



HAL
open science

Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet

Mathieu Andro, Imad Saleh

► To cite this version:

Mathieu Andro, Imad Saleh. Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet. Colloque International sur le Document Numérique, CIDE 17, Nov 2014, Fès, Maroc. <hal-02742735>

HAL Id: hal-02742735

<https://hal.inrae.fr/hal-02742735v1>

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet

Mathieu ANDRO (1,2), Imad SALEH (2)

(1) DV IST, Institut National de la Recherche Agronomique - mathieu.andro@versailles.inra.fr

(2) Paragraphe, Université Paris8

Andro, M., Saleh, I. (2014). Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet. in ZREIK Khaldoun, AZEMARD Ghislaine, CHAUDIRON Stéphane, DARQUIE Gaétan. Livre post-numérique : historique, mutations et perspectives. Actes du 17e colloque international sur le document électronique (CiDE.17), 2014, 152 p.

Résumé.

Cet article est la synthèse d'une thèse doctorale en sciences de l'information et de la communication, commencée en 2012, sur le sujet du crowdsourcing appliqué aux projets de numérisation du patrimoine culturel et, en particulier, de constitution de bibliothèques numériques. C'est dans ce cadre qu'un état de l'art préalable à toute expérimentation, a été réalisé, à partir de la littérature académique et professionnelle sur le sujet, rassemblée au sein d'un corpus de 120 publications, le plus souvent en anglais. Cet article résume et synthétise, dans une première partie théorique, les publications relatives à la définition, à la taxonomie, à la philosophie et à la critique du crowdsourcing et, dans une deuxième partie, les publications sur la conduite de projets, la communication, les motivations des contributeurs et sur des initiatives spécifiques de crowdsourcing, de gamification et de crowdfunding appliquées à des projets de numérisation du patrimoine culturel.

Abstract.

Digital libraries and crowdsourcing: a review of the international academic and professional literature

This paper is a summary of a PhD in information and communication sciences, begun in 2012, on crowdsourcing applied to digitization of cultural heritage and digital libraries. A state of the art on 120 publications from academic and professional literature is proposed. In a first theoretical part, definition, taxonomy, philosophy and criticism of crowdsourcing are introduced. In a second part, project management, communication, motivations of contributors, specific initiatives, gamification and crowdfunding are presented.

Mots clés. Crowdsourcing, crowdfunding, numérisation à la demande, correction participative de l'OCR, folksonomie, gamification, numérisation, bibliothèques numériques.

Keywords. Crowdsourcing, crowdfunding, digitization on demand, OCR correction, folksonomy, gamification, digitization, digital libraries.

Les humains passent de plus en plus de temps sur Internet. Avec le développement du web 2.0, ils sont désormais en capacité de contribuer activement au développement de contenus au lieu d'en être les consommateurs passifs. Si on considère, d'une part, que l'encyclopédie participative Wikipédia a bénéficié de près de 100 millions d'heures bénévoles en 2011 et, d'autre part, que les américains regardent la télévision 200 milliards d'heures par an, ils pourraient donc créer 2000 projets comme Wikipédia chaque année plutôt que de regarder la télévision (Shirky, 2010).

De leur côté, les bibliothèques disposent de moins en moins de ressources pour effectuer le travail nécessaire à l'achèvement de leurs projets documentaires. Elles pourraient donc, au lieu de sous-traiter une partie de leurs tâches auprès de prestataires ayant recours à de la main d'œuvre dans des pays à bas coûts, externaliser auprès de la foule des internautes, les tâches qui ne peuvent être exécutées automatiquement par des programmes et des algorithmes. Cette foule d'internautes compte des spécialistes dans tous les domaines et des individus susceptibles de s'engager pour des raisons aussi diverses que le développement personnel, la distraction, le jeu, l'autopromotion ou l'altruisme. Ces individus pourraient répondre aux appels à participation des bibliothèques qui bénéficient d'une bonne image, et disposent d'une tradition de bénévolat au service de l'intérêt général. Ils pourraient ainsi apporter travail, compétences, connaissances, créativité mais aussi argent et contribuer au développement de projets de numérisation pour la sélection des documents à numériser, pour la numérisation elle-même, pour la correction de l'OCR, le catalogage, l'indexation et la valorisation éditoriale. Ils pourraient même remplir des objectifs qu'il aurait été impossible d'imaginer et d'atteindre auparavant.

Origines du mot et définition

Le terme de crowdsourcing, qui pourrait être traduit en français par approvisionnement ou externalisation par la foule des internautes a été proposé par Jeff Howe en juin 2006 dans la revue Wired Magazine. Il pourrait être défini, en s'inspirant des travaux d'une étude spécifique sur ses diverses définitions comme un type d'activité participative en ligne pour laquelle un individu, une institution, une organisation à but non lucratif ou une société propose à un groupe hétérogène et indéfini d'individus de compétences variables, à travers un appel ouvert, la volontaire ou involontaire sous-traitance d'une tâche ou de micro tâches. L'externalisation de cette tâche, et pour laquelle, une foule d'internautes pourraient participer, apportant leur travail, leur argent, leurs connaissances et / ou expériences bénéficie toujours mutuellement à tous les associés. Les usagers recevront la satisfaction d'un type donné de besoins, qu'ils soient économiques, basés sur la reconnaissance sociale, l'estime de soi, ou le développement de compétences individuelles (inspiré de Estellés-Arolas, 2012 et González-Ladrón-de-Guevara) .

Le crowdsourcing ressemble donc à de l'externalisation d'autant qu'il fait suite à un appel à participation qui est une sorte d'appel d'offres, mais il est non contractuel et ses travailleurs ne sont pas définis. Il se distingue à la fois de l'open innovation car il fait appel à l'extérieur de l'entreprise, du user innovation car l'initiative vient de l'entreprise, non de l'utilisateur et de l'open source car il peut tout aussi bien fonctionner sur un mode collaboratif que sur un mode compétitif.

Historiquement, le crowdsourcing peut trouver sa filiation dans divers événements comme le Longitude Act de 1714 qui proposait 20000 livres à qui permettrait de déterminer la longitude d'un navire en pleine mer, ou comme le concours lancé par Toyota en 1936 afin de définir le dessin le plus pertinent pour le choisir comme logo de sa marque, ou encore comme l'Opéra de Sydney qui fut construit à la suite d'un concours public en 1955 ou enfin, comme le guide de restaurants Zagat rédigé de manière participative en 1979 par un nombre important de collaborateurs.

L'origine philosophique du crowdsourcing visant à réhumaniser Internet et à rendre à l'humain une place centrale sur le web comme origine et finalité, peut aussi bien être considérée comme relevant de l'humanisme chrétien et de l'amour de son prochain, de l'altruisme positiviste, du socialisme et de son slogan « de chacun selon ses capacités à chacun selon ses besoins », de l'anarchisme libertaire et son rejet de l'autorité puisque la contribution de l'amateur et de l'autodidacte devient égale à celle du professionnel et de l'expert, ou encore du libéralisme et de son amour des libertés individuelles et de son esprit d'initiative et d'entreprise d'autant que Internet est aussi parfois qualifié de « libéralisme informationnel » (Loveluck, 2012). Cette confusion des origines philosophiques du crowdsourcing est particulièrement évidente dans le domaine de la gamification. La gamification consiste à récolter les contributions et les données des internautes et à encourager leur participation en les faisant jouer sur le web. La gamification est tout autant l'héritière du stakhanovisme et de l'émulation socialiste qui organise une compétition entre les ateliers et la récompense avec des médailles et des titres que du grand capitalisme américain, du « weasure », mélange du travail (work) et du loisir (leasure) avec son slogan de « fun at work » et qui récompense ses meilleurs employés avec toutes sortes de cadeaux (Nelson, 2012). Selon certains théoriciens, le crowdsourcing permettrait même l'émergence d'une économie participatiste de la contribution, la fin du salariat, la disparition de la séparation entre amateurs et professionnels, entre loisirs et travail, le loisir devenant un travail et le travail un loisir, entre consommation et production, la consommation devenant elle-même productive de valeur, et entre vie privée et vie publique.

Critique du crowdsourcing

D'autres penseurs, moins enthousiastes face à ces perspectives, considèrent que derrière le développement de l'idéologie du web 2.0 et la passion déterministe de ceux qui rêvent de modéliser le monde grâce au « big data », se cacherait un nouveau cauchemar totalitaire où la sphère publique nierait toute vie privée et intime et bafouerait les libertés individuelles. Ce mouvement participerait au relativisme nihiliste et à la négation de toute autorité. Il engendrerait l'exploitation du travail gratuit ou sous payé, qualifié de servuction, car échappant à toute règle, comme sur l'Amazon Mechanical Turk Marketplace qui permet à des sociétés de vendre des micro tâches, pour des « micro salaires », à des travailleurs connectés au réseau en dehors de tout cadre juridique et faisant une concurrence déloyale aux prestataires traditionnels (Fort et al., 2011). D'autres, moins radicaux dans leur critique, promettent toutefois de fiscaliser les données afin de rendre à la population et au

contribuable une partie de la valeur des données qu'ils ont librement produites, sous la forme de « travail invisible », pour YouTube, Facebook ou Google.

Du strict point de vue des résultats, le crowdsourcing pourrait ne pas être l'annonceur promis des lendemains qui chantent. Avec la numérisation participative, par exemple, des documents numérisés de très mauvaise qualité sont susceptibles de dégrader le niveau général et la cohérence des bibliothèques numériques. Le point de vue égocentré ou pire, malveillant des contributeurs anonymes pourrait, quant à lui, être considéré d'avantage comme une menace que comme une réelle opportunité. Les indexations obtenues, n'obéissant à aucune structuration risquent d'être inexploitable, n'obéissant à aucune règle et à aucune hiérarchie. Ainsi, des analyses des tags obtenus via le jeu Waisda? d'indexation des documents audiovisuels de l'Institut néerlandais du son et de l'image montrent que seulement 5,8 % des tags ont une occurrence dans le thésaurus de l'institut et que seulement 23,6 % d'entre eux sont présents dans la base Cornetto des mots de la langue néerlandaise (Oomen et al., 2010). Un travail important de contrôle, de modération et de correction risque donc de remplacer le travail bibliothéconomique traditionnel et, le crowdsourcing ne représenter, finalement, ni une économie de moyens ni une optimisation de résultats. Enfin, du point de vue des professionnels des institutions culturelles, le crowdsourcing pourrait aussi signifier l'appropriation individuelle du patrimoine collectif par quelques internautes se permettant de taguer ou de donner leurs points de vues profanes, informels, personnels, intimes, centrés sur eux-mêmes, banals, lambdas, triviaux et médiocres.

Cependant, l'internaute qui consulte un document numérisé connaît souvent bien mieux son contenu intellectuel puisqu'il en est l'usager, l'utilisateur et parfois même l'auteur, que le bibliothécaire, malgré sa bonne culture générale, le plus souvent littéraire. Il est, en tous cas, généralement plus qualifié pour le faire que le sous-traitant d'un pays à bas salaires comme Madagascar, le Viêt Nam ou l'Inde et qui a finalement été chargé du travail. La qualité de l'information que l'internaute est susceptible d'apporter est donc loin d'être négligeable comme le suggèrent, d'ailleurs, quelques études comparatives (Rorissa, 2010).

Les projets de crowdsourcing et les bibliothèques numériques

Dans le domaine des bibliothèques numériques, quasiment chaque étape de la chaîne de numérisation est susceptible d'être effectuée par des internautes bénévoles. Ainsi, la politique documentaire de sélection, d'acquisition et de numérisation peut être externalisée auprès de la foule des internautes, comme cela est réalisé, par exemple, sur Internet Archive, Commons Wikimedia ou Europeana 1914-1918. C'est le cas aussi de la numérisation à la demande grâce à des financements participatifs, ou crowdfunding, expérimentés par Ebooks on Demand, les amis de la BnF ou plus récemment, en France, par le projet Numalire. La correction participative de l'OCR a, quant à elle, été expérimentée par de trop nombreux projets pour pouvoir tous les évoquer. Nous pouvons toutefois mentionner Distributed Proofreader, Wikisource (Wikipedia), TROVE, Digitalkoot, ReCaptcha. Dans la mesure où la reconnaissance de caractères par des logiciels OCR donne des résultats très variables selon les typographies et l'état des documents originaux et selon les qualités de numérisation, ces projets participatifs permettent, par correction, d'obtenir des textes de meilleure qualité, mieux indexés par les moteurs de recherche et surtout, des fichiers compatibles avec la lecture sur liseuses. Ils évitent aussi de recourir à des prestataires pour corriger l'OCR brute et qui font généralement appel à des pays dont les travailleurs ont des salaires faibles relativement à ceux des pays

occidentaux. Enfin, la redocumentarisation, l'indexation participative ou folksonomie peuvent également être demandées aux internautes comme cela est proposé, par exemple par le [steve.museum](#), [Filckr the commons](#), [Google Image Labeler](#), [ESP Game](#), [Metadata Games](#)...

Parmi tous ces projets, certains comme [Digitalkoot](#), [Google Image Labeler](#) ou [ESP Game](#) font participer les internautes sous la forme de jeux sur le web. Il pourra s'agir, par exemple, avec le jeu [Digitalkoot](#), de ressaisir convenablement chaque mot provenant de l'OCR afin de construire chaque brique d'un pont permettant à une armée de petites taupes de pouvoir traverser une rivière et ce, en évitant un maximum de noyades et le maximum d'erreurs invariablement sanctionnés par l'explosion d'une brique du pont (Chrons et Sundell, 2011). D'autres jeux, comme [Google Image Labeler](#), consisteront à chercher à deviner et à trouver le même mot clé qu'un autre internaute partenaire sur le web pour décrire et indexer une même image afin de marquer un maximum de points, et ce, tout en ayant l'impossibilité d'utiliser certains mots tabous car déjà validés lors de parties précédentes et car cela a pour effet de récolter des indexations plus utiles et plus précises (Von Ahn, 2006). Ce faisant, par un système de confrontation des saisies des internautes, on obtient une indexation de qualité optimale. Ce type de projets fait appel à une démarche qualifiée de gamification, terme proposé en 2002 par Nick Pelling et qui pourrait être traduit en français par celui de ludification. D'autres auteurs comme Luis Von Ahn parlent aussi de « games with a purpose », expression qui pourrait être traduite par « jeux avec une finalité ». Von Ahn évoque aussi l'expression de « human computation » pour désigner la possibilité d'utiliser les cerveaux humains comme des processeurs dans un système distribué au sein duquel chacun pourrait contribuer à générer un calcul massif et obtenir des résultats inimaginables auparavant (Von Ahn, 2006). Le projet scientifique [fold.it](#), a ainsi permis d'améliorer considérablement la connaissance du repliement des protéines grâce à l'intelligence humaine collective mobilisée autour de jeux de puzzles. La gamification permettrait d'améliorer considérablement la participation des internautes dans le cadre de projets de crowdsourcing et représenterait un potentiel important si on considère que 65 % des ménages américains jouent aux jeux vidéos et qu'un américain moyen a déjà joué 10 000 heures à des jeux vidéo lorsqu'il a atteint l'âge de 21 ans (McCarthy, 2012). Selon Luis Von Ahn, l'intégralité des images de [Google Images](#) aurait ainsi pu être indexée en seulement 31 jours par 5000 personnes qui joueraient continuellement à [ESP Game](#) (Von Ahn et Dabbish, 2004).

D'autres projets, comme [reCaptcha](#), bénéficient de la contribution involontaire des internautes (« crowdsourcing implicite ») qui doivent ressaisir les images distordues de deux mots pour prouver qu'ils ne sont pas des robots malveillants (Human Interactive Proof) et pouvoir ainsi créer un compte sur un site web. Ce faisant, ils participent à corriger du texte ocrisé par [Google Books](#), l'un des deux mots à recopier étant un mot ocrisé dans le cadre du célèbre projet de numérisation et non reconnu par un dictionnaire tandis que l'autre mot est effectivement utilisé pour des raisons de sécurité. Avec son slogan « Stop spam, read books », [reCaptcha](#) permet à environ 100 millions de mots douteux d'être ainsi astucieusement corrigés chaque jour par les internautes (Von Ahn et al., 2008), par confrontation de leurs saisies et au bénéfice de projets comme [Google Books](#), [Google Maps](#) et prochainement aussi, [Internet Archive](#).

D'autres projets enfin font appel à d'autres types de ressources des internautes : leurs ressources financières. On parle dans ce cas de crowdfunding pour le projet européen [Ebooks on Demand \(EOD\)](#) ou pour le projet expérimental [Numalire](#) de la société [Yabé](#). A partir des catalogues en ligne des bibliothèques ou de vastes plateformes de métadonnées bibliographiques bien référencées par les

moteurs de recherche, le financement participatif de la numérisation des livres libres de droits est proposée aux internautes, aux institutions et aux mécènes qui peuvent en commander la numérisation via des liens renvoyant vers une interface de paiement. Les documents dont la numérisation est financée sont ensuite numérisés sur place par un opérateur, ou acheminés vers un atelier de numérisation, par navette ou par courrier. Une fois numérisés et mis en ligne ils pourront porter la mention de leurs sponsors et leur offrir un retour sur investissement en termes de publicité et de trafic web vers leurs propres sites qui compensera les coûts investis dans la numérisation. Ce retour sur investissement pourra être particulièrement intéressant dans le cas de livres susceptibles d'être consultés plusieurs millions de fois, comme c'est le cas, par exemple, de quelques livres diffusés sur Internet Archive. L'argent public peut ainsi se concentrer sur des documents dont la numérisation présente un intérêt patrimonial ou scientifique et laisser l'argent privé financer la numérisation de livres présentant un intérêt commercial et susceptibles d'attirer le grand public. De leur côté, les bibliothèques sont désormais en mesure d'offrir à leurs usagers, des services de reprographie numérique, sans avoir à en supporter le coût, et de compléter leur programmes de numérisation. Elles peuvent aussi désormais externaliser le difficile travail d'identification et de sélection des documents qui, au sein de leurs patrimoines documentaires, méritent d'être numérisés et ainsi, ouvrir et partager auprès du grand public, leur politique documentaire.

Les contributeurs de tous ces projets culturels faisant appel au crowdsourcing ressemblent d'avantage à une communauté définie de bénévoles dont le profil est d'ailleurs plutôt celui de retraités passionnés de généalogie ou d'histoire locale et familiale, ou de jeunes diplômés en recherche d'emploi, qu'à une foule d'internautes indifférenciés. Dans ces conditions, il serait donc plus approprié de parler de nichesourcing ou de communitysourcing plutôt que de crowdsourcing, comme le suggère certains auteurs (Causer, 2012). Lorsque ces internautes sont volontaires, ils peuvent l'être pour des raisons aussi bien intrinsèques qu'extrinsèques. Certains cherchent ainsi à se sentir utile pour un groupe, pour la société, pour le pays, pour la science, pour l'intérêt général, ou pour une cause, ou encore faire quelque chose de désintéressé, dans un esprit d'altruisme et de philanthropie ou avec un sentiment de redevabilité. Rose Holley (Holley 2010) mentionne ainsi les catégories suivantes :

- Recherchent plutôt leur développement personnel, à se cultiver et à apprendre, à satisfaire leur soif de connaissance.
- Cherchent à s'amuser, à se distraire, à jouer ou à tester une démarche innovante.
- Stimulés par l'esprit de compétition, le challenge ou par le besoin de se prouver quelque chose.
- Veulent avoir une bonne estime de soi, avoir l'impression d'avoir du pouvoir sur les choses, d'être auteur et acteur ou, tout simplement, cherchent à améliorer leur e-reputation sur le web.

Dans tous les cas, pour les institutions culturelles qui bénéficient de travail bénévole, le développement de ce type de projets nécessite des investissements importants pour communiquer, recruter, faire du community management, alimenter régulièrement la plateforme en contenu, motiver, récompenser, former les contributeurs, modérer, contrôler, évaluer la qualité et surtout, réintégrer les données produites dans les bibliothèques numériques car, dans le cas contraire, le crowdsourcing n'obéirait qu'à une logique de communication institutionnelle superficielle.

En France, quelques projets de transcriptions participatives ont déjà été conduits dans les Archives (Alpes Maritimes, Cantal, Ain, Seine Maritime, Normandie...), ou dans quelques bibliothèques en partenariat avec Wikisource (en 2008, à l'Ecole Nationale Vétérinaire de Toulouse pour 100 thèses numérisées, puis en 2010, à la Bibliothèque nationale de France pour un peu plus de 1400 livres numérisés). Mais le développement du crowdsourcing y demeure particulièrement difficile, dans un pays où les modes de fonctionnement participatifs ont parfois du mal s'imposer. Ainsi, la grande étude menée par l'Online Computer Library Center (OCLC) en 2012 sur le sujet (Smith-Yoshimura et Holley, 2012) ne mentionne aucun projet français sur 76 sites audités dans le monde et il aura fallu attendre février 2013 pour qu'une première étude (Moirez et al., 2013) soit publiée en France sur le sujet, par la Bibliothèque nationale de France dans le cadre du projet Ozalid. Le crowdsourcing est pourtant utilisé dans un nombre croissant de projets culturels dans le monde et est mentionné dans un nombre exponentiel de publications scientifiques.

Comme le montre une récente étude de Deloitte pour Google datée de 2013, 11 % des salariés français utilisent des outils collaboratifs contre 25 % en Hollande. La France semble, en effet, éprouver une difficulté à passer d'une culture de l'information individualiste, fermée, corporatiste, centralisée, procédurière, complexe, hiérarchique et verticale à une culture plus collective, ouverte, partagée et transversale. D'un certain point de vue, la France pourrait apparaître comme demeurant encore largement prisonnière d'une théorie quasi malthusienne de la valeur de l'information qui considère que moins de gens ont accès à la connaissance, plus elle a de valeur. Le culte pour le secret qui enferme l'information dans les seuls réseaux d'initiés y demeure assez répandu malgré le développement de la société de l'information, de l'open access et de l'ouverture des données. Dans les bibliothèques françaises, parfois nostalgiques du monopole d'accès à l'information et craintives de se faire piller par des intérêts privés, le crowdsourcing implique une évolution majeure dans la culture de travail d'une corporation assez conservatrice. Cette corporation pourrait se sentir dévalorisée si son travail était effectué gratuitement et bénévolement par des amateurs, le grand public, les profanes, le secteur privé ou pire, marchand. Cette évolution serait susceptible de remettre en question l'autorité, les compétences bibliothéconomiques et le statut de professionnels déjà sur la défensive.

Conclusions

Si la France envisage tardivement d'expérimenter le crowdsourcing dans le cadre de projets de numérisation du patrimoine, d'autres pays, comme l'Australie et son projet TROVE (Ayres, 2013) constatent, de leur côté, qu'un seuil aurait été atteint et qu'il devient difficile de développer d'avantage cette piste, d'autant que les progrès dans les logiciels de reconnaissance de caractère pourraient rendre caduque, à l'avenir, la correction participative de l'OCR, en particulier. Pour sa part, le géant Google, s'il continue à bénéficier de la correction involontaire de l'OCR par les internautes via les reCaptcha, il a néanmoins abandonné son projet d'indexation des images dans le cadre du jeu Google Image Labeler.

En tout état de cause, les organisations qui mettent en place, les premières, des fonctionnements de type crowdsourcing pourront en tirer un avantage concurrentiel, mais lorsque cette pratique se sera généralisée, il est probable que la gamification et la rétribution deviennent l'un des seuls moyens de capter la participation des internautes. Dans le cadre du travail de recherche doctorale dont cet

article est issu, des expérimentations, réalisées en collaboration avec le projet de crowdfunding Numalire, sont en cours afin d'en évaluer la faisabilité.

La participation des internautes est vitale pour des institutions qui ont grandement besoin d'exister sur le web, de communiquer et de mobiliser autour de leurs collections. L'ouverture sur l'amateur peut représenter, pour elles, comme pour les entreprises, une source importante d'innovations et d'inventions car ne cherchant pas à reproduire les modèles établis du métier avec lesquels les professionnels ont été formés, ils sont susceptibles de provoquer des ruptures innovantes. Ainsi, selon Von Hippel, 46 % des entreprises américaines dans des domaines innovants et qui se maintiennent au moins 5 ans ont pour origine un simple utilisateur (Guillaud, 2012). Mais cette ouverture, dans les institutions culturelles françaises, nécessitera des investissements importants en conduite du changement.

Bibliographie

Andro, M. (2013). "Faire appel à la charité" : Trove, les journaux australiens et la foule des internautes. Traduction française de : Ayres, M.-L. (2013). 'Singing for their supper': Trove, Australian newspapers, and the crowd. In IFLA World Library and Information Congress. Singapore.

Causser, T. (2012). Building A Volunteer Community: Results and Findings from Transcribe Bentham. Digital Humanities Quaterly, vol. 6, num. 2, 26 p.

Chrons, O. Sundell S. (2011). Digitalkoot: Making Old Archives Accessible Using Crowdsourcing. 3rd Human Computation Workshop, HCOMP 2011.

Estellés-Arolas, E. González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information Science, vol. 38, num. 2, 189-200.

Fort, K., Adda, G., Cohen, K. B. (2011). Amazon Mechanical Turk: Gold Mine or CoalMine? Computational Linguistics, vol. 37, num. 2, 413-420.

Guillaud, H. (2012). Von Hippel : le paradigme de l'innovation par l'utilisateur. InternetActu.net, disponible à : <http://www.internetactu.net/2012/06/28/von-hippel-le-paradigme-de-linnovation-par-lutilisateur>

Holley, R. (2010). Crowdsourcing: How and Why Should Libraries Do It?. D-Lib Magazine, vol. 16, num. 3-4.

Loveluck, B. (2012). La liberté par l'information : généalogie politique du libéralisme informationnel et des formes de l'auto-organisation sur internet. Thèse de doctorat de l'Ecole des Hautes Etudes en Sciences Sociales, EHESS.

Mccarthy S. (2012). Using gamification as an effective OCR crowdsourcing motivator. Disponible à <http://sjmccarthy.com/?portfolio=research-proposal-using-gamification-as-an-effective-ocr-crowdsourcing-motivator>.

Moirez, P. Moreux, J.P. Josse, I. (2013). Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques. Livrable L-4.3.1 du projet de R&D du FUI 12 pour la conception d'une plateforme collaborative de correction et d'enrichissement des documents numérisés.

Nelson, M. J. (2012). Soviet and American Precursors to the Gamification of Work. In Proceedings of the 16th International Academic MindTrek Conference.

Oomen, J., Brinkerink, M., Heijmans, L., Van Exel, T. (2010). Emerging Institutional Practices: Reflections on Crowdsourcing and Collaborative Storytelling. J. Trant and D. Bearman (eds). In Museums and the Web 2010: Proceedings. Toronto: Archives & Museum Informatics.

Rorissa, A. (2010). A Comparative Study of Flickr Tags and Index Terms in a General Image Collection. *Journal of the American Society for Information Science and Technology*, vol. 61, num. 11, 2230-2242

Shirky, C. (2010). *Cognitive Surplus: Creativity and Generosity in a Connected Age*. Penguin Books.

Smith-Yoshimura, K., Holley, R. (2012). *Social Metadata for Libraries, Archives, and Museums: Recommendations and Readings*. OCLC Research.

Von Ahn, L., Dabbish, L. (2004). Labeling Images with a Computer Game. In *ACM Conference on Human Factors in Computing Systems, CHI*.

Von Ahn, L. (2006). Games With A Purpose. *IEEE Computer Magazine*, vol. 39, num 6, 96-98.

Von Ahn, L., Maurer, B., Mcmillen, C., Abraham, D., Blum, M. (2008). reCAPTCHA: Human-Based Character Recognition via Web Security Measures. *Science*, num. 321, 1465-1468.