

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. hal-01164625

HAL Id: hal-01164625
https://enpc.hal.science/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 ?

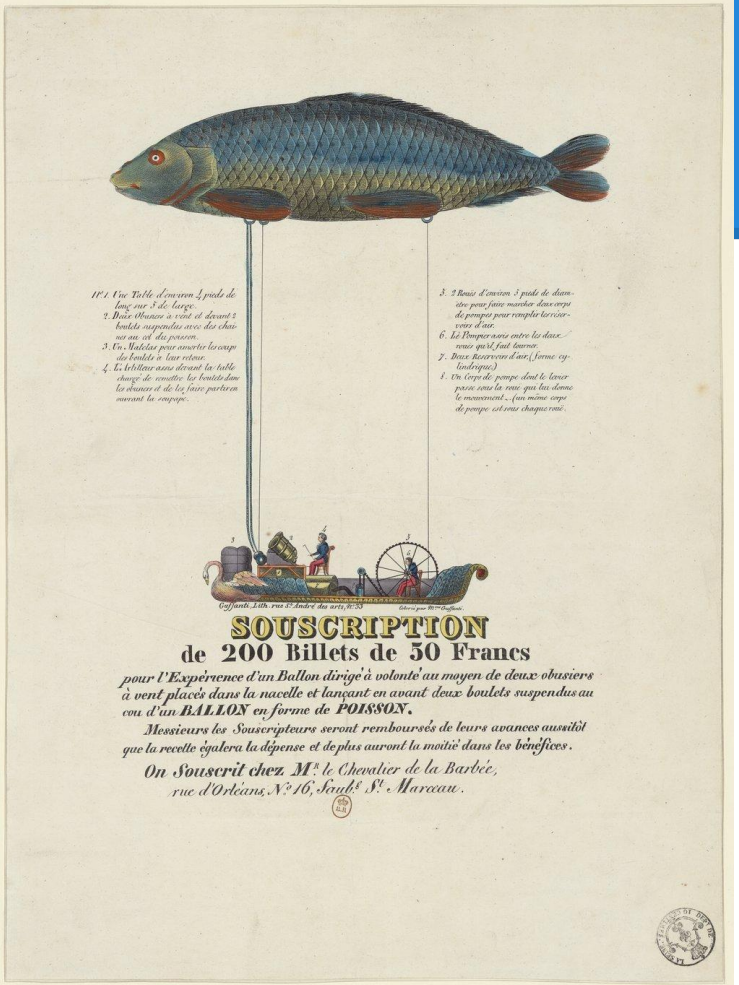
→ travail, argent, compétences ou connaissances moins limitées

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

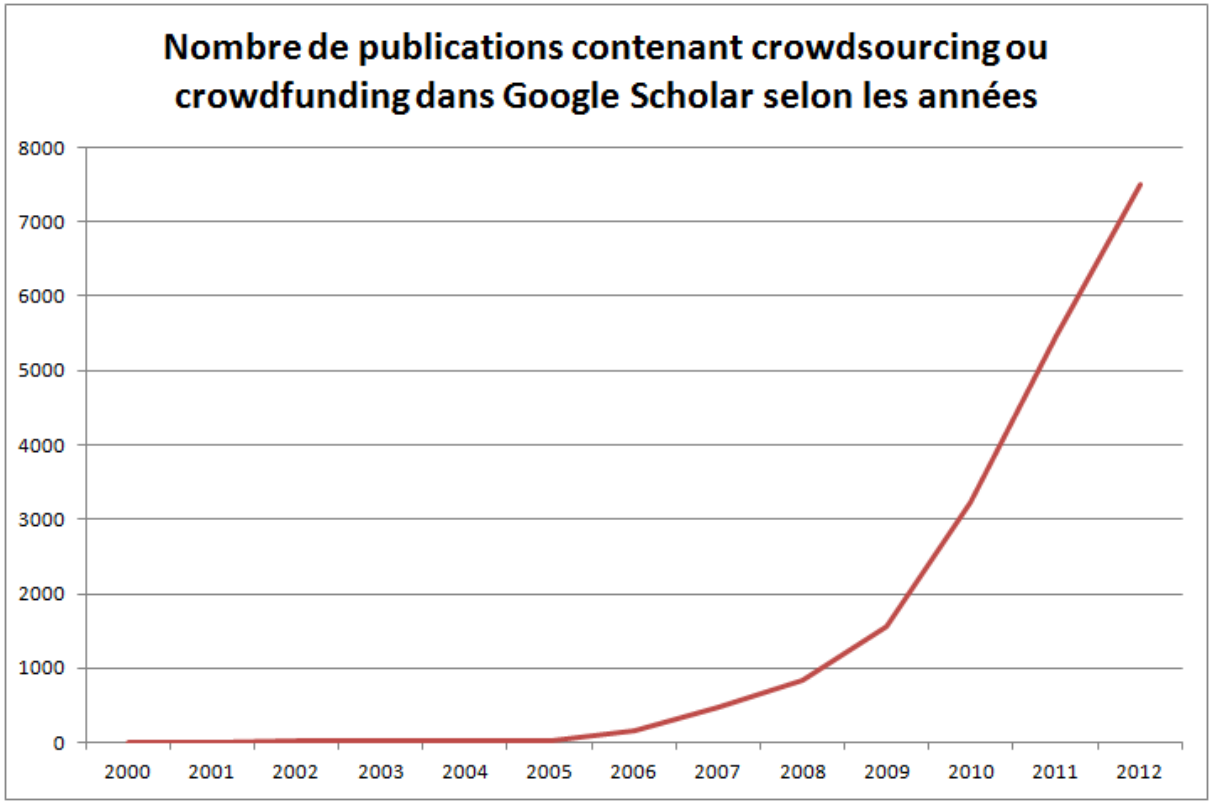


Le crowdsourcing, une manne ?

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).

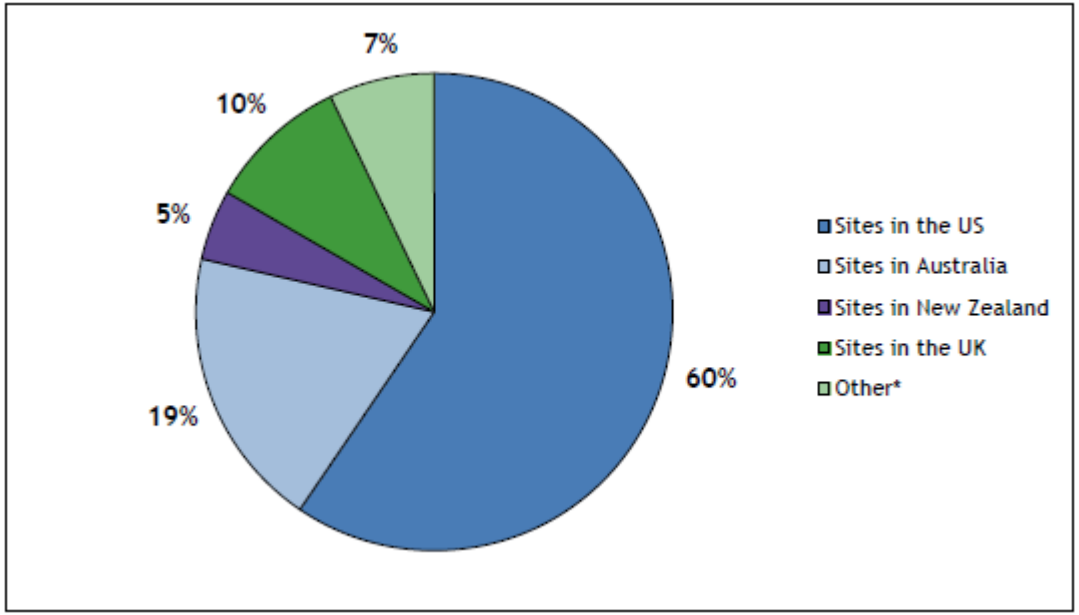
Le crowdsourcing, un sujet d'étude



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

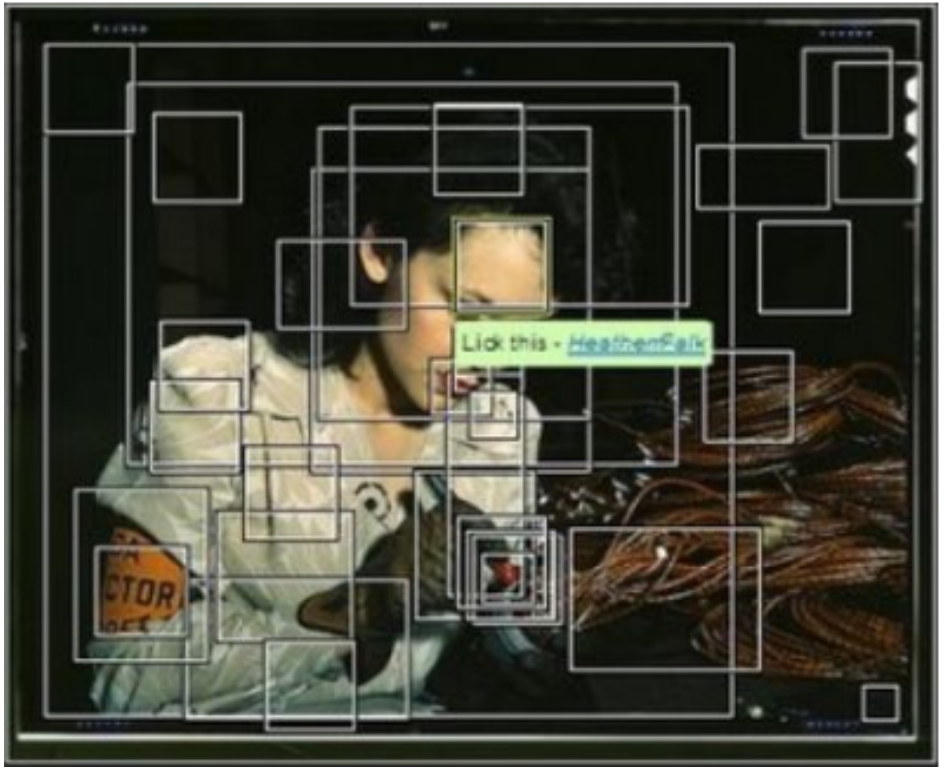
- Cauchemar totalitaire de l'idéologie 2.0
- Déterminisme de ceux qui rêvent de modéliser le monde
- Exploitation du travail gratuit ("servuction") ou sous payé car échappant à toute règle
- Fiscaliser les données ?



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

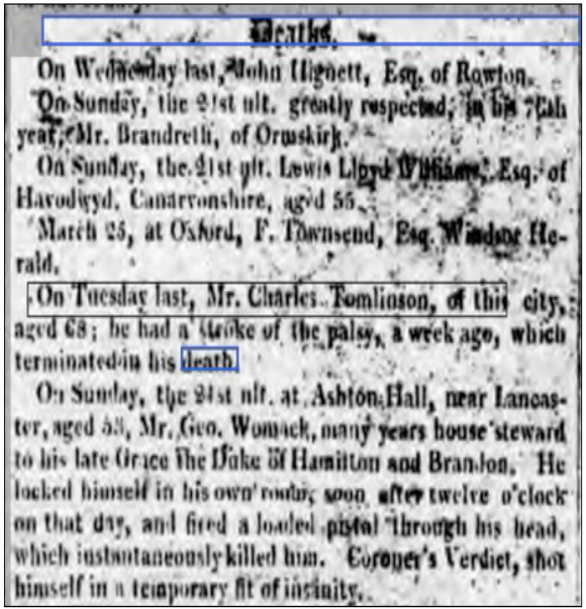
- Phénix Editions
- Chapitre.com
- Ebooks on Demand
- Adopter un livre pour Gallica
- Yabé : Numalire
- L’impression à la demande

La correction de l'OCR

raw OCR text

Deaths. **lln»rieff**, Esq. of <c .. Qn. Sunday, the till. greatly **Drandrellt**, of **Orms4\irJi.**- ~ ; ;√ ' • * On **ijfr r inn ljjjil F iij '11 f Havodivyd**, **Carnarvonshire**, S ; **" *- ' « ' March **Oxford**, **F. Tfovmeud**, **Uerald. » • V**. •On Tncsdav 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



Deaths.

On Wednesday last, John Liggett, Esq. of Rawton.

On Sunday, the 21st ult. greatly respected, in his 74th year, Mr. Brandreth, of Orrnskirk.

On Sunday, the 21st ult. Lewis Lloyd Williams, Esq. of Havodivyd, Carnarvonshire, aged 55.

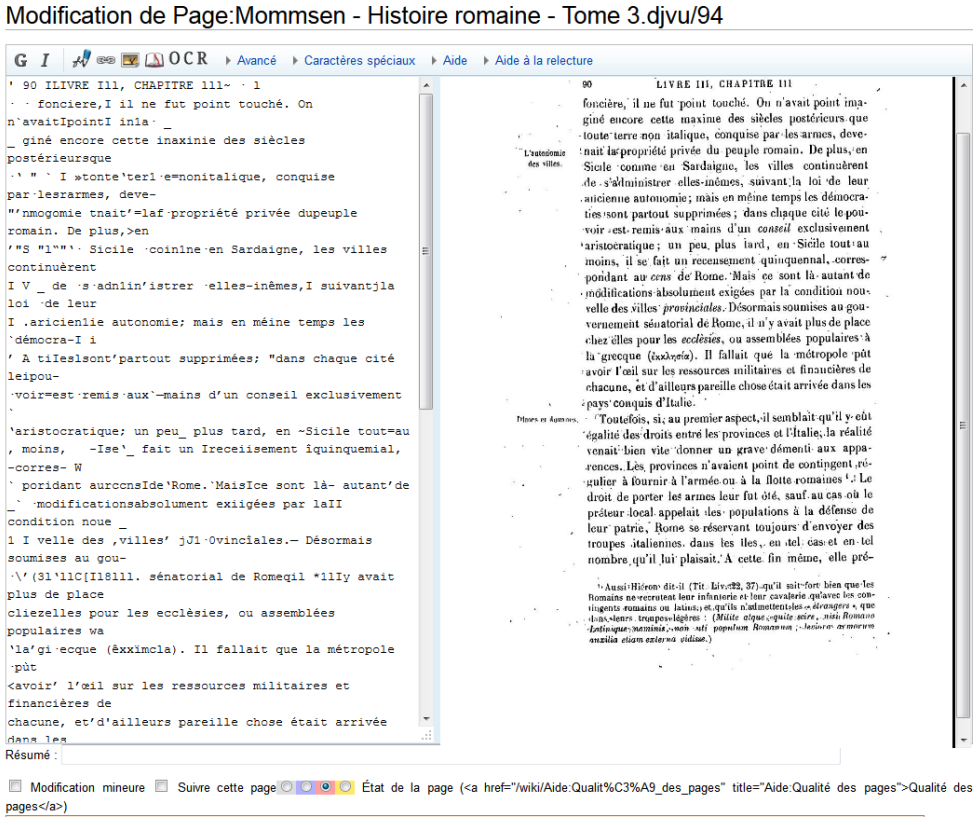
March 25, at Oxford, F. Townsend, Esq. Windsor Herald.

On Tuesday last, Mr. Charles Tomlinson, of this city, aged 68; he had a stroke of the palsy, a week ago, which terminated in his death.

On Sunday, the 21st ult. at Ashton Hall, near Lancaster, aged 53, Mr. Geo. Womack, many years house steward to his late Grace the Duke of Hamilton and Brandon. He locked himself in his own room, soon after twelve o'clock on that day, and fired a loaded pistol through his head, which instantaneously killed him. Coroner's Verdict, shot himself in a temporary fit of insanity.

Wikisource

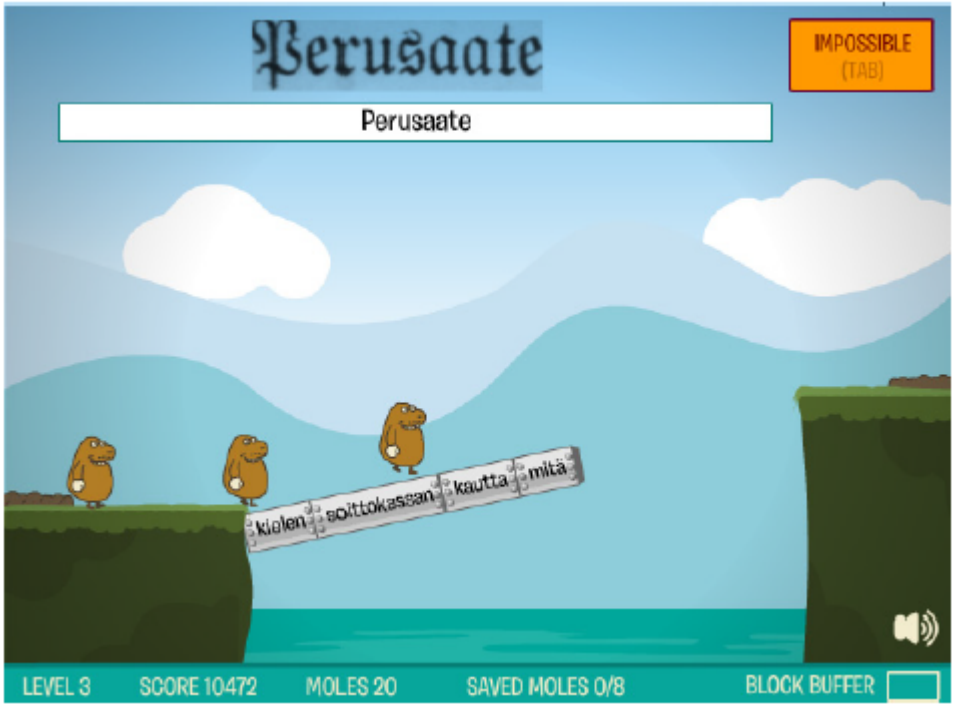
- 2008 : 100 thèses de l'École Nationale Vétérinaire de Toulouse
- 2010 : 1400 livres de la Bibliothèque nationale de France



Bibliothèque Nationale de Finlande (Digitalkoot)

Entre le 8 février 2011 et le 15 septembre 2011 :

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

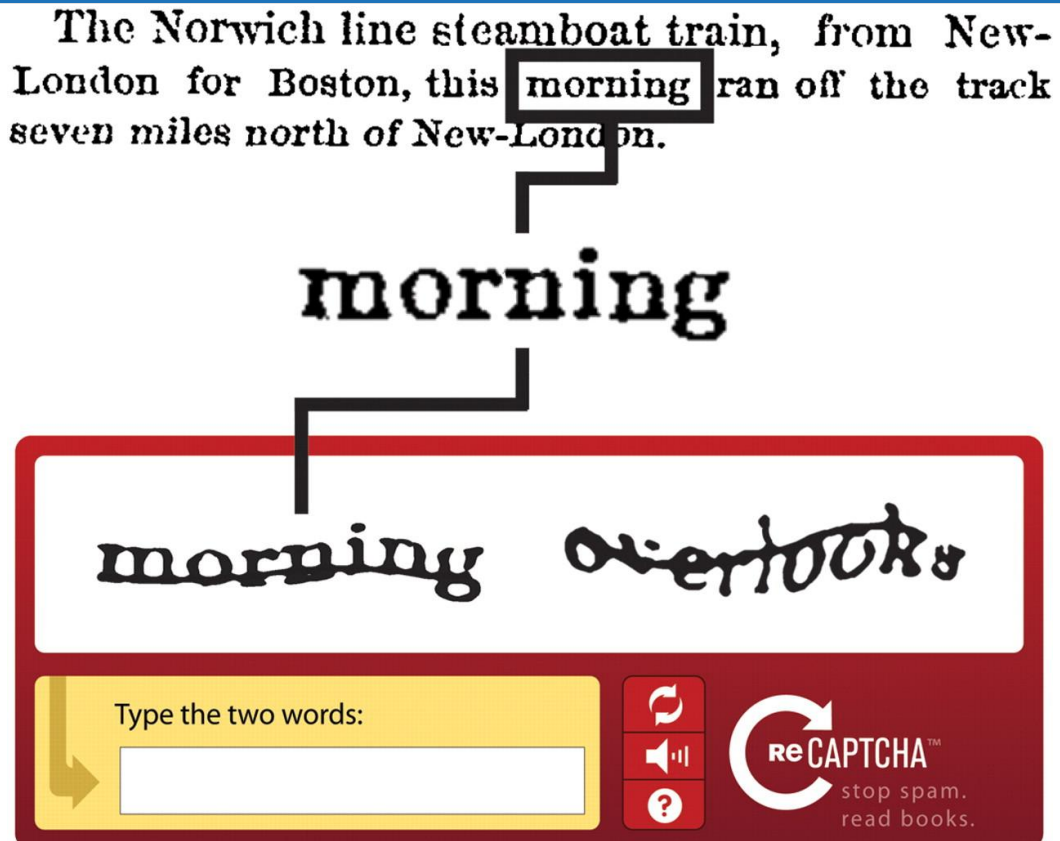


ReCAPTCHA

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.



Google Image Labeler

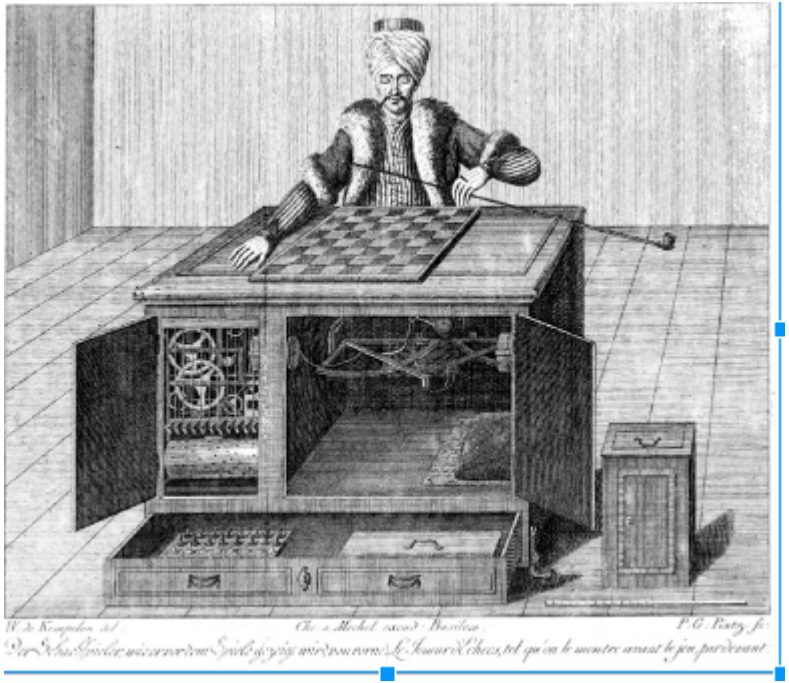
- ESP game
- GWAP
- Peekaboom
- Art Collector
- metadata games



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

	Volontaire	**Gamification**	**Non volontaire ou implicite**
Rémunéré	Amazon Turk Marketplace	*Récompenser les meilleurs scores ?*	*fiscaliser les données ?*
Non rémunéré	Wikisource	Google Image Labeler	ReCaptcha

Les clés du succès

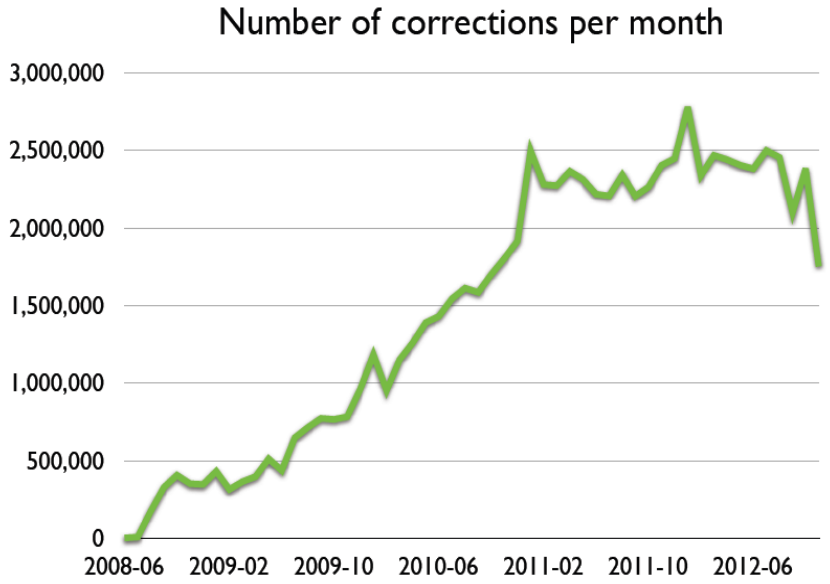
- Communication et recrutement
- Motivations, récompenses
- Gamification
- Community Management
- Contrôle, évaluation de la qualité, formation
- Réintégration des données produites
- Statistiques, indicateurs

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

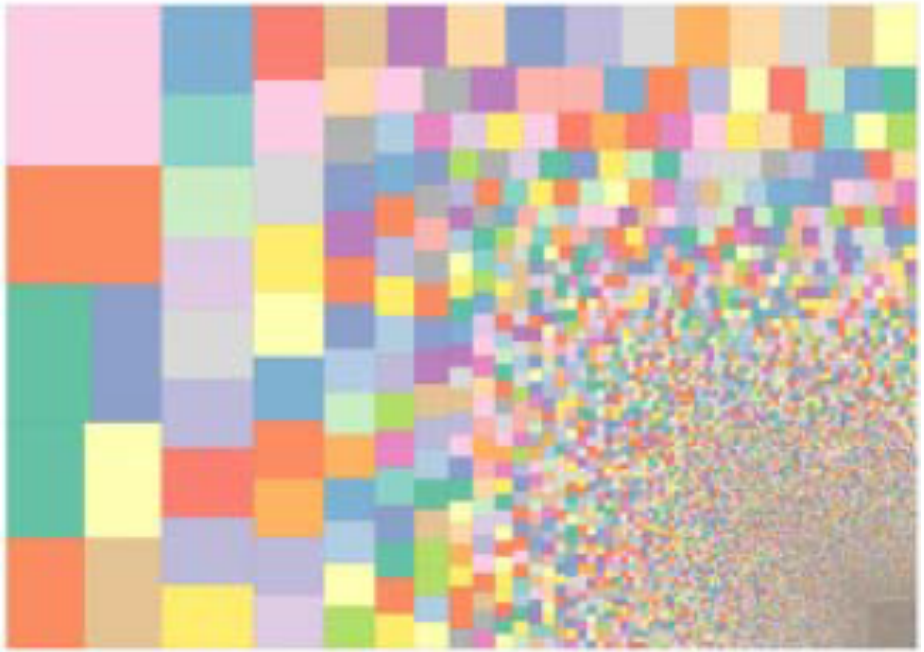
La fin du crowdsourcing sur OCR ?

- 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'usager : 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 : un million de photographies d’ astronomie qualifiées par près de 100 000 internautes
- 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

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., Maingreaud, 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).