



HAL
open science

Bibliothèques numériques et crowdsourcing : un état de l'art

Mathieu Andro, Imad Saleh

► **To cite this version:**

Mathieu Andro, Imad Saleh. Bibliothèques numériques et crowdsourcing : un état de l'art. Samuel Szoniecky; Nasreddine Bouhaï. Intelligence collective et archives numériques : Vers des écosystèmes de connaissances, 2, ISTE, pp.135-160, 2017, Collection Systèmes d'information, web et société, Série Outils et usages numériques, 978-1-78405-255-3. hal-01573272

HAL Id: hal-01573272

<https://hal.science/hal-01573272>

Submitted on 28 Dec 2018

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.

Bibliothèques numériques et crowdsourcing : un état de l'art

Mathieu Andro et Imad Saleh

Introduction

Saisies de notices bibliographiques, indexation et correction de l'OCR de documents numérisés, les bibliothèques ont souvent externalisé certaines de leurs activités à des prestataires ayant recours à de la main d'œuvre à bas coût dans des pays en voie de développement comme Madagascar, l'Inde ou le Viêt-Nam. Mais elles pourraient désormais plutôt faire appel aux foules d'internautes, c'est-à-dire au crowdsourcing afin de réaliser des tâches pour lesquelles leurs propres forces sont insuffisantes.

Le développement du crowdsourcing en bibliothèques est particulièrement important dans le domaine de la correction de l'OCR. En effet, les logiciels de reconnaissance de caractères qui convertissent les images des pages numérisées de livres en textes ne donnent pas des résultats fiables à 100 % et, en fonction de la qualité du document original, de celle de sa numérisation, de la typographie, de la présence d'éventuelles annotations manuscrites, il sera nécessaire de corriger les textes obtenus avec l'aide de dictionnaires. La correction de l'OCR est nécessaire afin de permettre des recherches en texte intégral plus efficaces dans les textes numérisés, un meilleur référencement des contenus par les moteurs de recherche, la production de fichiers Ebooks aux formats EPUB ou MOBI pour pouvoir être lus sur des liseuses, des extractions de données via des technologies de text mining ou encore des exploitations scientifiques liées à la culturomique.¹

Cette question du recours au crowdsourcing se pose aujourd'hui de plus en plus aux bibliothèques, aux plus grandes d'entre elles comme aux plus petites. Afin d'y apporter des éléments de réponses et afin d'apporter une contribution conceptuelle originale autour du crowdsourcing en bibliothèques, nous avons rédigé le présent état de l'art, lui-même issu d'un travail de thèse.

Il offrira des éléments conceptuels afin de comprendre ce phénomène, une taxonomie et un panorama des projets et des analyses du point de vue des sciences de l'information et de la communication.

4.1. Le concept de crowdsourcing en bibliothèques

4.1.1. Définition du crowdsourcing

Littéralement le crowdsourcing signifie l'externalisation (« outsourcing » auprès de foules (« crowd ») d'internautes selon l'expression de Jeff Howe proposée dans la revue Wired Magazine en juin 2006. Selon une définition du crowdsourcing qui fait autorité et que nous avons traduite et adaptée le crowdsourcing est

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 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 » (traduit de Estellés-Arolas, *Towards an integrated crowdsourcing definition*. Journal of Information Science, 2012. [EST 12])

Contrairement à ces auteurs, nous pensons que le crowdsourcing peut aussi exister sous la forme d'une participation qui n'est pas obligatoirement et strictement volontaire comme c'est le cas de projets où les internautes contribuent en jouant et qui sont qualifiés de gamification. Nous pensons même que le crowdsourcing peut aussi faire appel à la participation involontaire ou inconsciente des internautes, comme c'est par exemple le cas avec le projet

¹ La culturomique pourrait être définie comme l'analyse quantitative de la culture, notamment au travers de grands volumes de textes numérisés.

reCAPTCHA. Les millions de livres numérisés par Google Books sont OCRisés. Les mots n'ayant pas d'occurrences dans des dictionnaires sont ensuite renvoyés vers les internautes qui, pour des raisons de sécurité, sont contraints de ressaisir des mots déformés pour prouver qu'ils ne sont pas des robots. Ce faisant, en créant leurs comptes sur des sites web, ils contribuent involontairement à la correction de l'OCR des programmes Google Books et Google Maps. Nous qualifions cette participation involontaire des internautes de crowdsourcing implicite.

Ayant défini ce qu'est le crowdsourcing, il reste à préciser ce qu'il n'est pas. Le crowdsourcing ne doit pas être confondu avec l'externalisation car s'il y a bien une sorte d'appel d'offres sous la forme d'un appel à contribution, la relation avec le contributeur est non contractuelle. Il ne doit pas non plus être confondu avec le « user innovation » car l'entreprise reste à l'initiative du projet, ni avec l'open source car le mode de contribution n'est pas obligatoirement collaboratif mais peut, au contraire, faire appel à la compétition.

4.1.2. Origines historiques du crowdsourcing

Ce modèle économique trouve sa filiation dans

- les appels au peuple lancés par les gouvernements pour résoudre des problèmes scientifiques contre des récompenses dès le 18^e siècle ;
- les concours et les souscriptions publiques ;
- le libre service et le libre accès qui ont permis au consommateur de prendre en charge une partie du travail du producteur jusqu'au modèle du « à la demande » qui lui a permis de prendre en charge la décision de mise en production elle-même.

Nous proposons ci-après une chronologie des origines historiques du crowdsourcing et des sciences citoyennes. Cette chronologie est le résultat d'une analyse de la littérature sur le sujet. Elle a été constituée à partir d'une collecte des événements les plus significatifs sur le plan des dimensions fondatrices du crowdsourcing (appel à participation, récompenses, travail collectif, microtâches, externalisation, sagesse des foules). Les événements rassemblés sont également les plus représentatifs de la taxonomie que nous proposons dans la suite de l'article.

1714	Le gouvernement anglais lance un appel aux scientifiques afin qu'ils trouvent une solution pour déterminer la longitude en mer d'un bateau. John Harrison, un charpentier et horloger remporte les 20 000 livres de récompense parmi plus d'une centaine de concurrents dont Cassini, Huygens, Halley et Newton
1726	Une ordonnance du Roi de France, Louis XV demande aux capitaines de navires de rapporter des plantes et des graines des pays étrangers qu'ils visitent
1750	L'astronome britannique Nevil Maskelyne calcule la position de la lune pour la navigation en mer grâce à la confrontation des calculs de deux astronomes ayant effectué les calculs deux fois chacun puis ayant été vérifiés par un tiers
1758	Le mathématicien Alexis Clairaut parvient à calculer l'orbite de la comète de Halley en divisant les tâches de calculs entre trois astronomes
1794	L'ingénieur français Gaspard de Prony organisa des microtâches d'additions et de soustractions auprès de quatre-vingt coiffeurs au chômage afin de développer des tables logarithmiques et trigonométriques détaillées
1810	Nicolas Appert reçut 12 000 francs du gouvernement français suite à un appel à contributions pour ses nouvelles méthodes de conservation des aliments qui aboutiront aux boîtes de conserve
1852	Le magasin "Au bon marché" est le premier magasin en libre service. Désormais les consommateurs accèdent directement aux marchandises sans

	passer par l'intermédiaire du marchand et prennent ainsi en charge une partie du travail du producteur
1857	L'Oxford English Dictionary bénéficie, suite à un appel à contributions bénévoles, de plus de 6 millions de documents contenant des propositions de mots et des citations d'usages
1884	La statue de la Liberté est financée par une souscription publique
1893	Francis Galton constate, à l'occasion d'un concours lancé sur un marché de bétail et pour lequel il s'agissait de deviner le poids d'un bœuf, que la moyenne des estimations d'une foule est plus proche de la vérité que l'estimation des experts, laissant supposer l'existence d'une "sagesse des foules"
1900	La National Audubon Society (USA et Canada) organise un décompte annuel des oiseaux, le "Christmas bird count"
1895	Le bibliothécaire James Duff Brown invente le libre accès en bibliothèque. Les lecteurs de la bibliothèque publique de Clerkenwell ont désormais accès directement à une partie des collections
18??	Dans le domaine de l'édition, se multiplient les souscriptions publiques afin de financer la publication de livres
1936	Toyota rassemble 27 000 personnes et choisit le meilleur dessin proposé comme logo de sa marque
1938	Aux Etats-Unis, le Mathematical Tables Project mobilise 450 chômeurs victimes de la dépression économique et pilotés par un groupe de mathématiciens et de physiciens afin de calculer des tables de fonctions mathématiques, bien avant l'invention de l'ordinateur
195?	L'ingénieur industriel de Toyota, Taiichi Ōno, invente le modèle du "juste à temps" ancêtre du modèle "à la demande" et qui permettait de produire sans stocks ni invendus, en flux tendus, en fonction de la demande. Il s'agit, d'une certaine manière, d'externaliser auprès du consommateur la décision de mise en production elle-même. Ce modèle est à l'origine de la numérisation à la demande par crowdfunding et de l'impression à la demande
1954	Le premier Téléthon aux Etats-Unis permet de recueillir un financement pour lutter contre l'infirmité motrice cérébrale
1955	L'Opéra de Sydney est conçu et construit à la suite d'un concours public, qui encouragea les gens ordinaires de 32 pays à contribuer à ce projet de conception
1979	Le sondage Zagat (un guide de restaurants) fonda ses évaluations sur une grande quantité de testeurs. Le projet fut racheté en septembre 2011 par Google
1981	Le guide de voyages Lonely Planet est rédigé, pour sa 3ème édition, de manière participative par des voyageurs indépendants
1996	Naissance de Internet Archive
1997	Le livre à la carte : reproduction par fac similés des livres conservés dans les bibliothèques (numérisation et impression à la demande)
1997	Le groupe de Rock Marillion finance sa tournée aux USA grâce aux dons pour un total de 60 000 \$ de ses fans
1998	L'annuaire Dmoz, propose un contenu généré par ses utilisateurs. Le web 2.0 était né
2000	Les plateformes philanthropiques de <i>crowdfunding</i> justgiving.com et celle de financement participatif d'artistes artistshare.com voient le jour. Elles seront

	suivies par de multiples initiatives jusqu'à aujourd'hui
2000	Distributed Proofreader : premier projet de transcription participative de livres
2001	Naissance de Wikipédia
2003	ESP Game : un jeu pour l'indexation des images
2005	Amazon lance la plateforme de <i>crowdsourcing</i> Amazon Mechanical Turk Marketplace pour ses propres besoins et permet aussi de mettre en relation des sociétés et des institutions à la recherche et des travailleurs sur le web autour de microtâches
2006	Espresso Book Machine pour l'impression à la demande in situ
2006	Jeff Howe propose le terme de « <i>crowdsourcing</i> » dans la revue Wired Magazine en juin 2006
2007	Google Books utilise reCAPTCHA pour faire corriger son OCR brute par les internautes
2008	Le projet de gamification Fold.it permet de progresser, grâce à des jeux de puzzle, dans la connaissance des protéines
2011	Le projet Good Judgement Project mobilise la sagesse de foules d'internautes à travers leurs prévisions géopolitiques qui rivalisent avec celles des experts du renseignement
2011	Digitalkoot pour la correction de l'OCR sous la forme de jeux d'arcade
2013	le jeu video Star Citizen récolte la somme de 30 044 586 \$

Tableau 4.1. Chronologie du crowdsourcing en bibliothèques

4.1.3. Origines conceptuelles du crowdsourcing

Le crowdsourcing trouve sa filiation conceptuelle dans des idéologies aussi diamétralement opposées que les idéologies socialistes, libertaires, humanistes ou libérales [AND 14a], dont l'idéologie californienne réaliserait la synthèse la plus favorable au développement du crowdsourcing.

Idéologies socialistes et marxistes	Production « de chacun selon ses capacités à chacun selon ses besoins » et dépassement de la loi de la valeur, l'argent n'est plus le moteur principal, gratuité, esprit de sacrifice, travail au service de l'humanité, émulation socialiste, stakhanovisme, collectivisme, abolition de la séparation fondamentale entre travail nécessaire et surtravail, remise en cause du salariat et des classes sociales chacun pouvant être tour à tour employeur ou employé, dépassement de la propriété privée par l'usage partagé, « communaux collaboratifs », économie participatiste de la contribution pair à pair
Idées libertaires et anarchistes	Critique de l'autorité des castes dominantes et du totalitarisme, démocratie directe, contribution de l'amateur égale à celle de l'expert, disparition de la frontière entre producteurs et consommateurs, le travail devient un loisir, <i>weisure</i> (<i>work + leisure</i>) ou <i>playbour</i> (<i>play + labour</i>)
Humanisme	Humanités numériques, réhumanisation d'Internet et restitution à l'humain d'une place centrale sur le web comme origine et finalité, confiance dans les capacités de l'homme supérieures à celles des algorithmes, altruisme et amour de son prochain, souci du plus faible face au puissant

Tableau <i>origines</i>	Libéralisme	Externalisation, intégration du consommateur dans la production, méritocratie, accroissement des libertés individuelles, défense de la liberté d'expression, esprit d'initiative et d'entreprise, « <i>fun at work</i> », universalisme, mondialisme, démocratie, main invisible, ordre spontané (Friedrich Hayek) de Wikipedia qui fonctionne par l'action autonome d'individus sans aucune planification, confiance dans le marché, autoentreprenariat, remise en cause des monopoles, « ubérisation »	4.2. Les
Idéologie californienne : philosophie libérale libertaire, libertarienne d'entrepreneurs méritocrates hippies, et de digerati (digital literati)			

conceptuelles du crowdsourcing

4.1.4. Critiques du crowdsourcing. Vers une ubérisation des bibliothèques ?

Le *crowdsourcing* appliqué aux bibliothèques pourrait aussi être considéré comme une forme d'ubérisation des bibliothèques. L'ubérisation pourrait être définie comme la remise en question de sociétés et de professions établies par l'émergence de plateformes web permettant à des non professionnels d'offrir des services concurrents. Dans le domaine des bibliothèques, elle pourrait prendre la forme d'un remplacement du professionnel producteur de données faisant autorité par un bénévole travaillant gratuitement ou encore un travailleur sous payé produisant des données de piètre qualité et en dehors de tout cadre juridique [FOR 11]. Cette exploitation du travail invisible des prolétaires du web est parfois qualifiée de « servuction ». Elle est accusée de concurrence déloyale par les prestataires traditionnels. Elle provoquerait un désengagement des travailleurs, comme de ceux qui emploient cette main d'œuvre interchangeable. Elle générerait des rapports impersonnels et des escroqueries.

En conséquence, certains penseurs, comme Bernard Stiegler, parlent de créer un « revenu contributif » [STI 15], d'autres de fiscaliser les données afin de rendre aux citoyens une part de la valeur qu'ils ont produite par leur travail invisible de production de données, d'autres encore de faire des données produites par les foules, des biens communs.

Pour leur part, les bibliothèques restent parfois trop centrées sur la constitution de collections en tant que fin en soi plutôt que sur la satisfaction des besoins des lecteurs. Avant la numérisation de masse, elles jouissaient d'une forme de monopole dans l'accès à l'information, et leurs cadres, formant une corporation dotée d'une relative homogénéité idéologique, bénéficiaient de titres prestigieux de conservateurs. Dans ces conditions, une externalisation du travail de l'expert auprès de l'amateur, une ouverture sur le privé sous une forme renouvelée de partenariat public / privé, risque bien d'être considérée comme une remise en cause, une perte de contrôle, une concurrence déloyale, une attaque contre les acquis sociaux, et finalement, une ubérisation des bibliothèques.

La question de la qualité des contributions, des coûts liés au contrôle de la qualité, de l'appropriation individuelle du patrimoine collectif par le profane, l'inculte et de la malveillance possible des internautes sera ainsi mise en avant contre les projets de crowdsourcing qui ne pourront se développer sans une importante conduite du changement.

4.2. Taxonomie et panorama du crowdsourcing en bibliothèques

La plupart des auteurs dressent une typologie des projets de crowdsourcing en fonction de degré d'engagement du public. Ainsi, avec le crowdsourcing participatif ou contributif, les internautes se contentent de produire des données pour des institutions qui conçoivent les projets, en pilotent le développement et en cadrent la participation du public qui reste limitée à des microtâches ne demandant qu'un investissement individuel faible. Avec le crowdsourcing collaboratif, les internautes peuvent également interagir entre eux. Avec la co-création, cet investissement individuel est encore plus fort puisque les internautes pourront peser activement sur la politique et la définition du périmètre et des objectifs des projets, et même parfois être à l'origine des projets eux-mêmes.

Au-delà de cette distinction quantitative, nous avons été amenés à proposer une taxonomie plus qualitative des projets de crowdsourcing en bibliothèques. Nous distinguons ainsi les grands types suivants :

Le crowdsourcing explicite	Définition	Projets recensés
Crowdsourcing bénévole	Recours au travail volontaire des internautes bénévoles	<p>Mise en ligne et curation participatives : Oxford's great war archive, Europeana 1914-1918, Internet Archive, Commons Wikimedia, Wir waren so frei, Open Call - Brooklyn Museum, Pin-a-tale, Make history, Click! A Crowd-Curated Exhibition, The Changing Faces of Brooklyn, ExtravaSCANza.</p> <p>Correction participative de l'OCR : TROVE, Distributed Proofreader, Wikisource, California Digital Newspaper Collection, Correct, Franscriptor.</p> <p>Transcription participative de manuscrits : Transcribe Bentham, What's on the Menu ?, Ancient lives, ArchHIVE, What's the score, Transkribus, les Herbonautes, Do it yourself History, Monasterium Collaborative Archive, Citizen Archivist Dashboard, National Archives Transcription Pilot Project, Field Notes of Laurence M. Klauber, Notes from Nature, Transcribe Bushman, Smithsonian Digital Volunteers Transcription Center.</p> <p>Folksonomie : Flickr : The Commons, steve.museum, GLAM Wikimedia, Glashelder!, VeleHanden, 1001 Stories Denmark, Historical Maps Pilot, Mtagger, PennTags, Social OAC, Describe me, Tag! You're It!, Freeze tag!, Your Paintings Tagger, Operation war diary...</p>
Crowdsourcing rémunéré	Recours au travail d'internautes rémunérés	<p>Tous types de travaux : Amazon Mechanical Turk Marketplace, 99design, CloudCrowd, Cloud-Flower, CrowdFlower, eLance, Foule Factory, Freelancer, Guru, Innocentive, ManPower, Mob4hire, MobileWorks, oDesk, Postmates, quora.com, Samacource, sparked.com, TaskRabbit, Topcoder, Trada, Turkit, uTest...</p>

Le crowdsourcing implicite et la gamification	Définition	Projets recensés
Crowdsourcing implicite	recours au travail involontaire des internautes	Correction de l'OCR : reCAPTCHA
Gamification "human computation" "games with a purpose"	recours au travail des internautes sous la forme de jeux	Correction de l'OCR : Digitalkoot, COoperative eNginne for Correction of ExtRacted Text, TypeAttack, Word Soup Game, Smorball, Beabstalk Indexation : Art Collector, Google Image Labeler, ESP Game, GWAP, Peekaboom, KissKissBan, PexAce, museumgam.es, Metadata Games, SaveMyHeritage, Picaguess, Wasida

	Définition	Projets recensés
Le crowdfunding	recours aux contributions financières des internautes	Numérisation à la demande : ebook en demand (EOD), le livre à la carte, Phénix éditions, Chapitre.com, les amis de la BnF, Numalire, revealdigital, Lyrasis, FeniXX, unglue.it, Maine Shared Collections Strategy, International Amateur Scanning League, "Sauvez nos reliures" Impression à la demande : Espresso Book Machine, Electronic Library, Higher Education Resources ON Demand, Amazon Book Surge, CreateSpace, Jouve, lulu.com, Lightning source, Virtual Bookworm, Wingspan press, iUniverse, Xlibris.

Tableau 4.3. *Taxonomie du crowdsourcing et panorama des projets pour les bibliothèques numériques*

Chacune de ces formes de crowdsourcing exposées de manière synthétique et introductive dans les tableaux qui précèdent font l'objet d'une analyse développée dans la partie qui suit et qui reprend la structuration de cette taxonomie.

4.2.1. *Le crowdsourcing explicite*

4.2.1.1. *Le crowdsourcing bénévole*

C'est la forme la plus évidente et classique de crowdsourcing. Mais, le recours aux bénévoles pourrait vite atteindre ses limites face à la multiplication des projets. Par ailleurs, rien n'indique que les futures générations de retraités, qui forment une part parfois importante des contributeurs, auront les mêmes centres d'intérêt dans l'avenir.

Modification de Page:Faulon - De la névrotomie plantaire.djvu/21

OCR text (left sidebar):

```

{{tiret|simul|tanément,}} le cheval pointe tantôt d'un
pied, tantôt de
l'autre, et le temps d'appui sur chacun d'eux est
toujours
très court. Les membres postérieurs s'engagent
sous le centre de gravité en même temps que les reins
se voussent, le décubitus devient fréquent et est souvent
le point de départ de terminaisons funestes. Quant
l'animal
sort de l'écurie, ses membres antérieurs sont
très raides, ses épaules ''chevillées'' et ses pieds
rasent le
sol ; il fléchit sur ses boulets et quelquefois même il
butte et tombe. Cependant, à mesure qu'il s'échauffe,
on voit les membres récupérer peu à peu leur souplesse
qu'ils semblaient avoir perdue, les épaules paraissent
plus libres et les allures finissent par être plus
relevées. Mais dès qu'il est refroidi, tous les symptômes
précédents reparaissent avec la plus grande intensité.

En résumé la symptomatologie de la maladie naviculaire
est caractérisée : 1° par l'absence d'altérations
physiques appréciables extérieurement auxquelles on
puisse rattacher comme à leur cause suffisante l'attitude
du pointer et la claudication. — 2° Plus tard, le sabot
accuse le plus souvent soit par des cercles, soit par son
resserrement, soit par des bleimes, la maladie profonde
qui l'affecte ; de plus les tendons et les muscles du
membre s'altèrent fortement. — 3° Enfin, à une époque
encore plus avancée, la difficulté des mouvements des
membres antérieurs imprime aux allures, un cachet
particulier tout à fait pathognomonique, qui caractérise
si bien la maladie naviculaire, qu'il est impossible de

```

OCR text (main content area):

tanément, le cheval pointe tantôt d'un pied, tantôt de l'autre, et le temps d'appui sur chacun d'eux est toujours très court. Les membres postérieurs s'engagent sous le centre de gravité en même temps que les reins se voussent, le décubitus devient fréquent et est souvent le point de départ de terminaisons funestes. Quant l'animal sort de l'écurie, ses membres antérieurs sont très raides, ses épaules chevillées et ses pieds rasant le sol ; il fléchit sur ses boulets et quelquefois même il butte et tombe. Cependant, à mesure qu'il s'échauffe, on voit les membres récupérer peu à peu leur souplesse qu'ils semblaient avoir perdue, les épaules paraissent plus libres et les allures finissent par être plus relevées. Mais dès qu'il est refroidi, tous les symptômes précédents reparaissent avec la plus grande intensité.

En résumé la symptomatologie de la maladie naviculaire est caractérisée : 1° par l'absence d'altérations physiques appréciables extérieurement auxquelles on puisse rattacher comme à leur cause suffisante l'attitude du pointer et la claudication. — 2° Plus tard, le sabot accuse le plus souvent soit par des cercles, soit par son resserrement, soit par des bleimes, la maladie profonde qui l'affecte ; de plus les tendons et les muscles du membre s'altèrent fortement. — 3° Enfin, à une époque encore plus avancée, la difficulté des mouvements des membres antérieurs imprime aux allures, un cachet particulier tout à fait pathognomonique, qui caractérise si bien la maladie naviculaire, qu'il est impossible de

Résumé :

Modification mineure Suivre cette page [G](#) [O](#) [C](#) [R](#) État de la page (Qualité des pages)

→ Vérifiez la typographie ; Respectez l'orthographe d'origine.
→ Débutants, lisez la FAQ.

En cliquant sur « Publier », vous acceptez de placer votre contribution sous licence [Creative Commons BY-SA 3.0](#) et [GFDL](#). Vous acceptez d'être crédité par les ré-utilisateurs au minimum via un hyperlien ou une URL vers l'article sur lequel vous contribuez. Voyez les conditions d'utilisation pour plus de détails.

Publier Prévisualiser Modifications en cours Annuler Aide (ouvre une nouvelle fenêtre)

Figure 4.1. Page d'une thèse ancienne conservée à l'Ecole Nationale Vétérinaire de Toulouse dont la correction de l'OCR est proposée via Wikisource

4.2.1.2. Le crowdsourcing rémunéré

Les premiers utilisateurs de la plus grande plateforme de crowdsourcing rémunéré, l'Amazon Mechanical Turk Marketplace, sont les laboratoires de recherche américains. Cette plateforme met en relation ceux qui viennent proposer et ceux qui viennent rechercher du travail en ligne, généralement sous la forme de microtâches. Elle permet de mobiliser en quelques minutes, sans démarches administratives, à des coûts déterminés librement par l'offre et la demande, des foules de travailleurs dignes des plus grandes multinationales, avec des profils diversifiés, parmi plus de 500 000 internautes disponibles en permanence dans près de 200 pays, principalement aux USA et en Inde [IPE 10]. Elle permet ainsi de réaliser en une demi-journée des travaux qui auraient nécessité des années de travail ingrat auparavant et généré des « burn out ». Pour leur part, les travailleurs de la plateforme sont libres de travailler où ils le souhaitent, quand ils le souhaitent, autant qu'ils le souhaitent, pour qui ils le souhaitent, en fonction de leurs propres centres d'intérêts ; d'être tour à tour employeur et employé et de travailler pour un client plutôt que pour un patron.

Le nom de « Amazon Mechanical Turk Marketplace » est astucieusement inspiré d'un automate joueur d'échecs conçu au XVIIIe siècle et qui était prétendument doué d'une intelligence artificielle alors que, en réalité, un humain était caché derrière. De la même manière, derrière des résultats que l'on croit être réalisés par des algorithmes puissants, il peut, en réalité, y avoir des foules d'humains cachés derrière, en particulier via le crowdsourcing rémunéré.

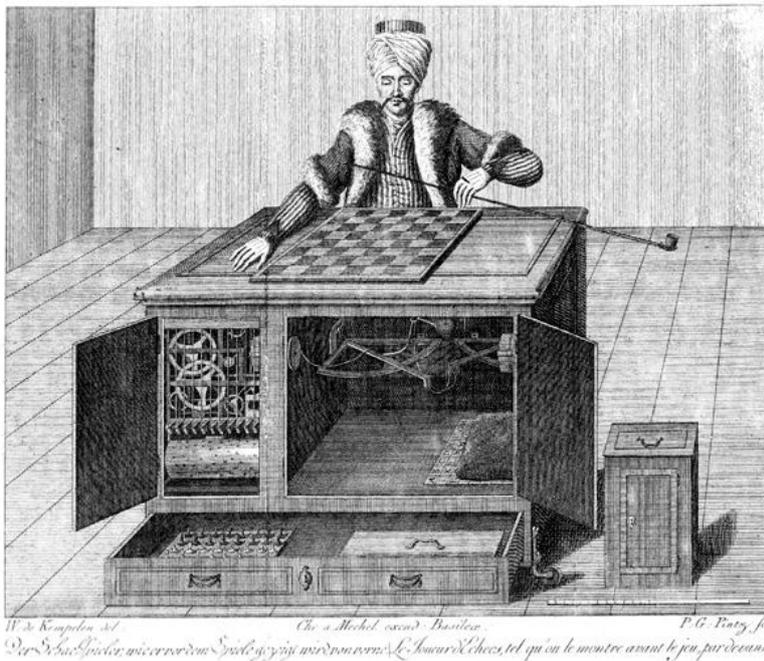


Figure 4.2. « Tuerkischer schachspieler windisch4 » par Karl Gottlieb von Windisch. 1783. Public domain via Wikimedia Commons

4.2.2. La gamification et le crowdsourcing implicite

La volonté de contribuer n'est pas forcément le but premier pour les participants de ces formes de *crowdsourcing*. Elles font appel à l'envie de jouer des internautes pour obtenir du travail de leur part (*gamification*) ou les font travailler sans qu'ils le sachent (*crowdsourcing* implicite).

4.2.2.1. La gamification

La *gamification*, qui pourrait être traduite en français par le terme de ludification, consiste à faire travailler des internautes sous la forme de jeux ayant ainsi une finalité utile et productive (« *Games with a purpose* »). Elle pourrait être définie comme le fait d'appliquer des éléments de design, de psychologie et de mécanismes du jeu vidéo dans d'autres contextes [DET 11].

Le fait de simplement attribuer des points pour la participation des internautes ne doit donc pas être confondu avec de la *gamification* mais résulte plutôt d'une forme de « pointification ». La *gamification* se distingue également des « *serious games* » car elle ne vise pas à se former pour un développement personnel mais à réaliser des objectifs extérieurs à soi comme corriger de l'OCR ou indexer des photographies numérisées. [AND 15, b]

Contrairement au *crowdsourcing* explicite, la réalisation de micro-tâches réalisées de manière aléatoire et hors contexte dans un jeu est généralement moins propice au développement personnel et à l'acquisition de connaissances mais elle pourrait permettre de faire faire plus facilement des tâches parfois assez ingrates.



Figure 4.3. Capture d'écran du jeu de correction de l'OCR *Digitalkoot* [CHR 11].

4.2.2.2. Le crowdsourcing implicite

La notion de *crowdsourcing* implicite a été conceptualisée par [HAR 13] mais le terme est encore très peu répandu dans la littérature. Les internautes qui participent le font involontairement ou inconsciemment. Le *crowdsourcing* implicite pourrait donc être considéré comme moins éthique que le *crowdsourcing* explicite et volontaire des bénévoles du web, une intrusion du web marchand cherchant à instrumentaliser les internautes.

Le projet le plus emblématique de cette forme de *crowdsourcing*, dans le domaine des bibliothèques numériques, est reCAPTCHA. Afin de créer un compte sur un site web et éviter toute attaque et submersion par de robots, les sites web demandent aux internautes de ressaisir des mots déformés et prouver ainsi qu'ils ne sont pas des robots malveillants. Les programmes Google Books et Google Maps ont ainsi astucieusement utilisé ce système afin de faire corriger l'OCR brute issue de leurs campagnes de numérisation par des foules d'internautes en confrontant leurs saisies. 200 millions de mots seraient ainsi saisis chaque jour, 12 000 heures de travail involontaires seraient ainsi récoltées et, d'après nos calculs, 146 millions d'euros par an ne seraient pas dépensés par Google en prestations de correction de textes. [AND 15a].

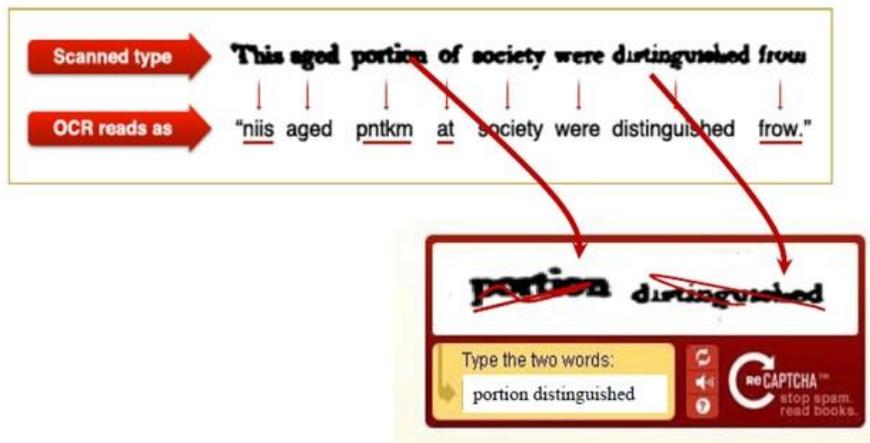


Figure 4.4. Schéma expliquant le fonctionnement de reCAPTCHA d'après <https://www.google.com/recaptcha>

4.2.3. Le crowdfunding

Parfois qualifié de mendicité institutionnelle [AYR 13], le *crowdfunding* est bien une forme de *crowdsourcing* qui fait appel, non pas au travail des internautes, mais à leurs ressources financières. En bibliothèque, il peut être utilisé pour acquérir des documents ou pour en financer la numérisation.

4.2.3.1. La numérisation à la demande

La numérisation à la demande permet aux bibliothèques d'offrir des services de reproduction numérique en faisant supporter le coût aux internautes, d'externaliser en partie, auprès des internautes l'ingrate et coûteuse tâche de sélection des documents qui méritent encore d'être numérisés, et évidemment de compléter leurs programmes de numérisation. L'utilisateur se retrouve ainsi placé au centre de la politique des bibliothèques [GST 11] qui ont pu, par le passé, avoir parfois tendance à le négliger en se concentrant principalement sur les collections. La politique documentaire de la bibliothèque numérique devient ainsi une co-construction entre les bibliothécaires et le grand public puisque la politique d'acquisition est désormais partagée. Pour les internautes, la numérisation à la demande permet d'avoir accès à un service de reproduction numérique. Pour d'éventuels mécènes et investisseurs, la numérisation à la demande pourrait être l'occasion de financer la numérisation de livres susceptibles d'intéresser tel ou tel public, et éventuellement de rechercher un retour sur investissement à travers le trafic web généré par ces livres sur le modèle de publicités Google Adwords.

Ce modèle économique pourrait permettre à l'argent public devenu plus rare de se concentrer sur la numérisation de documents à intérêt patrimonial, historique et scientifique n'intéressant pas le privé et de laisser l'argent privé de particuliers ou de mécènes financer la numérisation d'ouvrages intéressants le grand public ou des communautés d'érudits. Ce faisant, les bibliothèques auraient la possibilité de mieux se recentrer sur les domaines d'expertise qui sont les leurs et de mieux valoriser les compétences des conservateurs dans le domaine patrimonial, historique et scientifique.

La numérisation à la demande par *crowdfunding* est une nouvelle forme de souscription publique permettant de donner une nouvelle vie à une œuvre. Elle est particulièrement adaptée à la situation actuelle où il ne reste que des miettes à numériser après le passage des grands programmes de numérisation de masse.

La principale difficulté des projets de numérisation à la demande consiste à évaluer automatiquement les coûts de numérisation des documents. On constate, en effet, que, pour ces projets, il est nécessaire d'effectuer un devis à chaque demande. Ce devis sert à évaluer le coût de numérisation. La production du devis nécessite de vérifier la présence du document, son état, son nombre réel de feuillets, son format et son angle d'ouverture qui détermineront le nombre de vues à produire et type de scanner à utiliser, donc le coût de numérisation. Après réception du devis par l'internaute, celui-ci ne concrétise malheureusement que très rarement sa demande par une commande car son envie d'achat est passée, et car il peut être surpris par le coût à payer. Au final le temps de travail consacré à produire des devis coûte autant que l'argent collecté auprès des internautes et, en l'absence d'un calcul automatique des coûts de numérisation ou d'une subvention d'argent publique, les projets de numérisation à la demande ne guère viables. [AND 14b].

4.2.3.2. L'impression à la demande et les bibliothèques

Bien qu'il ne s'agisse pas de *crowdsourcing*, le modèle économique est identique à celui de la numérisation à la demande dont il est souvent indissociable. Il s'agit, cette fois, de ressusciter un imprimé à partir du document numérisé. Ce modèle de plus en plus utilisé dans l'édition pour produire en flux tendus, sans stocks ni invendus a été appliqué aux programmes de numérisation des bibliothèques [AND 15c].

Comme nous l'avons vu, avec la numérisation à la demande, la politique documentaire des bibliothèques numériques et la constitution de collections numériques sont désormais d'avantage une co-construction, elles sont désormais aussi en partie l'œuvre des internautes. Avec l'impression à la demande, on pourrait également aller jusqu'à imaginer une bibliothèque physique directement constituée d'imprimés demandés par ses lecteurs, imprimés en quelques minutes via une Espresso Book Machine et constituant, après restitution par le lecteur, une collection construite par l'utilisateur et constituée d'ouvrages ayant tous été consultés au moins une fois. [LEW 10]. Ce mode de fonctionnement serait radicalement différent de la politique d'acquisition des bibliothèques qui repose actuellement essentiellement sur l'anticipation des besoins et l'achat de livres au cas où ils intéressent un jour un lecteur, une politique qui n'est donc pas exclusivement centrée sur l'utilisateur. Elle prolonge ainsi la possibilité déjà offerte par les bibliothèques à leurs lecteurs de suggérer des acquisitions.



Figure 4.5. L'Espresso Book Machine d'après <http://ondemandbooks.com>

Au-delà de cette taxonomie, il y a des formes de *crowdsourcing* qui restent probablement encore à inventer comme la *gamification* rémunérée en fonction des résultats obtenus au cours des jeux, le recours à des citoyens éventuellement rémunérés pour numériser eux-mêmes les documents au sein des bibliothèques ou encore, la mise en œuvre d'un reCAPTCHA au bénéfice des bibliothèques publiques ou d'un reCAPTCHA facturant pour l'OCR corrigé et partageant ses bénéfices avec les sites ayant accepté de l'implémenter.

4.3. Analyses du crowdsourcing en bibliothèque du point de vue des sciences de l'information et de la communication

4.3.1. Pourquoi les bibliothèques ont recours au crowdsourcing et quelles sont les conditions nécessaires ?

Clay Shirky considérait que si les américains consacraient leur temps à des projets comme Wikipédia au lieu de regarder la télévision, ils pourraient créer 2000 projets de la même ampleur que la célèbre encyclopédie participative [SHI 10]. Luis Von Ahn, quant à lui, prétendait que les 425 millions d'images de Google Images auraient pu être indexées en seulement 31 jours par 5000 internautes jouant à ESP Game [VON 08]. Quoi qu'il en soit, il existerait un important réservoir de bonnes volontés dont les bibliothèques pourraient bénéficier d'autant qu'elles disposent déjà d'une expérience dans l'animation de communautés, l'encadrement de bénévoles, qu'elles ont une bonne image auprès du public pour qui elles apparaissent comme digne de confiance et au service de l'intérêt général et qu'elles pourraient, par conséquent, plus facilement recruter.

Elles pourraient donc avoir recours au *crowdsourcing*, c'est à dire à l'externalisation de micro-tâches auprès de foules d'internautes pour réduire leurs coûts ou pour décupler leurs ressources humaines et réaliser un travail « de fourmi », de « petites mains » qu'elles n'ont plus les moyens d'assumer, ou encore pour achever, réaliser ou rendre possibles des projets jusqu'à présent inachevables, irréalisables, ou même inimaginables auparavant [HOL 10]. En bénéficiant de la collaboration des internautes, les bibliothèques pourraient bénéficier de connaissances et de compétences illimitées, bien au-delà de celle de leurs équipes restreintes et limitées et ce, malgré l'excellente culture générale de ses cadres. Le *crowdsourcing* interroge ainsi les frontières de l'organisation car il permet de créer de la valeur au delà de ses frontières [REN 14]. Les bibliothèques pourraient ainsi capter la puissance qui est plus forte à l'extérieur qu'à l'intérieur de leurs organisations et mobiliser tel ou tel érudit ou spécialiste pour identifier un lieu ou une personne sur une photographie, reconnaître une reliure, un atelier d'incunables, dater un objet d'un cabinet de curiosités... Les bibliothèques pourraient ainsi s'engager dans un processus de redocumentarisation participative et voir leurs collections revisitées et réinventées. Au-delà de ces motivations, les bibliothèques pourraient aussi chercher à engager en profondeur l'utilisateur autour de leurs collections, à démocratiser la conservation du patrimoine sous la forme d'un « empowerment », d'une émancipation et d'un changement des relations entre le patrimoine et la société au nom d'un droit à l'information et à la participation, pour améliorer leur image, pour chercher à innover ou malheureusement aussi pour faire de la communication institutionnelle autour d'un sujet à la mode.

Bien qu'il soit tout à fait possible de communiquer sur une authentique utilisation des données produites par des usagers, il semble exister deux conceptions différentes dans les bibliothèques. L'une considère plutôt que les bibliothèques ont besoin des internautes et l'autre que ce sont les internautes qui ont besoin des bibliothèques. La première est utilitariste et économique. Elle reconnaît humblement son besoin de renfort et cherche véritablement à produire un résultat grâce à la collaboration entre patrimoine et société. La seconde est plus liée à des considérations démocratiques et recherche d'avantage à faire du *crowdsourcing* pour faire du *crowdsourcing*, de la publicité autour de la démarche et de la participation du public comme fin en soi. Lorsque cette seconde conception prédomine, le travail produit par les internautes n'est finalement guère valorisé et utilisé et les métadonnées qu'ils ont saisies ne seront que rarement intégrées dans les systèmes d'informations

des bibliothèques, ce qui peut décourager et être vécu comme une trahison par les bénévoles comme nous le verrons dans un chapitre suivant.

Quoi qu'il en soit, les bibliothèques ne pourront toutefois s'engager dans la voie du *crowdsourcing* qu'à la condition que les tâches concernées puissent être réalisées par internet sous la forme de micro-tâches, qu'elles ne mobilisent pas de données confidentielles, qu'elles puissent être prises en charge de manière indépendante et en ne nécessitant que peu d'interaction, de formation et de communication, et enfin, qu'elles puissent être accomplies par des amateurs non experts.

Il existe un intérêt croissant pour le *crowdsourcing* en bibliothèques dans la littérature mondiale comme l'illustre le diagramme suivant :

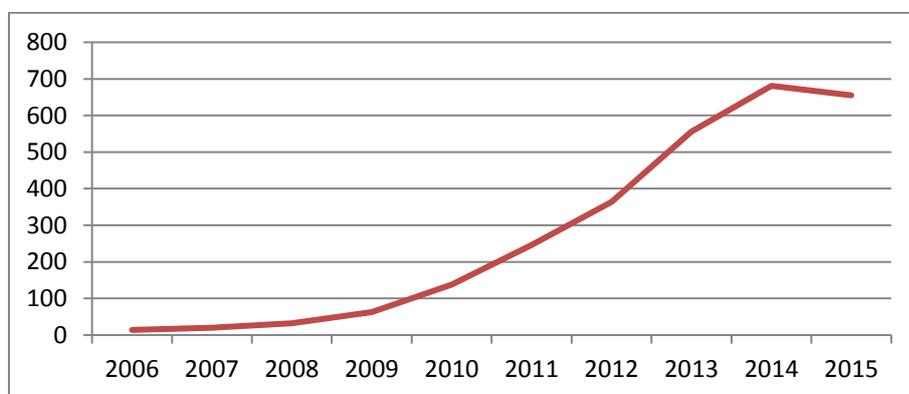


Figure 4.6. Nombre de publications indexées dans Google Scholar, en fonction des années de publication et répondant à l'équation de recherche *crowdsourcing AND library AND digitization*

Malgré cet intérêt croissant dans la littérature, on constate cependant un relativement faible intérêt pour le sujet en France. Ainsi, la bibliographie de 216 publications réalisée dans le cadre d'une thèse sur le sujet par les auteurs de ces lignes ne contient que 7 articles ayant au moins un auteur français comme l'illustre le diagramme suivant :

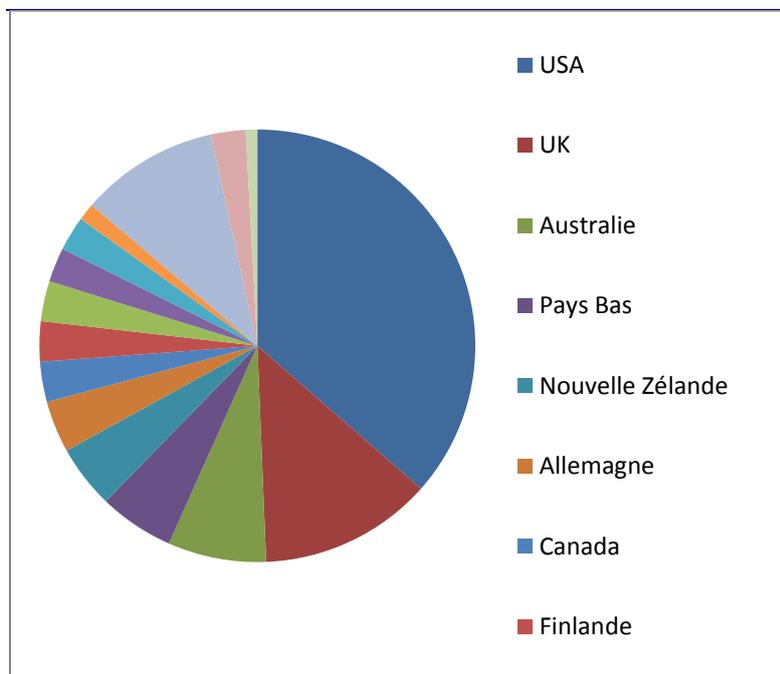


Figure 4.7. Poids relatif des pays dans la bibliographie de la thèse (216 publications)

Par ailleurs un état de l'art de l'OCLC faisant autorité sur le sujet et répertoriant les projets de *crowdsourcing* en bibliothèques [SMI 11] ne mentionne jamais la France sur 350 pages et il aura fallu attendre février 2013 pour qu'une première étude soit publiée en France sur le sujet [MOI 13], dans le cadre d'un projet de la Bibliothèque Nationale de France.

4.3.2. Pourquoi les internautes contribuent ? Taxonomie des motivations des internautes

Parmi les différents ressorts qui amènent les internautes à contribuer à des projets de *crowdsourcing* en bibliothèques, on distingue principalement les motivations intrinsèques qui poussent l'individu à agir de manière désintéressée et pour le seul plaisir que le travail lui procure et les motivations extrinsèques qui le poussent à travailler pour les effets et résultats obtenus grâce à ce travail comme la reconnaissance, les récompense, ou la rémunération. Dans la littérature, les motivations suivantes ont été identifiées et organisées par nos soins :

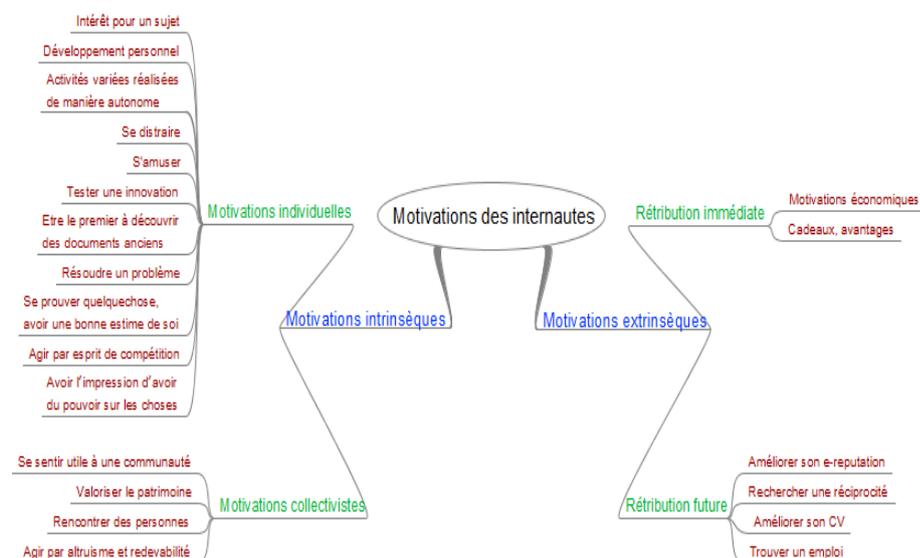


Figure 4.8. Taxonomie des motivations des internautes qui participent à des projets de crowdsourcing en bibliothèques

En fonction des types de *crowdsourcing* et des types d'individus et de cultures, les types de motivations peuvent être très divers. Ainsi, si le *crowdsourcing* bénévole mobilise des internautes à la recherche d'un développement personnel, la *gamification* fait plutôt appel au besoin de se distraire utilement. Le ressort des motivations extrinsèques et, en particulier, des récompenses symboliques, virtuelles ou réelles peut parfois se faire au détriment de la qualité des données produites et des motivations intrinsèques, c'est à dire celles qui le poussent à agir pour le seul intérêt du travail lui-même. Dès 1975, une expérience de Edward L. Deci avait ainsi montré que si des personnes étaient rémunérées pour réaliser des puzzles, elles perdaient tout intérêt pour cette activité si elles ne percevaient plus aucune rémunération pour s'y adonner.

4.3.3. Des récompenses symboliques aux rémunérations concrètes

Comme le montre la carte conceptuelle des motivations qui précède, les projets de *crowdsourcing* se font toujours nécessairement au bénéfice mutuel de l'institution et de l'internaute. Outre les motivations intrinsèques, les récompenses peuvent aller de récompenses symboliques (classements, grades, médailles) à des récompenses bien réelles, des cadeaux ou même à des rémunérations. Ainsi, des bénévoles du projet Foldit ont été publiquement remerciés dans un article de l'illustre revue académique *Nature Structural & Molecular Biology* vol. 18, 2011 pour avoir permis de découvrir la structure d'une enzyme très importante. D'autres bénévoles d'autres projets ont été mentionnés dans des newsletters, invités à témoigner dans le cadre de conférences ont bénéficié de formations diplômantes, d'abonnements, de livres, de T-shirts, de lecteurs MP3, de bons d'achats, de visites ou de voyages.

4.3.4. La communication pour recruter des contributeurs

Les institutions culturelles bénéficient d'une bonne image auprès de public et apparaissent comme dignes de confiance, au service de l'intérêt général. Elles disposent donc de solides atouts pour recruter des bénévoles. Parmi les moyens de communication utilisés pour les projets de *crowdsourcing*, on peut relever les campagnes dans la presse spécialisée, associative, locale, nationale, la publication d'articles et de posters, la distribution de dépliants, le collage d'autocollants et d'affiches diffusés notamment à l'occasion de conférences et de salons, l'organisation de réunions publiques ou d'événements spécifiques, la présence à la radio et à la télévision, mais aussi, la production de vidéos, l'utilisation des réseaux sociaux, des forums, des listes de diffusion, du mailing, des sites institutionnels, et enfin, l'achat de mots spécifiques dans le cadre de campagnes Google Adwords.

Un site de *crowdsourcing* doit toujours avoir une page d'accueil qui décrit le projet, en explicite clairement et simplement la finalité, son état d'avancement et invite immédiatement à la participation des bénévoles en leur montrant en quoi leur participation sera utile et comment ils seront accompagnés et reconnus [MCK 15].

4.3.5. Le community management pour maintenir les contributeurs

La majorité des données produites dans le cadre de projets de *crowdsourcing* a été produite par une minorité bien déterminée de participants et non par des foules d'anonymes.

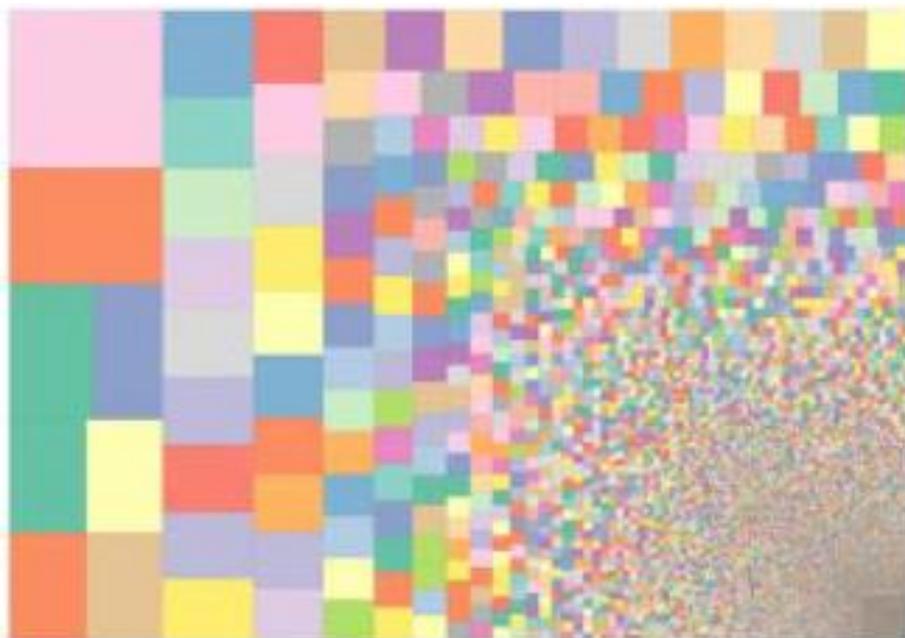


Figure 4.9. Quelques internautes produisent la majeure partie des contributions (selon le blog de Brumfield manuscripttranscription.blogspot.fr en 2013)

A l'instar de ce qui apparaît dans le précédent diagramme où chaque carré correspond aux contributions d'une seule personne et où la taille de chaque carré est proportionnelle à la quantité de contributions, tous les projets de *crowdsourcing* bénévole rapportent aussi que la majeure partie des contributions est le fait d'une minorité et qu'il ne s'agit donc pas véritablement de foules anonymes, mais plutôt d'une communauté bien définie de bénévoles [OWE 13]. Dans ces conditions, il serait donc plus judicieux de parler de

communitysourcing [CAU 12] ou encore de *nichesourcing* et de rechercher à recruter des personnes bien ciblées plutôt que de s'adresser à des foules sans visages.

Si le profil type d'un bénévole est un homme diplômé avec un haut niveau d'études, de catégorie socio-professionnelle élevée et qui vient de finir ses études ou qui est à la retraite, les communautés de contributeurs regroupent aussi des profils plus diversifiés et forment des communautés de pairs ayant des affinités et des objectifs communs dont il faudra maintenir le dynamisme par un *community management*. En effet, les bénévoles devront être recrutés par des actions de communication, encadrés, managés, soutenus par des hotlines, des helpdesks, des forums, formés par des manuels, des tutoriels, motivés, tenus régulièrement informés, leurs contributions modérées, leur qualité contrôlée, les statistiques de participation suivies, et enfin, les données produites réintégréées.

4.3.6. La qualité et la réintégration des données produites

La question de la qualité et du vandalisme est un argument central des opposants à l'externalisation de la production de données auprès d'internautes amateurs. Il existe cependant de nombreux moyens éprouvés afin de garantir l'une et de se prémunir de l'autre, à partir de robots, d'autorégulation, de contrôles par des professionnels ou par des bénévoles, d'évaluations, de votes, de tests d'aptitude ou encore, de confrontation des saisies.

Quoi qu'il en soit, nombreuses sont les études qui montrent que les données produites par ceux qui consultent un document numérisé sur le web peuvent être de meilleure qualité car celui qui consulte le document est généralement quelqu'un qui connaît bien le sujet sur lequel porte le document qu'il consulte. Par ailleurs, comme le prétendent les tenants de la « sagesse des foules », la diversité de profils constitués par une foule d'internautes et la « loi des grands nombres » auraient pour effet de neutraliser les erreurs individuelles dans la masse des données justes [BOE 12] et même parfois de donner des meilleurs résultats que ceux obtenus avec des experts. Cela semble être confirmé par de multiples études comparatives [ROR 10] [OOM 11].

Néanmoins, les données produites par les internautes ne sont pas toujours réintégréées par les institutions qui visent parfois d'avantage une démocratisation du patrimoine ou encore une communication institutionnelle autour d'un sujet à la mode qu'une réelle participation des internautes. Lorsqu'elles sont réutilisées, ces données issues de contributions bénévoles et gratuites, devraient toujours être diffusées sous des licences juridiques autorisant la plus large réutilisation possible.

4.3.7. L'évaluation des projets de crowdsourcing

Bien qu'il soit difficile de recueillir des données chiffrées dans la littérature hormis pour [CAU 12], il semble que les projets de *crowdsourcing* soient loin d'être tous rentables au sens où les bénéfices en données récoltées ne soient pas toujours au niveau des coûts en gestion de projets, en développement de plateformes, en administration, en maintenance, en hébergement, en communication, en *community management* et réintégration des données produites.

Dans une étude précédente [AND 15a], nous avons estimé que le projet California Digital Newspaper Collection avait récolté du travail de correction d'OCR pour plus de 1500 € par mois, le projet Digitalkoot pour plus de 2000 € par mois, le projet TROVE pour près de 35 000 € par mois, et le projet reCAPTCHA pour plus de 12 millions d'euros par mois.

Les projets qui semblent le mieux fonctionner sont ceux qui reposent sur des finalités claires, mènent une campagne de communication efficace, et disposent de communautés managées et motivées. Ceux qui échouent font généralement appel à des tâches trop complexes, à des connaissances trop spécialisées qui nécessitent un

trop grand investissement en formation, qui ne communiquent pas suffisamment avec les bénévoles [RID 13] et parfois, qui ont négligé la conduite du changement en interne.

Conclusions sur l'intelligence collective et la sagesse des foules

Si on fait la moyenne des estimations individuelles d'une foule concernant le poids d'un bœuf, le nombre de billes dans un récipient, la température dans une pièce, la réponse à une question de culture générale dans le cadre d'un jeu télévisuel comme « Qui veut gagner des millions ? », on constate que cette moyenne est très proche de la vérité. Ce phénomène identifié depuis très longtemps sous l'adage « vox populi vos dei » et bien compris par Machiavel qui écrivait que :

« ce n'est pas sans raison qu'on dit que la voix du peuple est la voix de Dieu. On voit l'opinion publique pronostiquer les événements d'une manière si merveilleuse, qu'on dirait que le peuple est doué de la faculté occulte de prévoir et les biens et les maux. Quant à la manière de juger, on le voit bien rarement se tromper. » [MAC 37]

Ce phénomène est aujourd'hui qualifié de « sagesse des foules » [SUR 04]. C'est sur ce phénomène que s'appuient nombre de projets de *crowdsourcing* afin d'obtenir des données de qualité ou même parfois une véritable expertise. Ainsi, une agence de renseignements américaine, avec le « Good Judgement Project », s'est appuyée sur les prévisions géopolitiques de foules d'internautes estimant quantitativement la probabilité ou l'improbabilité de tel ou tel événement. De la même manière, le *big data* et l'analyse par des technologies de *text mining* des lieux géographiques en co-occurrence avec le nom « Ben Laden » dans la presse mondiale a montré que ces lieux étaient proches de celui où il se cachait. Lorsque les données font foule, elles pourraient ainsi faire science et, grâce au *crowdsourcing*, les cerveaux humains pourraient être reliés comme des processeurs surpuissants contribuant à un calcul beaucoup plus massif que celui des algorithmes. On pourrait ainsi parler de « *Human Computation* » [VON 06].

Néanmoins, dans l'histoire, les foules ne sont pas toujours illustrées par leur sagesse, mais plutôt parfois par l'irresponsabilité criminelle des individus qui composent les masses, comme le relevait déjà Gustave Le Bon, dans sa *Psychologie des foules*, notamment à la lumière des événements liés à la terreur sous la révolution française, et ce, bien avant les expériences totalitaires du XX^e siècle. Aujourd'hui, sur le web, si la folie des foules n'a pas la même ampleur, les rumeurs, les théories du complot et les paranoïas collectives semblent parfois aussi infirmer l'existence d'une quelconque sagesse des foules.

Mais, et au delà de la simple production de données, le recours aux foules d'amateurs de profils diversifiés qui ne cherchent pas à reproduire les modèles établis avec lesquels les professionnels ont été formés peut aussi être source d'heureuses coïncidences (sérendipité), de « lecteurs inattendus », de découvertes accidentelles, de ruptures innovantes. Elle favorise, en tous cas, le développement d'un écosystème de l'innovation. Ainsi, selon Eric Von Hippel, théoricien de la « user innovation », 46 % des entreprises américaines dans les secteurs innovants auraient pour origine un simple consommateur bricoleur [VON 05].

- [AND 14a] ANDRO M., SALEH I., « Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet », dans K. Zreik, G. AZEMARD, S. CHAUDIRON, G. DARQUIE (dir), *Livre post-numérique : historique, mutations et perspectives. Actes du 17e colloque international sur le document électronique (CiDE.17)*, Fès, Maroc, 2014.
- [AND 14b] ANDRO M., RIVIERE P., DUPUY-OLIVIER A., GROPALLO F., MAINGREAU D., « Numalire, une expérimentation de numérisation à la demande du patrimoine conservé par les bibliothèques sous la forme de financements participatifs (crowdfunding) », *Bulletin des Bibliothèques de France*, contribution du 2 octobre, [9] p., 2014.
- [AND 15a] ANDRO M., SALEH I., « La correction participative de l'OCR par crowdsourcing au profit des bibliothèques numériques », *Bulletin des Bibliothèques de France*, contribution du 16 juin, [8] p., 2015.
- [AND 15b] ANDRO M., SALEH I., « Bibliothèques numériques et gamification : panorama et état de l'art », *I2D - Information, données & documents*, vol. 52, n° 4, p. 70-79, 2015.
- [AND 15c] ANDRO M., KLOPP S., « L'impression à la demande et les bibliothèques », *Bulletin des Bibliothèques de France*, contribution du 13 février, [7] p., 2015.
- [AYR 13] AYRES, M.-L., « Singing for their supper⁷: Trove, Australian newspapers, and the crowd », *IFLA World Library and Information Congress*, Singapore, 2013.
- [BOE 12] BŒUF G., ALLAIN Y.-M., BOUVIER M., « L'apport des sciences participatives dans la connaissance de la biodiversité », rapport remis à la Ministre de l'Ecologie, 2012.
- [CAU 12] CAUSER T., WALLACE V., « Building A Volunteer Community: Results and Findings from Transcribe Bentham », *Digital Humanities quarterly*, vol. 6, n° 2, 26 p, 2012.
- [CHR 11] CHRONS O., SUNDELL S., « Digitalkoot: Making Old Archives Accessible Using Crowdsourcing », *HCOMP 2011: 3rd Human Computation Workshop*, San Francisco, USA, 2011.
- [DET 11] DETERDING S., DIXON D., KHALED R., NACKE L. E., « Gamification : toward a Definition », *ACM CHI Gamification Workshop*, New York, USA, 2011.
- [EST 12] ESTELLÉS-AROLAS E., GONZALEZ-LADRON-DE-GUEVARA F., « Towards an integrated crowdsourcing definition », *Journal of Information Science*, 14 p, 2012.
- [FOR 11] FORT K., ADDA G., COHEN K. B., « Amazon Mechanical Turk: Gold Mine or Coal Mine? », *Computational Linguistics*, vol. 37, n° 2, p. 413-420, 2011.
- [GST 11] GSTREIN S., MÜHLBERGER G., « Producing eBooks on Demand: A European Library Network », 2011.
- [HAR 13] HARRIS C. G., « Applying human computation methods to information science », PhD dissertation, University of Iowa, 2013.
- [HOL 10] HOLLEY R., « Crowdsourcing: How and Why Should Libraries Do It? », *D-Lib Magazine*, vol. 16, n° 3-4, 2010.
- [IPE 10] IPEIROTIS P. G., « Demographics of Mechanical Turk », 2010.
- [LEW 10] LEWIS D. W., « The User-Driven Purchase Giveaway Library », *Educause Review*, vol. 45, n° 5, p. 10-11, 2010.
- [MAC 37] N. MACHIAVEL, *Œuvres complètes*, Tome premier. Paris : Auguste Desrez, 1837.
- [MCK 15] MCKINLEY D., « Heuristics to support the design and evaluation of websites for crowdsourcing the processing of cultural heritage assets », 2015.
- [MOI 13] MOIREZ P., MOREUX J. P., JOSSE I., « 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, 2013.
- [OOM 11] OOMEN J., AROYO L., « Crowdsourcing in the cultural heritage domain: opportunities and challenges », 5th International Conference on Communities & Technologies, Brisbane, Australia, June-July 2011.
- [OWE 13] OWENS T., « Digital Cultural Heritage and the Crowd », *Curator: The Museum Journal*, vol. 56, p. 121-130, 2013.
- [REN 14] RENAULT S., « Crowdsourcing : La nébuleuse des frontières de l'organisation et du travail », *RIMHE : Revue Interdisciplinaire Management, Homme(s) & Entreprise*, n° 11, p. 23-40, 2014.

-
- [RID 13] RIDGE M., « From tagging to theorizing: deepening engagement with cultural heritage through crowdsourcing », *Curator: The Museum Journal*, vol. 56, n° 4, p. 435-450, 2013.
- [ROR 10] RORISSA A., « 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, n° 11, p. 2230-2242, 2010
- [SHI 08] SHIRKY C., *Here comes everybody: the power of organizing without organizations*, Penguin Books, London, 2008.
- [SMI 11a] SMITH-YOSHIMURA K., SHEIN C., « Social Metadata for Libraries, Archives and Museums Part 1: Site Reviews », OCLC Research, 2011.
- [SMI 11b] SMITH-YOSHIMURA K., GODBY C. J., HOFFLER H. VARNUM K. YAKEL E., « Social Metadata for Libraries, Archives, and Museums: Survey analysis », OCLC Research, 2011.
- [SMI 12a] SMITH-YOSHIMURA K., « Social Metadata for Libraries, Archives, and Museums: Executive Summary », OCLC Research, 2012.
- [SMI 12b] SMITH-YOSHIMURA K. HOLLEY R., « Social Metadata for Libraries, Archives, and Museums: Recommendations and Readings », OCLC Research, 2012.
- [STI 15] STIEGLER B., *La société automatique. 1, l'avenir du travail*, Fayard. Paris, 2015.
- [SUR 04] SUROWIECKI J., *La sagesse des foules*, traduction de the wisdom of crowds, J.-C. Lattès, Paris, 2004.
- [VON 06] VON AHN L., « Games With A Purpose », *IEEE Computer Magazine*, vol. 39, n° 6, p. 96-98, 2006.
- [VON 08] VON AHN L., DABBISH L., « Designing Games With A Purpose », *Communications of the ACM*, vol. 51, n° 8, p. 58-67, 2008.
- [VON 05] VON HIPPEL E., *Democratizing innovation*, MIT Press, Cambridge, 2005.