



## Crowdsourcing et bibliothèques numériques

Mathieu Andro

► **To cite this version:**

Mathieu Andro. Crowdsourcing et bibliothèques numériques. Semaine internationale du libre accès, conférence sur le thème de la science comme bien commun, Oct 2014, Champs-sur-Marne, France. 2014. <hal-01164625>

**HAL Id: hal-01164625**

**<https://hal-enpc.archives-ouvertes.fr/hal-01164625>**

Submitted on 17 Jun 2015

**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.

# Crowdsourcing et bibliothèques numériques

Mathieu Andro - INRA DV-IST - Paris 8  
ENPC - mardi 21 octobre 2014  
(10 h - 10 h 30)

# L'origine du mot

Jeff Howe, Wired Magazine, juin 2006

- Crowd = foule (des internautes)
- (out)sourcing = externalisation

Au lieu d'externaliser vers des pays à main d'oeuvre à bas coût, pourquoi ne pas bénéficier des internautes ?

→ travail, argent, compétences ou connaissances moins limitées

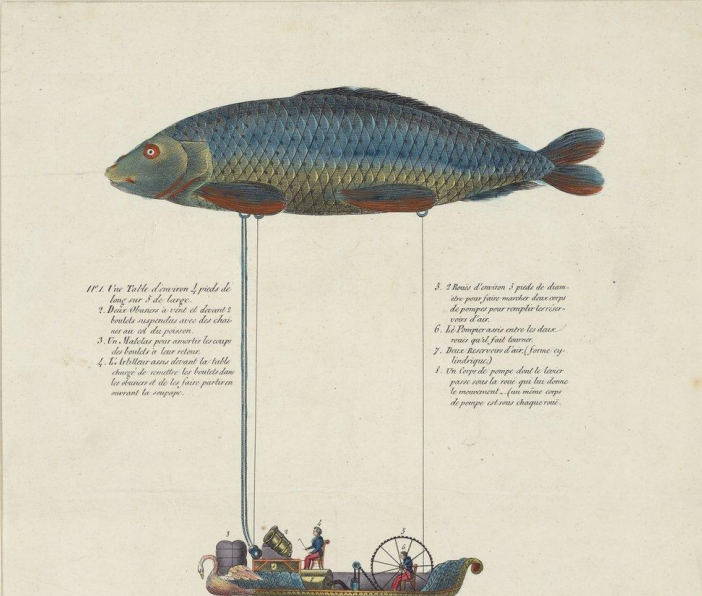
# Histoire du crowdsourcing

- 1714 : Longitude Act : 20 000 livres à qui permettrait de déterminer la longitude d'un navire en pleine mer
- 1884 : statue de la Liberté financée par souscription publique (125 000 personnes).
- 1936 : Toyota rassemble 27 000 personnes et choisi le meilleur dessin proposé comme logo de sa marque
- 1979 : guide de restaurants Zagat

# Parfois originaux...

Une forme originale de  
crowdfunding

D'après <http://gallica.bnf.fr/ark:/12148/btv1b8509563b>



16. Une Table d'encre de 2 pieds de long sur 3 de large.  
17. Deux Obusiers à vent et devant 2 boulets suspendus avec des chaînes au au. et des porcelaines.  
18. Un Matras pour recevoir le corps de la Table et leur résidu.  
19. Les Articles sans devant la Table chargé de recevoir les boulets dans le bassin et de les faire passer en avant le corps.


20. Deux Lances 3 pieds de diamètre pour faire marcher deux corps de pompe pour remplir les réservoirs d'eau.  
21. La Pompe sans entre les deux réservoirs qui s'est écartée.  
22. Deux Réservoirs d'eau (forme cylindrique).  
23. Un Corps de pompe dont le levier pour soulever les réservoirs qui les donne le mouvement. son même corps de pompe est sous chaque réservoir.

**SOUSCRIPTION**  
de 200 Billets de 50 Francs

pour l'Expérience d'un Ballon dirigé à volonté au moyen de deux obusiers à vent placés dans la nacelle et lançant en avant deux boulets suspendus au bout d'un **BALLON** en forme de **POISSON**.

Messieurs les Souscripteurs seront remboursés de leurs avances aussitôt que la recette égalera la dépense et de plus auront la moitié dans les bénéfices.

On s'inscrit chez **M<sup>rs</sup> le Chevalier de la Barbic**,  
rue d'Orléans, N<sup>o</sup> 16, Sculp<sup>ts</sup> S<sup>rs</sup> Marcq.

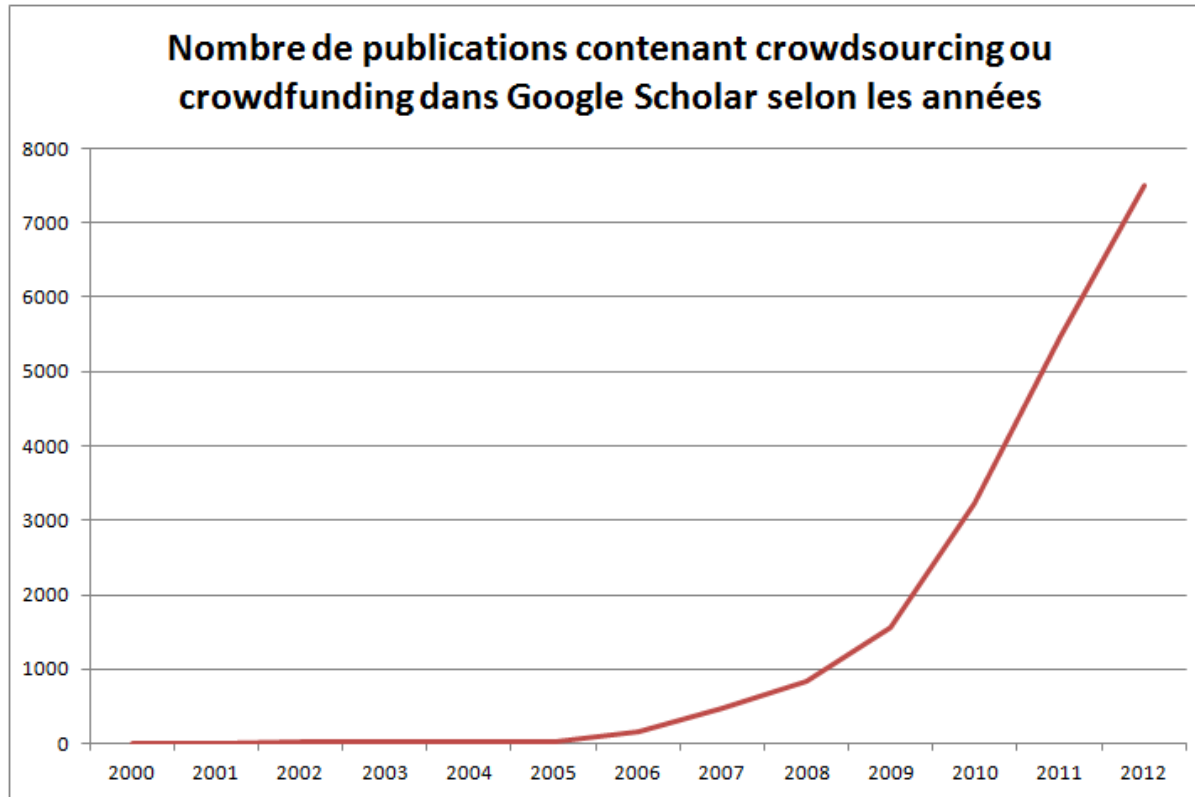


# Le crowdsourcing, une manne ?

Wikipedia = 100 millions d'heures bénévoles et 15,3 millions de dons en 2011. Mais aussi YouTube, Facebook...

Les américains regardent la télévision 200 billions d'heures par an. Ils pourraient donc créer 2000 projets comme Wikipédia chaque année au lieu de regarder la télévision (Shirky, 2010).

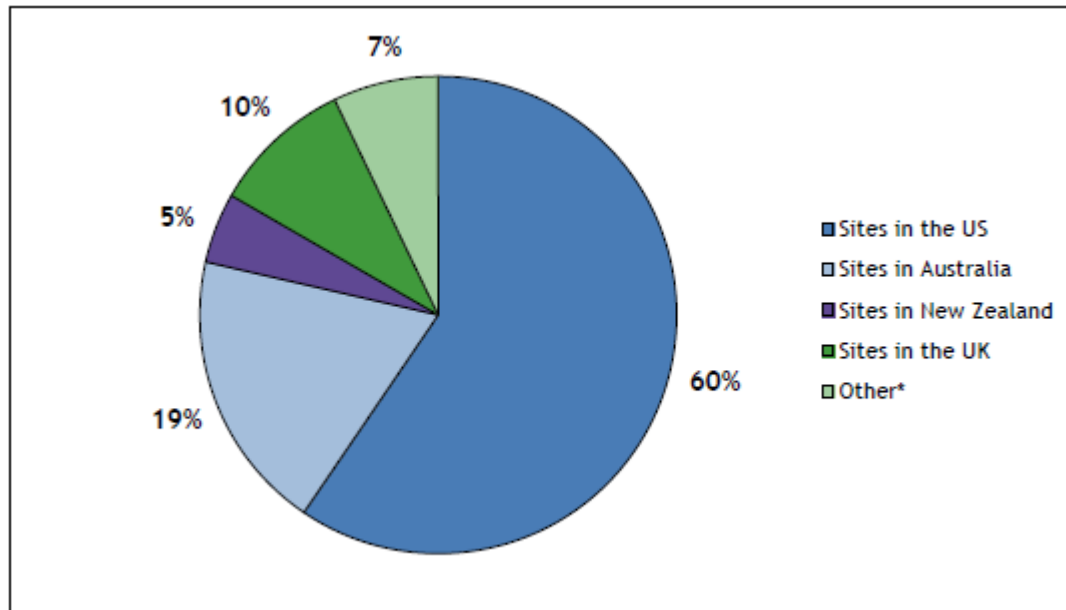
# Le crowdsourcing, un sujet d'étude



# Une opportunité pour les projets de numérisation du patrimoine

Mais la France tarde à en profiter :

Pas d'études avant 2013.  
Représentations des pays dans l'enquête menée par l'OCLC à propos des métadonnées sociales (d'après Smith-Yoshimura, 2011)





# Philosophie du crowdsourcing

- Marxisme et collectivisme ? “*de chacun selon ses capacités à chacun selon ses besoins*” (Louis Blanc repris par Karl Marx)
- Anarchisme libertaire ? (anti autoritaire, démocratie directe)
- Humanisme chrétien ? (amour de son prochain)
- Libéralisme ? (libertés individuelles émancipées des règles, esprit d’initiative et d’entreprise)

# Philosophie du crowdsourcing



Stakhanovisme de l'émulation socialiste qui organise une compétition entre les ateliers

=

Gamification du travail, fun at work ?



# Philosophie du crowdsourcing

- Économie participatiste
- Fin du salariat ?
- Disparition de la frontière entre loisirs et travail, entre consommation et production, entre vie privée et vie publique

# Les critiques

- Cauchemar totalitaire de l'idéologie 2.0
- Déterminisme de ceux qui rêvent de modéliser le monde
- Exploitation du travail gratuit ("servuction") ou sous payé car échappant à toute règle
- Fiscaliser les données ?

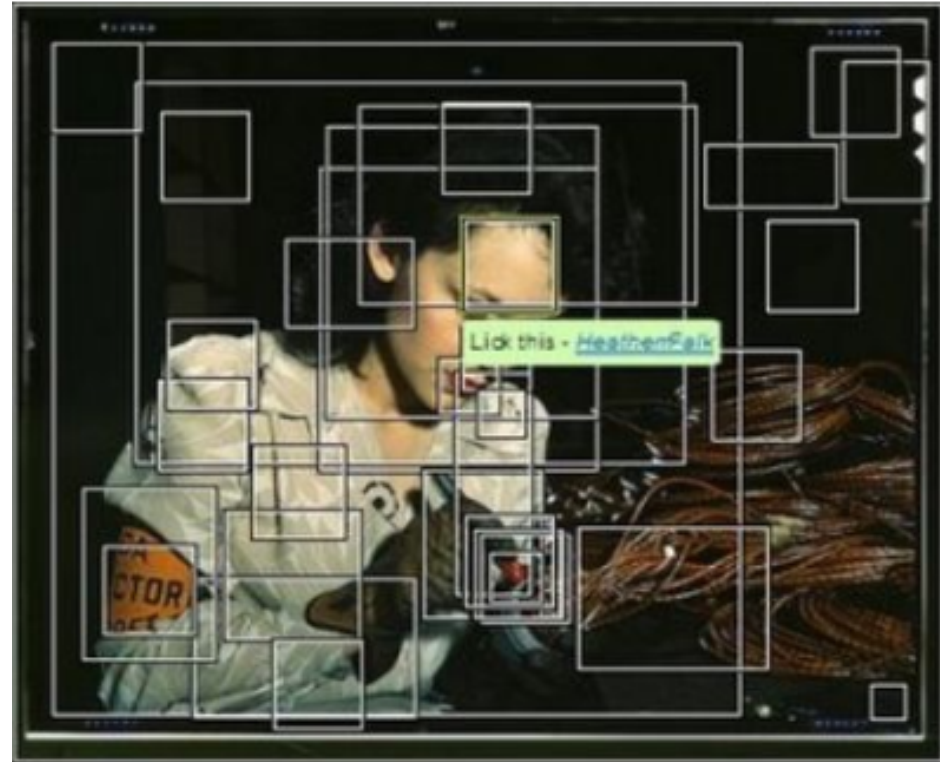


Ecce Homo (Elias García)  
avant et après restauration amateur

# Les critiques

Capture d'écran d'une photo taguée  
par la foule sur Flickr The Commons

<http://northwesthistory.blogspot.fr>



# La qualité en question

Appropriation individuelle du patrimoine collectif  
par le lambda, le trivial ?

Mais :

- L'internaute qui consulte connaît mieux le sujet
- Etudes comparatives

# En France

- Difficultés 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.
- Culture du secret, réseaux d'initiés, rétention d'information



11 % des salariés français utilisent des outils collaboratifs contre 25 % en Hollande (Etude Deloitte 2013)

# Nos institutions culturelles

- Hostilité idéologique à l'économie privée
- Nostalgie du monopole d'accès à l'information
- Crainte de se faire piller, donc copyright
- Refus de voir le travail effectué gratuitement par l'amateur, le profane, le privé qui remettent en question l'autorité, les compétences et le salaire des professionnels



# Mise en ligne participative

- Internet Archive
- Commons Wikimedia
- Europeana 1914-1918

# La numérisation à la demande

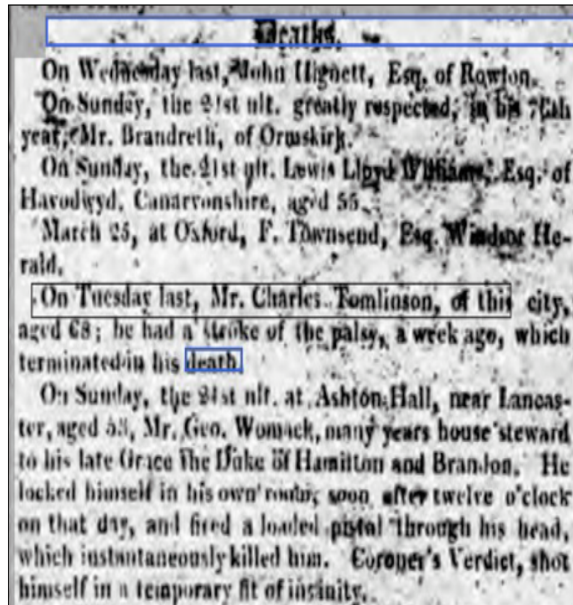
- Phénix Editions
- Chapitre.com
- Ebooks on Demand
- Adopter un livre pour Gallica
- Yabé : Numalire
- L'impression à la demande

# La correction de l'OCR

## raw OCR text

Deaths. In»rieff, Esq. of <c . Qn. Sunday, the till. greatly Drandrellt, of Orms4irJi.- ~ ; ✓ ' . \* On ijfr r inn ljjil F iij '11 f Havodivyd, Carnarvonshire, S ; \*#\* \*- ' « ' March Oxford, F. Tfovmeud, Uerald. » • V . • On Tnesdav last, Mr. Charles. IWilinson, this 8 ; had vf thesis#,, a week ago, which terminate<i'iu his death. . / ' ■ O'i Sunday, dJst nit. at AsbtCnvHall, mar Lancaster, Mr.,Geo. Worn ick, many years house'steward hit late Once The Hamilton and Brandon. He locked himself h»oWn'r<wte<: soon. twelve o'clock" that dny, and fii>d a loaded pistol "through Ins bead, 1 which instantaneously killed him. Coronet's Verdict, shot himself in a temporary fit of Friday week,

## newspaper image



# Wikisource

- 2008 : 100 thèses de l'École Nationale Vétérinaire de Toulouse
- 2010 : 1400 livres de la Bibliothèque nationale de France

## Modification de Page:Mommsen - Histoire romaine - Tome 3.djvu/94

G I

90 ILIVRE III, CHAPITRE III - 1  
foncière, il ne fut point touché. On n'avait point imaginé encore cette maxime des siècles postérieurs que toute terre non italique, conquise par les armes, devenait la propriété privée du peuple romain. De plus, en Sicile comme en Sardaigne, les villes continuèrent de s'administrer elles-mêmes, suivant la loi de leur ancienne autonomie; mais en même temps les démocraties sont partout supprimées; dans chaque cité le pouvoir est remis aux mains d'un conseil exclusivement aristocratique; un peu plus tard, en Sicile tout au moins, il se fait un recensement quinquennal, correspondant au *cens* de Rome. Mais ce sont là autant de modifications absolument exigées par la condition nouvelle des villes provinciales. Désormais soumises au gouvernement sénatorial de Rome, il n'y avait plus de place chez elles pour les *ecclesiae*, ou assemblées populaires à la grecque (*ἐκκλησία*). Il fallait que la métropole pût avoir l'œil sur les ressources militaires et financières de chacune, et d'ailleurs pareille chose était arrivée dans les pays conquis d'Italie.

Toutefois, si, au premier aspect, il semblait qu'il y eût égalité des droits entre les provinces et l'Italie, la réalité venait bien vite donner un grave démenti aux apparences. Les provinces n'avaient point de contingent régulier à fournir à l'armée ou à la flotte romaines. Le droit de porter les armes leur fut ôté, sauf au cas où le préteur local appelait des populations à la défense de leur patrie, Rome se réservant toujours d'envoyer des troupes italiennes dans les îles, en tel cas et en tel nombre, qu'il lui plaisait. A cette fin même, elle pré-

\* Anselm Hübner dit-il (Tit. Liv. 22, 37), qu'il n'est point bien que les Romains ne recrutent leur infanterie et leur cavalerie qu'avec les contingents romains ou italiens, et qu'ils n'admettent les « étrangers », que dans leurs troupes légères : (Militia etiam opibus aëre, nisi Romanis Italique; munitionibus nisi populis Romanorum; - latrocinio ac mercedem auxilia etiam externis viduae.)

Résumé :

Modification mineure  Suivre cette page État de la page ([«<a href="/wiki/Aide:Qualité/C3%A9\\_des\\_pages" title="Aide:Qualité des pages">Qualité des pages</a>](#))

# Bibliothèque Nationale de Finlande (Digitalkoot)

Entre le 8 février 2011 et le  
15 septembre 2011 :

- 80 000 vистeurs
- 4000 heures de jeu /  
travail
- 5 millions de saisies de  
mots



# ReCAPTCHA

Par jour :

- 200 millions de mots dont 88 validés
- 12 000 heures
- 1147 ebooks à 99,1%

Corrigés par des pays à bas coût, il faudrait environ 146 million d'euros / an.

The Norwich line steamboat train, from New-London for Boston, this **morning** ran off the track seven miles north of New-London.

morning

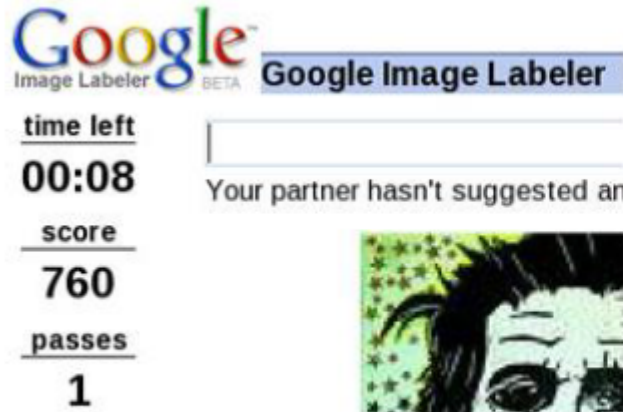
morning overtook

Type the two words:



# Google Image Labeler

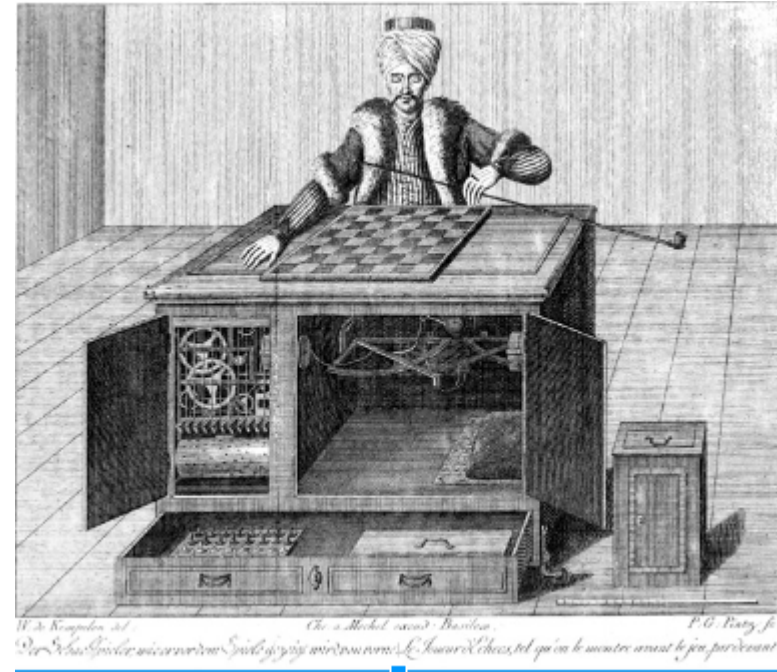
- ESP game
- GWAP
- Peekaboom
- Art Collector
- metadata games



Captures d'écran de Google Image Labeler et ESP Game (d'après Vohn Ahn, 2004)

# Amazon Mechanical Turk Marketplace

- 500 000 travailleurs de 190 pays
- En avril 2010 : 6 701 406 HITs proposés par 9436 sociétés pour une valeur marchande de 529 259 \$
- 700 000 HITs par semaine
- Jeunes américains + indiens



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



# Taxonomie du crowdsourcing

	<b>Volontaire</b>	<b>Gamification</b>	<b>Non volontaire ou implicite</b>
<b>Rémunéré</b>	Amazon Turk Marketplace	<i>Récompenser les meilleurs scores ?</i>	<i>fiscaliser les données ?</i>
<b>Non rémunéré</b>	Wikisource	Google Image Labeler	ReCaptcha

# Les clés du succès

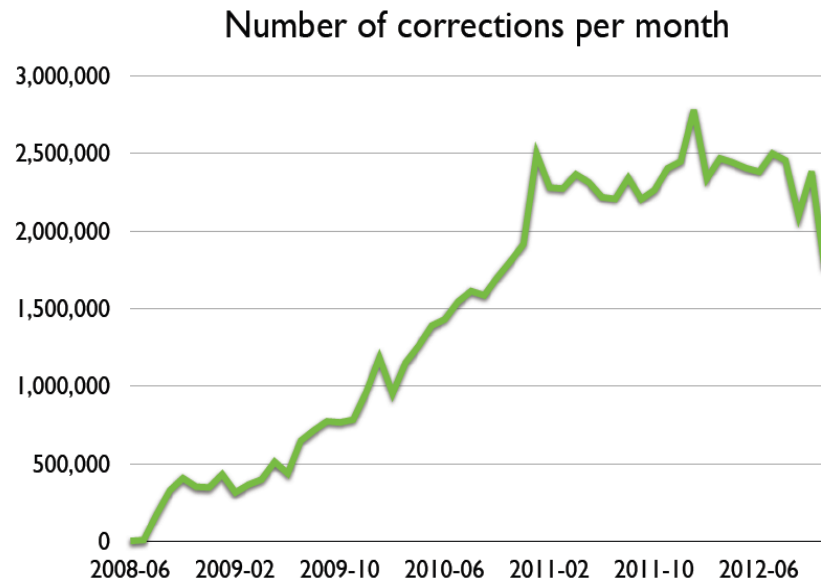
- Communication et recrutement
- Motivations, récompenses
- Gamification
- Community Management
- Contrôle, évaluation de la qualité, formation
- Réintégration des données produites
- Statistiques, indicateurs

# Pourquoi les internautes contribuent

- Se sentir utile pour le groupe, la société, le pays, la science, l'intérêt général, une cause
- Faire quelque chose de désintéressé, pas pour de l'argent dans un esprit d'altruisme et de philanthropie ou avec un sentiment de redevabilité
- Pour le développement personnel, se cultiver et apprendre, satisfaire sa soif de connaissance
- Pour s'amuser, se distraire, jouer (gamification) ou tester une démarche innovante
- Par esprit de compétition, de challenge ou pour se prouver quelque chose
- Avoir une bonne estime de soi, avoir l'impression d'avoir du pouvoir sur les choses, d'être auteur et acteur
- Améliorer son e-reputation

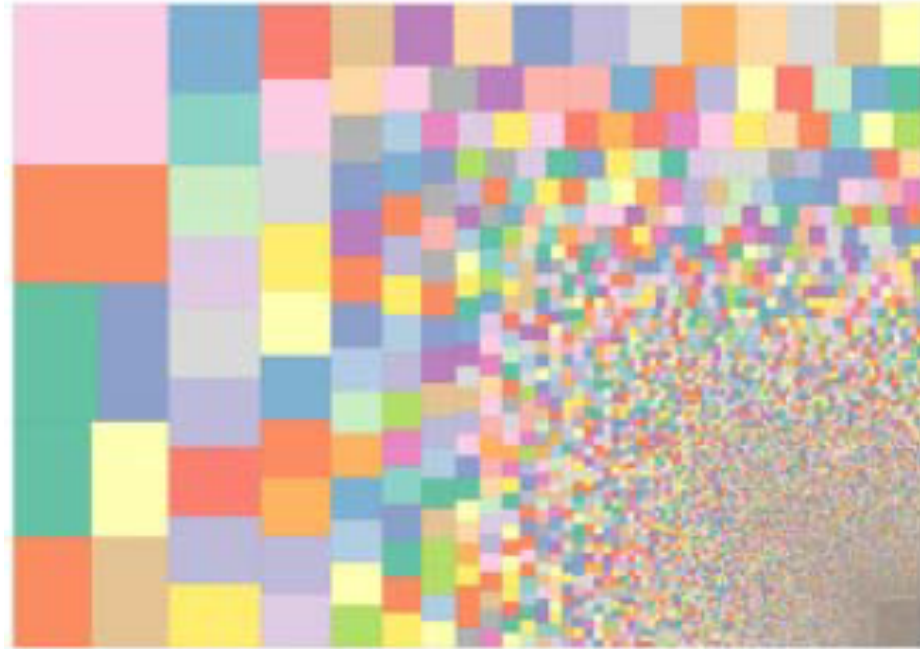
# La fin du crowdsourcing sur OCR ?

- Google Image Labeler abandonné par Google
- Les progrès dans l'OCR



L'évolution du nombre de corrections sur TROVE, d'après (Hagon, 2013)

# Communitysourcing plutôt que crowdsourcing ?



**Quelques internautes produisent la majeure partie des contributions (Brumfield, 2013)**

# Une expérimentation crowdfunding

Proposer aux internautes de participer au financement de la numérisation de tel ou tel livre parmi le patrimoine conservé par les bibliothèques via [Numalire](#)

- Pour l'utilisateur : service de reproduction numérique
- Pour le privé : prestataires, mécènes, investisseurs
- Pour l'Inra : nouveaux services sans coûts, enrichit la bibliothèque numérique, ouverture de la politique documentaire

# Numalire

## Résultat :

En 8 mois : 414 demandes de devis, 36 numérisations

## En réflexion :

- Calcul automatique de prix
- Diminution et partage des coûts
- Ouverture à d'autres bibliothèques
- Elargissement à d'autres types de documents

# Les sciences participatives

Dans d'autres domaines :

- [galaxyzoo.org](http://galaxyzoo.org) : un million de photographies d'astronomie qualifiées par près de 100 000 internautes
- [fold.it](http://fold.it) : améliorer la connaissance du repliement des protéines grâce à l'intelligence humaine collective mobilisée autour de jeux de puzzles
- Birdify : le Shazam des chants d'oiseaux



# Conclusion

- Amateurs = ruptures innovantes car ne cherchent pas à reproduire les modèles établis du métier
- Von Hippel 46 % des entreprises américaines dans des domaines innovants ont pour origine un simple utilisateur
- Conduite du changement

# Merci pour votre attention

Contact : [mathieu.andro@versailles.inra.fr](mailto:mathieu.andro@versailles.inra.fr)

Diaporama : <http://tinyurl.com/o2kzfj8>

# Une définition

*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 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.*

# Bibliographie

- Andro, M., Saleh, I. (2014). Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet. Colloque International sur le Document Numérique, CIDE 17 (sous presse)
- Andro, M., Rivière, P., Dupuy-Olivier, A., Gropallo, F., Maingreud, D. (2014). [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 2014, 9 p.
- Andro, M. (2014). [Crowdsourcing and digitization: Presentation for the Ebooks on Demand Conference 2014 \(Innsbruck University, April 11th 2014\)](#).