (supprimer les spams, les messages politiquement incorrects, les doublons *etc.*) ou alors, elle peut avoir un rôle plus affirmé en revêtant le statut d'animateur. Elle aura dès lors un comportement pro-actif, veillera à remercier les utilisateurs pour leur contribution, recentrera la discussion si celle-ci s'enlise dans des débats stériles ou s'égaré dans des débats hors-sujet, encouragera les participations si certains documents sont délaissés, synthétisera les différentes découvertes qui ont été faites si les échanges sont particulièrement longs et denses, *etc.*⁶.

Cette implication du personnel permet de stimuler la participation et d'agir de façon directe sur l'indispensable motivation de la communauté d'utilisateurs, mais elle sous-entend bien sûr un plus lourd investissement de l'équipe.

4. <http://www.flickr.com/photos/nlireland/16038214197/> (consulté le 16.05.15)

5. Avec toutefois quelques variations de l'orthographe.

6. Parfois, un membre de l'institution passe du statut de modérateur à celui de contributeur, allant jusqu'à mener l'enquête sur le terrain... « Comme le suggérait 00h20, le Père Murzeau a été approché. Il nous donne quelques éclairages qui risquent de mettre un point final à cette recherche. Il confirme qu'aucun des deux prêtres non identifiés ne peut être Maurice Tiberghien, né en 1899 et alors trop jeune. [...] » (Archives Vendée)

7.5 Destin des métadonnées

Globalement, le langage utilisé dans les commentaires (langage naturel) est très éloigné du langage scientifique que requiert la création de métadonnées « professionnelles », en témoigne par exemple le faible pourcentage de commentaires comportant des éléments techniques ou un jargon particulier (voir paramètre 6). Ce constat peut être perçu négativement ou positivement par l’institution, selon le point de vue adopté.

Négativement, car cela signifie qu’il existe un véritable décalage entre les métadonnées dites « sociales » (Moirez *et al.*, 2013) et les métadonnées « professionnelles ». Cet état de fait implique qu’il n’est pas réaliste de simplement déléguer leur production à des internautes en espérant obtenir le même résultat.

Positivement, parce que le fait que les commentaires soient rédigés dans un langage naturel relativement spontané peut également être considéré comme une preuve de succès. En effet, proposer un espace de parole qui soit accessible aux non-spécialistes est l’un des paris que certaines institutions ont l’ambition de relever.

La question du lieu de destination de ces métadonnées faisait partie des enjeux évoqués en première partie. En effet, l’institution devra s’interroger : quel destin vont connaître les contributions du grand public ?

Lorsque les contributions des utilisateurs sont destinées à être intégrées dans le catalogue en ligne, il s’agit également d’avoir des métadonnées rédigées dans un langage qui se rapproche le plus possible de celui des utilisateurs. Ceci afin d’avoir une plus grande adéquation entre les termes de recherches utilisés par le citoyen lambda et les termes décrivant le document dans le catalogue⁷(Moirez, 2014).

S’il est question de tags, il est facilement envisageable d’intégrer ces mots-clés au catalogue en ligne de l’institution. C’est d’ailleurs ce que fait le *Powerhouse Museum*⁸. À côté du document sélectionné, deux encarts présentent les tags proposés par des utilisateurs. Le premier regroupe les tags qui ont été ajoutés par des contributeurs directement sur le site de l’institution, tandis que le second encart reprend les tags issus de *Flickr*.

7. Une illustration de ce fossé sémantique pourrait être ce cas de figure iconique qui a émergé dans le cadre du *Steve project* : « An example will suffice to illustrate the discrepancy between official museum cataloging and the cultural literacy of everyday visitors : one of the historical impetuses for the Steve project museum was the realization that a web search for “Impressionism” on the Metropolitan Museum of Art’s collections site would have omitted most of the institution’s holdings. All of the paintings corresponding to that term were listed in the Met’s collections management database as “French”, “19th century”, and “oil on canvas” – but there was no field for “art movement”, and hence there were only scattered returns on the word most educated visitors would use to find them. The simple theory behind Steve and other efforts at cataloging by crowds is that, if museums use terms submitted by visitors to tag their artworks, other visitors will have an easier time finding them. » Trant (2010)

8. <http://www.powerhousemuseum.com/collection/database/menu.php> (consulté le 10.05.15)

En revanche, lorsqu'il s'agit de commentaires, c'est-à-dire de métadonnées non-structurées, qui sont parfois constituées de plusieurs paragraphes, la question de leur intégration ne va pas de soi. Des questionnements peuvent émerger concernant la façon de présenter ces commentaires et le rôle qu'il faut leur attribuer. De plus, sachant qu'il s'agit de métadonnées de qualité inégale, les personnes à l'origine du projet seront peut-être amenées à se demander s'il est plus pertinent d'effectuer un tri parmi les contributions. L'une des solutions envisageables consiste à intégrer les données les plus pertinentes à la description préexistante du document, comme le fait le personnel des Archives de Vendée lorsqu'une enquête est clôturée. Cependant, si l'institution doit imaginer un processus de relecture, de vérification et de tri des commentaires, cela pourrait venir invalider l'argument principal du *crowdsourcing*. En effet, la gestion des données issues du *crowdsourcing* viendrait finalement alourdir le cahier des charges (des membres de l'institution) au lieu de représenter un gain de temps et/ou de permettre des économies financières.

CONCLUSION

Le point de départ de ce mémoire était la volonté d'une prise de distance face au succès croissant du *crowdsourcing* dans le secteur culturel. Afin de pouvoir dégager les possibilités et limites de cette production de métadonnées par le grand public, ce travail s'est articulé autour de deux parties.

La première, plus théorique, fut l'occasion de saisir dans quel contexte s'inscrit ce phénomène et de mieux comprendre quels sont ses enjeux, avant d'aboutir sur un état de l'art présentant une sélection des recherches marquantes effectuées durant la dernière décennie.

La seconde partie traitait plus spécifiquement des commentaires produits par le grand public. L'analyse comparative de trois échantillons a permis de confronter les commentaires issus de projets de différentes ampleurs : *Le L@boratoire des Internautes* (administré par Les Archives départementales de Vendée), le compte du *Powerhouse Museum* sur *The Flickr Commons* et la plateforme web *Britain from Above* (mise en ligne par *The English Heritage*). Cette analyse s'est structurée en deux volets. Le survol de données quantitatives a précédé une analyse qualitative plus poussée, dans le cadre de laquelle près de 900 commentaires ont été traités. Cette recherche s'est organisée au travers d'une grille d'analyse composée de 25 paramètres, qui étaient répartis en huit angles d'approche : apports personnels, langage, expertise, interactions humaines, information, « méta »-information et spam.

Les résultats, denses et nuancés, ont permis de dresser de multiples constats. Le plus frappant concerne sans doute le fait que le type de plateforme choisie par l'institution influence parfois davantage les pratiques que la taille du corpus. Ainsi, le cas B affiche des résultats parfois tout à fait étonnants qui semblent directement découler du fait que le projet prenne place sur *The Flickr Commons*. L'interprétation des résultats a servi de base à un ultime chapitre, qui énonce des pistes de réflexions et des recommandations destinées aux structures culturelles désireuses de mettre en place de tels projets.

Cette prise de recul, étayée par une recherche faisant appel à des données empiriques, permet finalement de relever trois grands paradoxes :

Premièrement, les cas étudiés montrent que l'utilisation d'un corpus de taille modeste favorise la production de métadonnées de qualité. Or, le recours au

crowdsourcing est généralement motivé par la nécessité de traiter un grand nombre de données. Si le corpus est vraiment réduit, alors quelle plus-value représente l'externalisation vers la foule ? En présence d'une institution dont l'ambition se limite strictement à la production de métadonnées pertinentes, déléguer ce travail à la foule apparaît alors comme une vaine dispersion d'énergie. En effet, s'il s'agit d'un corpus de petite dimension, alors ce travail pourrait être directement effectué par ses propres collaborateurs scientifiques, sans nécessiter la mise en place d'un dispositif de *crowdsourcing*.

Deuxièmement, la qualité des métadonnées produites est très inégale : si certaines sont riches et basées sur une solide argumentation, d'autres sont composées de mots isolés dont l'apport informationnel est nul ou insignifiant. Sachant que l'institution pourrait vouloir intégrer ces nouvelles informations aux catalogues de ses collections, un autre paradoxe émergerait alors : les ressources requises pour extraire, trier et vérifier les commentaires avant leur intégration viendraient peut-être « annuler » les gains (de temps ou financiers) permis par le *crowdsourcing*.

Enfin, il est apparu que des utilisateurs particulièrement actifs sont à l'origine d'un nombre très élevé de contributions (l'utilisateur le plus actif produit entre 12 et 21% des commentaires, selon les cas). En déduisant du total ces contributions ainsi que celles des trois ou quatre autres contributeurs les plus actifs, le nombre de commentaires réellement produits par « la foule » est dès lors beaucoup plus modeste, voire anecdotique.

En définitive, la recherche effectuée au sein de ce mémoire permet de penser que le format du commentaire n'est pas le plus adéquat si l'institution culturelle est principalement motivée par un objectif de réintégration des contenus. La forme non-structurée des commentaires, leur qualité très inégale, la fréquente absence de références et la délicate coexistence d'appréciations subjectives et d'informations factuelles signifient qu'un lourd travail de relecture, de tri et de vérification serait nécessaire avant toute réutilisation de ces contenus. En revanche, il est indéniable que le recours au *crowdsourcing* présente de puissantes qualités qui permettent aux institutions culturelles d'expérimenter à de multiples niveaux. Tout d'abord, cela leur permet de nouveaux types d'interaction avec leurs publics. De plus, en laissant la parole à leurs utilisateurs, elles peuvent en apprendre davantage sur leurs usages et intérêts et ainsi mieux adapter leur offre. C'est également une façon de proposer aux visiteurs une forme inédite d'appropriation des contenus, tout en les encourageant à développer leur curiosité et leurs compétences. Par ailleurs, l'analyse des données a montré que les utilisateurs interagissent entre eux et qu'un véritable phénomène de co-création s'élabore. D'autre part, la création d'un dispositif de *crowdsourcing* représente une façon originale et stimulante de valoriser les collections et d'augmenter leur visibilité.

Enfin, la présence d'amateurs extrêmement érudits et compétents pourrait conduire les institutions à envisager de nouvelles formes de collaboration.

Si cette recherche devait être poursuivie, il serait bien sûr intéressant de comparer les résultats obtenus aux conclusions auxquelles ont abouti d'autres chercheurs. D'autre part, il serait également constructif de pouvoir les confronter aux impressions des contributeurs et des initiateurs de ces projets.

Les motivations des utilisateurs, ainsi que leurs différents profils, mériteraient également que l'on s'y attarde davantage, afin de compléter les résultats déjà obtenus.

Par ailleurs, si de nouvelles analyses devaient être menées, il serait sans doute pertinent de comparer des cas qui partagent davantage de similitudes (taille, type de collections, dispositif de contribution).

En outre, il serait utile et instructif de poursuivre l'exploration des commentaires en traitant automatiquement les données textuelles à l'aide de logiciels comme *Semantria*⁹ ou encore *Unitex*¹⁰. Cela permettrait ainsi de tester de nouveaux outils d'analyse, capables de gérer des jeux de données beaucoup plus conséquents, tout en offrant une perspective nouvelle sur le contenu des commentaires issus du *crowdsourcing*.

Avant de conclure, j'aimerais encore souligner que les opérations de *crowdsourcing* ne sont que l'une des nombreuses manifestations du changement plus profond que vivent les institutions culturelles. Anderson affirme que si le musée (et, par extension, les autres institutions patrimoniales) veut continuer à rester pertinent et essentiel à la communauté, un changement fondamental est nécessaire, tant au niveau idéologique qu'au niveau pratique (Anderson, 2012).

Or, cette évolution soulève parfois des réserves. Lorsque la *National Library of Ireland* a interrompu la publication de nouvelles images sur *The Flickr Commons* sous prétexte de vouloir intégrer les informations déjà récoltées dans son catalogue, elle a laissé les utilisateurs pantois et impuissants. Cet incident possède une forte charge symbolique ; il révèle deux visions qui s'opposent : l'institution culturelle érigée comme un temple, ou comme un forum¹¹. Que désire privilégier l'établissement : un catalogue parfaitement entretenu par les détenteurs du savoir, quitte à interrompre un projet de façon abrupte, ou bien le maintien d'un espace d'expression et d'apprentissage pour le grand public ?

Finalement, en sortant de son mutisme et en acceptant la main que lui tendaient des contributeurs désireux d'apporter leur aide, la *National Library of Ireland* a prouvé qu'un dialogue est possible ; mieux encore : elle a initié un processus de co-création

9. *Semantria* utilise des méthodes de *sentiment analysis* (en français « analyse de la tonalité » (Noiret, 2012)), <http://www.semantria.com>

10. *Unitex* permet une analyse syntaxique automatique de textes rédigés en langages naturels, <http://www-igm.univ-mlv.fr/~unitex>

11. Formulation inspirée de l'article de Duncan F. Cameron, « The Museum, a Temple of the Forum » (Anderson, 2012)

qui ne concerne pas uniquement des documents, mais le dispositif de *crowdsourcing* lui-même. J'espère que cette issue heureuse pourra inspirer d'autres institutions et encourager le développement de nouvelles pratiques collaboratives.

Bibliographie

- ANDERSON, G. (2012). *Reinventing the Museum : The Evolving Conversation on the Paradigm Shift*. AltaMira Press, Lanham (USA). 540 p. (citations : p. 1-8).
- BALTUSSEN, L. B., OOMEN, J., MAARTEN, B. et TIMMERMANS, N. (2013). Open culture data : Opening glam data bottom-up. In PROCTOR, N. et CHERRY, R., éditeurs : *Museums and the Web 2013 : Proceedings*, Portland (Oregon, USA). [En ligne], <http://mw2013.museumsandtheweb.com/paper/open-culture-data-opening-glam-data-bottom-up/>, consulté le 11.04.2015.
- BENHAMOU, F. et THESMAR, D. (2011). Valoriser le patrimoine culturel de la france. Rapport technique, Conseil d'analyse économique. 167 p. (citation : p. 81).
- BENJAMIN, W. (2000). L'œuvre d'art à l'époque de sa reproductibilité technique. In *uvres III*. Gallimard.
- BEUDON, N. (2014). Les enjeux de la médiation numérique dans les institutions culturelles. *Bulletin des bibliothèques de France*, 1(3). [En ligne], http://bbf.enssib.fr/tour-d-horizon/les-enjeux-de-la-mediation-numerique-dans-les-institutions-culturelles_64898, consulté le 05.05.2015.
- BONNEY, R., BALLARD, H., JORDAN, R., MCCALLIE, E., PHILIPPS, T., SHIRK, J. et WILDERMAN, C. C. (2009). Public participation in scientific research : Defining the field and assessing its potential for informal science education. a caise inquiry group report. Rapport technique, Center for Advancement of Informal Science Education, Washington, D.C. (USA). [En ligne], <http://informal-science.org/documents/PPSR%20report%20FINAL.pdf>, consulté le 11.04.2015.
- BOYDENS, I. (2010). Hiérarchie et anarchie. dépasser l'opposition entre organisation centralisée et distribuée? *Les Cahiers du numérique*, 6(3). p. 77-101.
- BOYDENS, I. et van HOOLAND, S. (2011). Hermeneutics applied to the quality of empirical databases. *Journal of documentation*, 67(2). p. 279-289.

- BRABHAM, D. C. (2013). *Crowdsourcing*. Cambridge London. The MIT Press. 138 p. (citations : p. 1-4, 9-18,.
- BRAUER, S. (2014). Britain from above : User engagement between the digital and 'real' world. *In Discovering Collection, Discovering Communities 2014 : Proceedings*. The National Archives and Research Libraries UK, The Library of Birmingham, Cadbury Research Library, University of Birmingham and Arts Council England. [En ligne], <http://www.rluk.ac.uk/about-us/blog/britain-above-user-engagement-digital-real-world/>, consulté le 05.05.2015.
- BRULEAUX, A.-M. et GIOVANNINI, A. (s.d.). Politiques et pratiques de conservation et préservation. [En ligne], <http://www.piaf-archives.org/espace-formation/course/category.php?id=5>, consulté le 05.05.2015.
- CHAN, S. (2008). Commons on flickr - a report, some concepts and a faq - the first 3 months from the powerhouse museum. [En ligne], <http://www.freshandnew.org/2008/07/commons-on-flickr-a-report-some-concepts-and-an-faq-the-first-3-months-from-the-powerhouse-museum/>, consulté le 05.05.2015.
- CLAERR, T. et WESTEEL, I. (2011). *Manuel de la numérisation. Sous la direction de Thierry Claerr et Isabelle Westeel; préface de Michel Melot*. Ed. du Cercle de la librairie. 317 p. (citations : p. 17 et suivantes).
- COURTIN, A. (2014a). Brève introduction au linked open data [appliqué aux institutions culturelles]. [En ligne], <http://fr.slideshare.net/antoinecourtin/brve-introduction-au-linked-open-data-appliqu-aux-institutions-culturelles>, consulté le 05.05.2015.
- COURTIN, A. (2014b). Médiation numérique dans les glams : Crowdsourcing, réseaux sociaux, open data et autres initiatives... [En ligne], <http://fr.slideshare.net/antoinecourtin/mediation-numrique-dans-les-glams-crowdsourcing-rseaux-sociaux-open-data-et-autres-initiatives>, consulté le 11.04.2015.
- DETERDING, S., DIXON, D., KHALED, R. et NACKE, L. (2014). Questionner les mises en forme ludiques du web : gamification, ludification et ludicisation. du game design au gamefulness. [En ligne], <http://www.sciencesdujeu.org/pdf/7278.pdf>, consulté le 15.05.2015.
- DUNN, S. et HEDGES, M. (2013). Crowd-sourcing as a component of humanities research infrastructures. *International Journal of Humanities and Arts Computing*. p. 147-169.

- EARLE, E. F. (2014). Crowdsourcing metadata for library and museum collections using a taxonomy of flickr user behavior. mathesis, Cornell University, Cornell (USA). [En ligne], <https://ecommons.cornell.edu/handle/1813/36010>, consulté le 15.05.2015, 105 p.
- ESTELLÉS-AROLAS, E. et de GUEVARA, F. G.-L. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2):189–200. [En ligne], <http://www.crowdsourcing-blog.org/wp-content/uploads/2012/02/Towards-an-integrated-crowdsourcing-definition-Estell%C3%A9s-Gonz%C3%A1lez.pdf>, consulté le 11.04.2015.
- FILDES, E. et FOSTER, A. (2015). "what ths fonds?! the ups and downs of digitising tate's archive. *In Museums and the Web 2015 : Proceedings*, Chicago. [En ligne], <http://mw2015.museumsandtheweb.com/paper/what-the-fonds-the-ups-and-downs-of-digitising-tates-archive/>, consulté le 15.05.2015.
- FLICHY, P. (2010). *Le sacre de l'amateur. Sociologie des passions ordinaires à l'ère numérique*. Seuil. 96 p. (citation : p. 7-35).
- FLYVBJERG, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2). p. 219-245.
- FÜSSEL, S. (2008). *Villes du monde*. Taschen, Hong-Kong, Köln, Paris.
- GEIGER, D., ROSEMAN, M. et FIELT, E. (2011). Managing the crowd : Towards a taxonomy of crowdsourcing processes. *In Proceeding of the Seventeenth Americas Conference on Information Systems*, Detroit (USA). [En ligne], http://www.gess.uni-mannheim.de/fileadmin/gess_db/Publications/CDSB/Erscheinungsjahr_2011/Managing_the_Crowd__Towards_a_Taxonomy_of_Crowdsourcing_Processes.pdf, consulté le 15.05.2015, 11 p.
- GORDON-MURNANE, L. (2006). Social bookmarking, folksonomies, and web 2.0 tools. *Searcher*, 14(6). p. 26-38.
- HARVEY, R. (2010). *Digital Curation. A How-To-Do-IT Manual*. Neal-Schuman, New York. 225 p.
- HERITAGE, T. E. (2014). English heritage annual report and accounts 2013-2014. Rapport technique, The English Heritage. [En ligne], https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/332969/eh-ann-rep-accounts-2013-14.pdf, consulté le 05.05.2015, 67 p.

- HOLLEY, R. (2010). Crowdsourcing : How and why should libraries do it? *D-Lib Magazine*, 16(3/4). [En ligne], <http://www.dlib.org/dlib/march10/holley/03holley.html>, consulté le 05.05.2015.
- HOWE, J. (2006). The rise of crowdsourcing. [En ligne], <http://www.wired.com/wired/archive/14.06/crowds.html>, consulté le 11.04.2015.
- HOWE, J. (2008). *Crowdsourcing : Why the Power of the Crowd is driving the Future of Business*. Crown Business, New York. 312 p.
- JENKINS, H. (2013). *La culture de la convergence. Des médias aux transmédia*. Armand Colin, Paris. 336 p.
- JOHNSON, K. (2008). 3,344 people may not know art but know what they like. [En ligne], <http://www.nytimes.com/2008/07/04/arts/design/04clic.html>, consulté le 05.05.2015.
- JOHNSTON, C. (2008). Awesome photo - thanks!! or, what i've learnt from our flickr pilot. [En ligne], <http://librarytechnz.natlib.govt.nz/2008/07/awesome-photo-thanks-or-what-ive-learnt.html>, consulté le 05.05.2015.
- KELLY, K. (2013). Images of works of art in museum collections : The experience of open access. Rapport technique, Council on Library and Information Resources. [En ligne], <http://www.clir.org/pubs/reports/pub157>, consulté le 11.04.2015, 35 p.
- LANIER, J. (2006). Digital maoism : The hazards of the new online collectivism. [En ligne], <http://edge.org/conversation/digital-maoism-the-hazards-of-the-new-online-collectivism>, consulté le 05.05.2015.
- LEBRATY, J.-F. et LOBRE, K. (2010). Créer de la valeur par le crowdsourcing : la dyade innovation-authenticité. *Systèmes d'Information et Management*, 15(3):32. p. 9-40.
- MAAS (2013). Museum of applied arts & sciences annual report 2012-2013. Rapport technique, The Museum of Applied Arts and Sciences. [En ligne], http://maas.museum/app/uploads/2015/01/Annual_Report_12-13.pdf, consulté le 05.05.2015, 70 p.
- MESGUICH, V. et AMAR, M. (2012). *Bibliothèques 2.0 à l'heure des médias sociaux*. Électre-Éd. du Cercle de la Librairie, Paris. 217 p.
- MOIREZ, P. (2013). Bibliothèques, crowdsourcing et métadonnées sociales. *Bulletin des bibliothèques de France*, (5). [En ligne], <http://bbf.enssib.fr/consulter/bbf-2013-05-0032-007>, consulté le 05.05.2015.

- MOIREZ, P. (2014). Bibliothèques, crowdsourcing et métadonnées sociales. Intervention dans le cadre du stage ENSSIB de formation continue "Les évolutions des catalogues" (6 octobre 2014), [En ligne], <http://fr.slideshare.net/archivesmasala/enssib-evolution-descataloguesmetadonneessociales06102014>, consulté le 11.04.2015.
- MOIREZ, P., MOIREUX, J.-P. et JOSSE, I. (2013). État de l'art en matière de crowdsourcing dans les bibliothèques numériques. Rapport technique, BNF. [En ligne], http://www.bnf.fr/documents/crowdsourcing_rapport.pdf, consulté le 11.04.2015, 77 p.
- NAKATSU, R. T., GROSSMAN, E. B. et IACOVOU, C. L. (2014). A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science*, 40(6). p. 823-834.
- NÉROULIDIS, A. (2015). Flickrland 3. [En ligne], <http://phonothèque.hypotheses.org/14364>, consulté le 05.05.2015.
- NIELSEN, J. (2006). The 90-9-1 rule for participation inequality in social media and online communities. [En ligne], <http://www.nngroup.com/articles/participation-inequality/>, consulté le 05.05.2015.
- NOIRET, S. (2012). La digital history : histoire et mémoire à la portée de tous. In *Une introduction aux humanités numériques*. OpenEdition Press, Marseille. [En ligne], <http://books.openedition.org/oep/258>, consulté le 05.05.2015, p. 151-177.
- OOMEN, J. et AROYO, L. (2011). Crowdsourcing in the cultural heritage domain : Opportunities and challenges. In *Proceedings of the 5th International Conference on Communities and Technologies*, CT '11, New York (USA). ACM. [En ligne], http://www.iisi.de/fileadmin/IISI/upload/2011/p138_oomen.pdf, consulté le 11.04.2015, p. 138-149.
- O'REILLY, T. (2005a). Web 2.0 : Compact definition? [En ligne], <http://radar.oreilly.com/2005/10/web-20-compact-definition.html>, consulté le 26.04.2015.
- O'REILLY, T. (2005b). What is web 2.0; design patterns and business models for the next generation of software. [En ligne], <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>, consulté le 26.04.2015.
- OWENS, T. (2013). The metadata games crowdsourcing toolset for libraries archives : An interview with mary flanagan. [En ligne], <http://blogs.loc.gov/digitalpreservation/2013/04/the-metadata-games-crowdsourcing-toolset->

- for-libraries-archives-an-interview-with-mary-flanagan/, consulté le 05.05.2015.
- OWENS, T. (2014). Linked open crowdsourced description : A sketch. [En ligne], <http://www.trevorowens.org/2014/09/linked-open-crowdsourced-description-a-sketch/>, consulté le 11.04.2015.
- PASQUIER, D. (2014). Les jugements profanes en ligne sous le regard des sciences sociales. *Réseaux*, 1(183). p. 9-25.
- PECCATTE, P. (2009). Une typologie du crowdsourcing appliqué aux images. [En ligne], <http://blog.tuquoque.com/post/2009/11/15/Une-typologie-du-crowdsourcing-images>, consulté le 11.04.2015.
- PECCATTE, P. (2012). Photosnormandie a cinq ans - un bilan en forme de faq. [En ligne], <http://culturevisuelle.org/dejavu/1097>, consulté le 05.05.2015.
- PIROLI, F. et CRÉTIN-PIROLI, R. (2011). Web social et multimédia : propriétés d'une relation symbiotique. *Les enjeux de l'information et de la communication 2/2011*, 12(2). [En ligne], http://www.cairn.info/resume.php?ID_ARTICLE=ENIC_HS03_0006, consulté le 11.04.2015, p. 73-82.
- PISANI, F. et PIOTET, D. (2011). *Comment le Web change le monde : des internautes aux webacteurs*. Pearson, Montreuil, 2ème édition édition. 250 p.
- PROULX, S. (2009). L'intelligence du grand nombre : la puissance d'agir des contributeurs sur internet : limites et possibilités. *In Intelligence collective et organisation des connaissances, 7ème colloque du chapitre ISKO France*, Lyon. p. 3-14.
- RENAULT, S. (2014). Comment orchestrer la participation de la foule à une activité de crowdsourcing? la taxonomie des 4 c. *Systèmes d'Information et Management*, 19(1). p. 77-116.
- RIDGE, M. (2011). Playing with difficult objects : Game designs to improve museum collections. *In Museum and The Web 2011 : Proceedings*, Toronto (Canada).
- RIDGE, M. (2012). Frequently asked questions about crowdsourcing in cultural heritage. [En ligne], <http://www.openobjects.org.uk/2012/06/frequently-asked-questions-about-crowdsourcing-in-cultural-heritage/>, consulté le 05.05.2015.

- RIDGE, M. (2014). *Crowdsourcing Our Cultural Heritage : Introduction*. Digital Research in the Arts and Humanities. Ashgate, Farnham (Angleterre). 306 p., (citations : p. 1-9).
- RIEDER, B. (2010). De la communauté à l'écume : quels concepts de sociabilité pour le "web social" ? *tic&société*, 4(1). [En ligne], <http://ticetsociete.revues.org/822>, consulté le 05.05.2015, p. 33-53.
- SANDAHL, J., PARBY, J. I., SMITH, A., THORBEB, J. et BROE, L. K. (2011). Taking the museum to the streets. In TRANT, J. et BEARMAN, D., éditeurs : *Museums and the Web 2011 : Proceedings*, Toronto (Canada).
- SCHENK, E. et GUITTARD, C. (2012). Une typologie des pratiques de crowdsourcing : l'externalisation vers la foule, au-delà du processus d'innovation. *Management international*, 16(1). [En ligne], <http://id.erudit.org/iderudit/1012395ar>, consulté le 26.04.2015, p. 89-100.
- SHIRKY, C. (2005). Ontology is overrated : categories, links and tags. [En ligne], http://www.shirky.com/writings/ontology_overrated.html, consulté le 05.05.2015.
- SIEGAL, N. (2013). Masterworks for one and all. [En ligne], http://www.nytimes.com/2013/05/29/arts/design/museums-mull-public-use-of-online-art-images.html?_r=0, consulté le 05.05.2015.
- SPRINGER, M., DULABAHN, B., MICHEL, P., NATANSON, B., RESER, D., WOODWARD, D. et ZINKHAM, H. (2008). For the common good : The library of congress flickr pilot project. Rapport technique, The Library of Congress. [En ligne], http://www.loc.gov/rr/print/flickr_report_final.pdf, consulté le 05.05.2015, 50 p.
- SUROWIECKI, J. (2004). *The Wisdom of Crowds. Why the Many are Smarter Than the Few*. Doubleday, New York (USA). 336 p.
- SUROWIECKI, J. (2008). Reflections on click! by james surowiecki. [En ligne], <http://www.brooklynmuseum.org/community/blogosphere/bloggers/2008/08/08/reflections-on-click-by-james-surowiecki/>, consulté le 05.05.2015.
- TALLON, L. et WALKER, K. (2008). *Digital Technologies and the Museum Experience : Handheld guides and other media*. AltaMira Press, Lanham (USA). 264 p., (citations : p. x-xii, 110-111).
- TRANT, J. (2010). Tagging, folksonomy and art museums : Results of seteve.museum's research. [Rapport non publié, en ligne], <http://hdl.handle.net/10150/105627>, consulté le 05.05.2015, 108 p.

- van HOOLAND, S. (2006). From spectator to annotator : Possibilites offered by user-generated metadata for digital cultural heritage collections. *In Proceeding of CILIP Conference 2006*, Angleterre. CILIP Conference, University of East Anglia. [En ligne], http://www.academia.edu/2807203/Spectator_becomes_annotator_Possibilities_offered_by_user-generated_metadata_for_image_databases, consulté le 05.05.2015, 15 p.
- VIDAL, G. et CRENN, G. (2010). Les musées et le web 2.0 : Approches méthodologiques pour l'analyse des usages. *In PROULX, S. et RUEFF, J., éditeurs : Web social : mutation de la communication*. Presses de l'Université du Québec. p. 145-158.
- von ANH, L. et DABBISH, L. (2004). Labeling images with a computer game. *In Proceeding of the Conference on Human Factors in Computing Systems, CHI 2004*, Vienne. p. 319-326.

Annexe A

Exemples de commentaires d'utilisateurs

Réf.	Cas A/B/C	Commentaire	Nom de l'utilisateur	URL
E1	A	Bonjour, Vous avez tout à fait raison, il s'agit bien du moulin de la Tranche. De très nombreuses cartes postales le montrant sous toutes ses faces sont actuellement visibles sur le site http://cartes-postales.delcampe.fr . Un meilleur tirage de la photo ci-dessus, avec la même référence, précise qu'il est situé à La Tranche sur Mer. On le trouve aussi dénommé Moulin Vieux, Le vieux moulin du tènement du bourg... Une carte précise même que ce moulin a été bombardé par les anglais en 1816. Cordialement	Jean-Pierre Bouchet	http://laboratoire-archives.vendee.fr/Identifiez-les/Edifices-et-monuments/Savez-vous-ou-se-trouvent-ces-moulins/Moulin-non-identifie-5
E2	B	Great Great Great ... shot ... LOVE it	Nooura	https://www.flickr.com/photos/powerhouse_museum/2633077972
E3	B	wow that is awesome !	Randy806	https://m.flickr.com/#/photos/powerhouse_museum/2581845285/in/favorites-mitrm_da/
E4	B	Absolutely amazing ! Too bad our society does not look like this anymore. I don't like modern architecture.	Tyler Gillihan	https://www.flickr.com/photos/powerhouse_museum/2533268514/
E5	C	My (maternal) grandparents lived at No 1 Park Lane which was then a ham shop. After the start of WWII hams we [sic] not available so they turned the shop into bar (and as Steve Gregory as already mentioned, the fountain was removed). In 1946 (approx.) the film So Well Remembered was made in and around Park Green, and my grandparents provided tea and cakes to the star John Mills and his wife Mary.	Maurice Mottershead	http://www.britainfromabove.org.uk/image/epw016745
E6	C	Eureka	Class31	http://www.britainfromabove.org.uk/image/ew042936
E7	A	L'hypothèse du mariage de Pierre Pellerin et d'Emma Bernard est séduisante. Nous disposons, pour la confirmer ou l'infirmier, d'une information fort intéressante. Maurice Coulais, propriétaire au Breuil, qui fut maire de la commune entre 1941 et 1971 et s'éteignit dans sa centième année, était en effet témoin de la mariée. Un(e) Langonnais(e) le reconnaîtrait-il dans le cortège ?	Philippe Jaunet	http://laboratoire-archives.vendee.fr/Identifiez-les/Scenes-manifestations/2014-02-Noces-au-Langon
E8	B	'Andula' Built at Bartrams/launched 16.12.48 id 5501697.Reg owner. Soc Geral De Comercio Industria @ Transportes. Her Flag.PRT.Lisbon. In collision with, 'Statue of Libert') a Tanker, 8.6.65. Sank 9.6.65 while under tow Lisbon-Casablanca.	brumor	http://www.britainfromabove.org.uk/image/ew022799
E9	B	Thanks Dave, great spot and we'll arrange to swap the image around as well as get the correctly identified location information updated. Yours, Katy Britain from Above Cataloguing Team Leader	Katy Withaker	http://www.britainfromabove.org.uk/image/epw000152
E10	B	According to David Miller in "Charles Kerry's Federation Australia" ISBN 0908197330. These photographs, taken from a high viewpoint, were made by Willen van der Valden firstly for a commission Kerry had to document the streets of Sydney for the Australian Town and Country Journal. He used a wide angle lens and perched on top of a mobile tower. Ref P. 20	John Buxton	https://www.flickr.com/photos/powerhouse_museum/2363524214
E11	A	Au regard de la posture de ces hommes de troupe, et sachant que le port de la tenue militaire ne tolérerait pas plus, il y a un siècle qu'aujourd'hui, le moindre écart, je crains qu'il faille voir ici quelques camarades de retour d'un corso fleuri ou d'une procession de la Fête-Dieu..	le00h20	http://laboratoire-archives.vendee.fr/Identifiez-les/Scenes-manifestations/2014-01-Un-oeillet-a-la-boutonniere
E12	C	The building is still visible on google-earth, the words have faded from the roof!	Beachcomber	http://www.britainfromabove.org.uk/image/epw000367

FIGURE A.1 – Commentaires d'utilisateurs cités en exemples

Réf.	Cas A/B/C	Commentaire	Nom de l'utilisateur	URL
E13	B	My name is Jim Thompson and my great, great grand father was Captain David Thompson who purchased Clifton Gardens in 1871. He purchased what was then, the "Clifton Arms Hotel" which later became the "Marine Hotel". My family auctioned off the area in 1912. Unfortunately, the "Marine Hotel" was pulled down in 1967 due to the bad crowd it drew at that time. Captain David had built a wharf, pleasure grounds and a dance hall. Thompson St and David St are named after him [...]	Jim Thompson1	https://www.flickr.com/photos/powerhouse_museum/2533270724/
E14	C	Well Done Dave Brunt for getting this one with a twist ie flipped. I had given up on this one after many visits. I am very pleased that we all know where it is as it puts another one to bed. Well done.	Class31	http://www.britainfromabove.org.uk/image/epw000152
E15	B	@Wardie44B - I had another look, and still think the two photos are taken at the same time (within a few minutes). There are differences caused by strong sunshine and contrast in this picture - a building disappeared (was it the Hotel Australia?). This photo is a full plate glass negative; the other is half plate. In this photo the tram wires have been purposely whited out (see persistent URL). The best proof comes from the two carts on the left and kerb (see notes) - the one in front has the same arrangement of parcels in both pictures - see Persistent URL. The tram wires (post December 1899) and the absence of electric lights (pre 1904) should help date the photos. Or maybe I missed something ?!	beachcomber australia	https://www.flickr.com/photos/powerhouse_museum/2362685537
E16	A	Bonjour, je crois qu'il y a un problème avec l'hypothèse Souzy : avez-vous bien vérifié qu'il était amiral ? (Peu probable : il n'était promu capitaine de vaisseau qu'en 1876). Il ne figure nulle part comme tel... en particulier annuaires maritimes de 1877, 1878 et 1879 (au cas où, pour les nominations tardives de 1878).	de Baynast	http://laboratoire-archives.vendee.fr/Identifiez-les/Personnes/2011-07-Cet-amiral-serait-il-Henri-Libaudiere-1820-1892
E17	A	Le polonais est une langue slave avec un alphabet latin agrémenté de diacritiques, et des digraphes (ch cz dz ...) qu'on ne voit pas dans le texte. Celui-ci est en cyrillique. Dans ce texte écrit en cursive on remarque que tous les "m" sont surmontés d'un tiret. On retrouve ce tiret dans le cyrillique macédonien et serbe.	MARQUET	http://laboratoire-archives.vendee.fr/Identifiez-les/Personnes/2014-05-Un-soldat-polonais
E18	C	A bit of an educated(?) guess but I'm thinking that this is an open cast coal mine site. My reasoning is : (a) Appearance in some of this set of images of a dark seam well below the original ground level and this dark material (coal?) being removed by lorry; (b) LAING contractors sign on some earthmoving equipment (John Laing Son had a contract for some such work round about 1948. The location is far less certain but one POSSIBLE area is Heanor/Carrington Farm /Shiple Hall/Mapperley/Smalley area, west of Nottingham. There were a few open-cast mines being worked here in the late 1940s. One of them - Carrington - used a very large dragline excavator, on two 'feet' rather than caterpillar tracks. To see these, search BritainfromAbove using "opencast" as the key word.	ewnmcg	http://www.britainfromabove.org.uk/image/eaw015494
E19	A	Les deux photos m'ont fait "tilt" .Sur la première je reconnais parfaitement Louis Victor baron Alquier vice amiral,2-12-1831/13-11-1906, à la carrière militaire prestigieuse,fut maire de la Flocelliere,dans sa propriété du château de la Bergeliere à la Flocelliere.Sur la seconde photo, de gauche à droite, Laurence, contesse de Froidefond, née de Hillerin,Louis ou Albert coupvent des Graviers, Clotilde Abaudret,Laure Alquier, Soeur de l'amiral,Marie louise valérie marquise de Hillerin, née Alquier,mère de Laurence, soeur de l'amiral,assis, l'amiral Louis Victor Alquier,photo prise aux environs de 1890, au château de la Bergeliere à la Flocelliere,je possède des reproductions de photos prêtées par des descendants de familles de l'ancienne bourgeoisie Flocéenne, Je suis tres intéressé pour visionner le reste du lot,étant administrateur à l'association la Boulite à la Flocelliere. Michel Rambaud.	RAMBAUD MICHEL	http://laboratoire-archives.vendee.fr/Identifiez-les/Personnes/2014-11-Des-photographies-du-Dr-Epron-de-La-Chataigneraie

FIGURE A.2 – Commentaires d'utilisateurs cités en exemples

Réf.	Cas A/B/C	Commentaire	Nom de l'utilisateur	URL
E20	B	Students of Victorian photo art history might find this of interest - this looks suspiciously like a tourist set up studio style photo - they were becoming very popular in the late Victorian era. Clues are : the unruffled hats and tidy bows, neatly turned out fashionable day wear dresses, with lace, women in stays and perhaps even corsets, and not sports clothing. The group shows no signs of exertion, or cold weather sporting. Very similar to other tourist photos using false scene backdrops in studios - EX : families sittings in boats with tropical ocean scene behind them. British Queen Alexandra daughters made one of that style in late 1880s - Also suspect is the difference in perspective and resolution from the men and buildings in the background in respect to the subjects in the foreground. Wonder if anyone else noticed - Pam Eisenberg, photo archivist - OC Register, Santa Ana, CA	Palma J. Eisenberg	https://www.flickr.com/photos/powerhouse_museum/2554627055/
E21	B	Sometime after Decemebtr 899 when the trams were introduced to George Street and 1907-08 when the span poles were removed to enable wide-bodied trams to run in the street.	John Cowper	https://www.flickr.com/photos/powerhouse_museum/2376884726/
E22	A	Tout à fait d'accord! Allez sur http://www.delcampe.net/page/item/id,151482512,var,85-SAINT-MICHEL-MONT-MERCURE-VIEUX-CASTEL-DE-LA-BONNELIERE,language,F.html et vous y verrez le même cliché du même photographe avec la bonne légende. C'est un site de vente de cartes postales.	Dominique Pénisson	http://laboratoire-archives.vendee.fr/Identifiez-les/Edifices-et-monuments/2014-09-Un-chateau-vendeeen
E23	C	Much of the area is covered by the image EPW050203 which is of New housing estates and adjacent countryside, Old Basford, from the south-west, 1936	Class31	http://www.britainfromabove.org.uk/image/epw050204
E24	B	Here's Fred Leist's Sydney Mail illustration of the Vincennes at Manly (see pp. 1490-1491) - news.google.com/newspapers?id=sYNVAAAIBAJ&sjid=zZUDA... ; the main Sydney Mail coverage had appeared in the previous week's issue.	quasymody	https://www.flickr.com/photos/powerhouse_museum/2801040719/
E25	A	Comme il s'agit d'un petit éditeur il a tout naturellement travaillé avec une grosse maison d'édition. Jules Robuchon a ses débuts travaille avec Neurdein également à Nancy et on trouve sur ces cartes ce petit cachet rond. Je serais tenté de dire (si le derrière de la carte est non divisé) que la carte (en fonction de la grandeur de la photo sur le Recto qui laisse un blanc) que nous sommes à la fin de la période précurseur vers 1902 . Si la carte est divisée au verso nous sommes au début de la période dite de " l'âge d'or" et Phelippeau ne s'est pas encore mis à la mode (vers 1905) Il ne s'agit ni d'un mariage paysan ni d'un mariage aristocratique mais d'un mariage bourgeois. Et au Langon il n'y a qu'un petit "château" maison bourgeoise . Dans ce cas on pourrait avoir affaire au mariage d'une Demoiselle DEVERTEUIL ? peut être avec le docteur Eugène GOURMAUD ?	Maurice Bedon	http://laboratoire-archives.vendee.fr/Identifiez-les/Scenes-manifestations/2014-02-Noces-au-Langon

FIGURE A.3 – Commentaires d'utilisateurs cités en exemples

Réf.	Cas A/B/C	Commentaire	Nom de l'utilisateur	URL
E26	C	<p>Dear all,</p> <p>As this image is currently the most commented-upon picture on the site, I thought I'd take a closer look prompted by all your contributions.</p> <p>Clearly we need to change the spelling of Hayes in our caption - so I've put that in hand and the change will appear here in due course (probably at the next data upload scheduled for late October/early November).</p> <p>On the 1877 25" Ordnance Survey map of the area the road name is "Heys Lane" and shows very clearly the different route that John Ellis describes. The lane bent round The Heys and Heys View. Heys View has a large area of glasshouses on the 1898 25" map. By the time of the publication of the 25" map in 1910 the old lane was still there but the newer, wider, straighter course - labelled as Heyes Lane - has been driven through. "The Heys" has gone by the third revision of 1935 and Heys View is then marked as "Fairlie Nurseries".</p> <p>I wonder if, given the range of detailed knowledge here, anyone would be interested in writing an entry in the Wiki tab for this image; which would help to bring all the information together in a history.</p>	Katy Withaker	http://www.britainfromabove.org.uk/image/epw017453
E27	B	Such an amazing BW shot ! Please accept this invitation to [image] Black and white with some creative art work [Hyperlien]	NBAH M	https://www.flickr.com/photos/powerhouse_museum/2555451906/
E28		<p>Having read the above I understand where you are and I am sure you took a long time to make this decision.</p> <p>That said - I believe you have made a bad decision.</p> <p>Any other Library in the world would give their right arm to have a commons presence half as good as you have.</p> <p>Your operation is world class and a great credit to the NLI and all who work there.</p> <p>There must be a solution that would allow the photostream continue, could the regular commentators help?</p> <p>I would ask the people who regularly look at this photostream to join me in asking you to reconsider this decision.</p> <p>p.s. Nice Dog</p>	Danny M8	https://www.flickr.com/photos/nlireland/12584715825
E29		Please don't stop this project! I am just an ordinary person in the USA, but have become fascinated with your beautiful country and her history. Please reconsider your decision to take a break	Meglatron	https://www.flickr.com/photos/nlireland/12584715825
E30		The greatest enemy of data gathering on old photographs is Time. [...] I can only add that every week that goes by must make it more difficult to gather personal information and reminiscences. Great photo and a great doggy yarn.	Mike	https://www.flickr.com/photos/nlireland/12584715825
E31		I will add that I worked in a museum and too much of the collection remained archived and not seen. There is a limited time to gather the information from the generation that have been helping you with this project on Flickr... they will die. At this time their memories go back to the 1950s and provide a link to the changes that have occurred at many of the locations photographed.	Vab2009	https://www.flickr.com/photos/nlireland/12584715825

FIGURE A.4 – Commentaires d'utilisateurs cités en exemples

