

# Sciences citoyennes

## le crowdsourcing appliqué à la recherche

Mathieu Andro (DIST)



Wikipedia = 100 millions d'heures bénévoles.  
Télévision = 200 milliards d'heures / an (USA).

*Les américains pourraient donc créer  
2000 projets comme Wikipédia chaque année  
au lieu de regarder la télévision  
(Shirky, 2010).*

# C'est quoi le crowdsourcing ?

Juin 2006, Jeff Howe, Wired Magazine.

**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, compétences ou connaissances moins limitées.

→ argent (crowdfunding).

# Philosophie du crowdsourcing

- Marxisme et collectivisme ?
  - Anarchisme libertaire ?
  - Humanisme chrétien ?
  - Libéralisme ?
- Idéologie californienne (libérale libertaire)

# Philosophie du crowdsourcing



Stakhanovisme  
émulation socialiste  
compétition entre les  
ateliers

=

Gamification du  
travail, fun at work ?



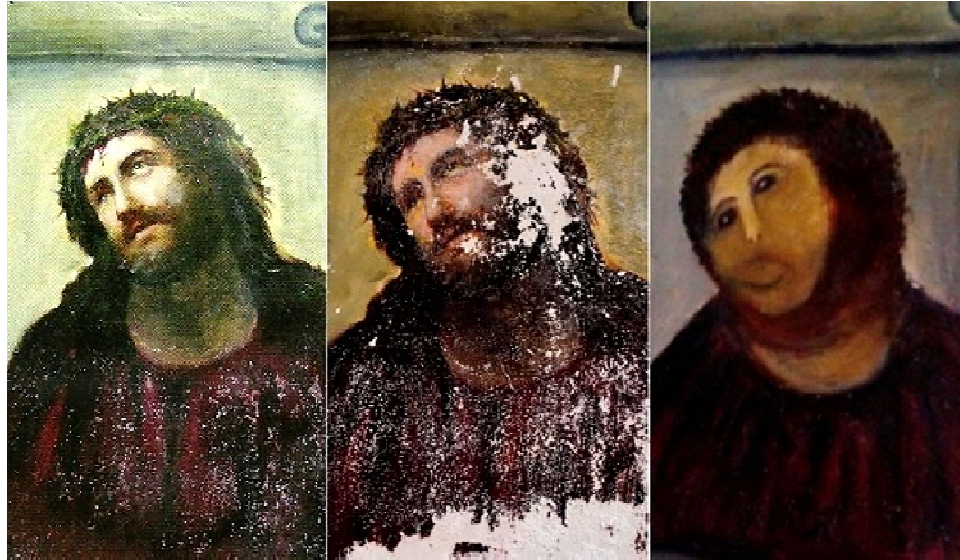
# Philosophie du crowdsourcing

- Économie participatiste, fin du salariat ?
- Disparition de la frontière entre :

|                     |                         |
|---------------------|-------------------------|
| ludique             | sérieux                 |
| loisirs             | travail                 |
| privé               | public                  |
| amateur<br>bénévole | expert<br>professionnel |
| consommation        | production              |

# Les critiques de fond

- Cauchemar totalitaire de l'idéologie déterministe 2.0 ?
- Remplacer les salariés par des bénévoles ?
- 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 sur les résultats

- Capture d'écran d'une photo taguée par la foule sur Flickr The Commons
- Rapport entre les coûts et les bénéfices
- Communication institutionnelle



<http://northwesthistory.blogspot.fr>



# Les origines historiques du concept

1707 : accident de Cloudeley Shovell aux îles Scilly.

1714 : Longitude Act  
(20 000 livres à qui permettrait de déterminer la longitude d'un navire en pleine mer et d'éviter de nouveaux accidents).

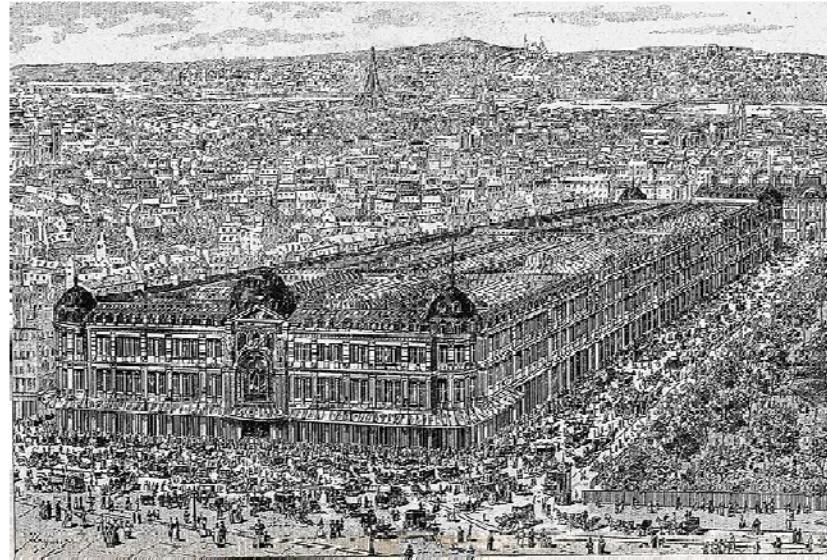


Michael Dahl [Public domain], via Wikimedia Commons

# L'invention du libre service

1852 : “Au bon marché”.

Le consommateur est progressivement intégré au processus de production : caisses automatiques, distributeurs de billets, cafétérias, fast food, VeLib, Ikea...

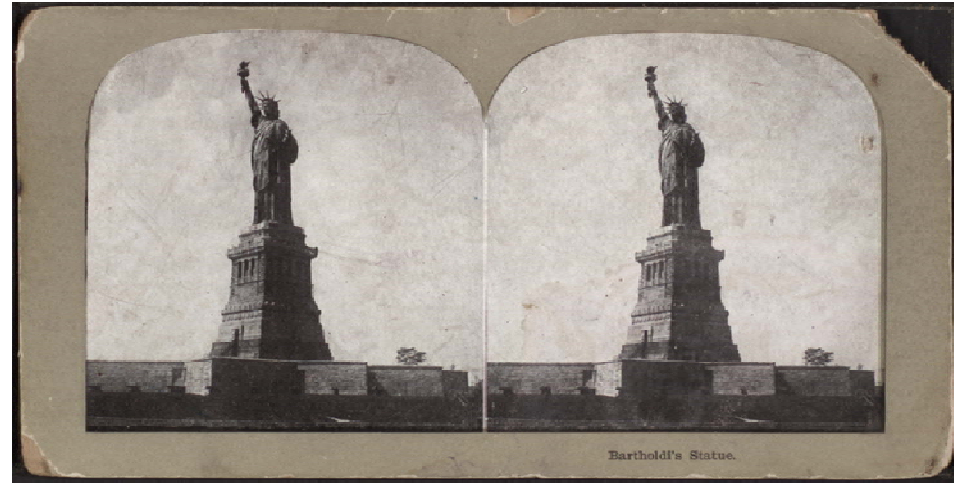


Fonds Boucicaut [CC0], via Wikimedia Commons

# L'invention de la souscription publique

1884 : statue de la Liberté.

Elle est financée suite à une souscription publique de 125 000 personnes qui commença en 1875 en France.

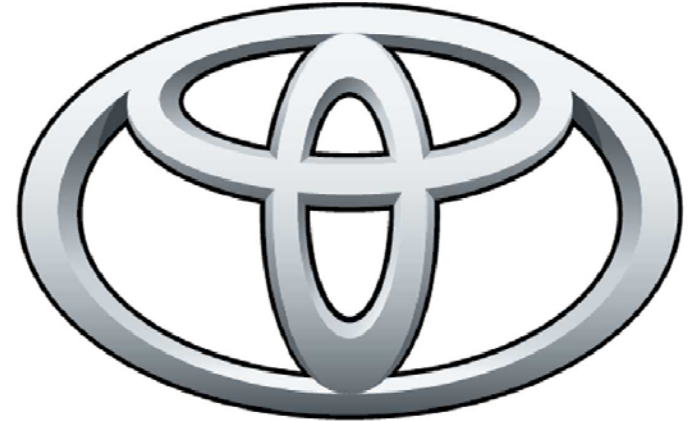


[Public domain], via Wikimedia Commons

# Les premiers concours créatifs

1936 : Logo de Toyota.

Toyota rassembla 27 000 personnes et choisit le meilleur dessin proposé comme logo de sa marque.



By Tom777 (Own work) [CC BY-SA 4.0 (<http://creativecommons.org/licenses/by-sa/4.0>)], via Wikimedia Commons

# Le “just in time” puis le “on demand”

Années 1950 : modèles “just in time” puis “on demand”.

- Toyota
  - sans stocks ni invendus
  - en flux tendus, offre en fonction de la demande
  - centré sur le consommateur
- Externaliser auprès du consommateur, la décision de mise en production elle-même.

# Le crowdsourcing explicite classique

2001 : Wikipedia.

Web 2.0 (DiNucci, 1999 ;  
Tim O'Reilly, 2005)



**WIKIPÉDIA**  
*L'Encyclopédie libre*

By Wikipedia-logo-de.png: Nohat (concept by Paullusmagnus); Text by Andre Riemann, Text in last revision by Mandavi derivative work: Otourly (Wikipedia-logo-de.png) [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0/>)], via Wikimedia Commons

# Les sciences citoyennes

Entre le 22 septembre  
2012 et le 22 février  
2013, 62 133  
contenus de champs  
ont été complétés sur  
2 456 parts d'herbier

The screenshot displays the 'Les herbonautes' website interface. At the top, there are navigation links: 'Pourquoi ce site?', 'L'Herbier de Paris', 'FAQ', 'Consignes', and 'Quel herbonaute êtes-vous?'. The main header includes the site logo 'Les herbonautes' with the tagline 'L'herbier numérique collaboratif citoyen', a search bar with the text 'Accès rapide', and a 'Se connecter' link.

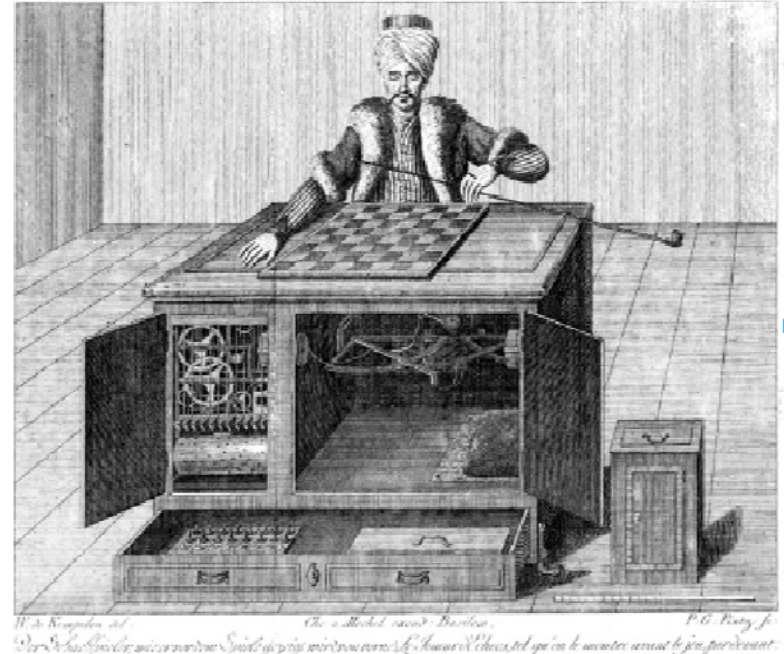
The central part of the page shows a user profile for 'Protectrice d'Afrique (et de Madagascar)'. The profile includes a profile picture, the name 'Leucadendron', a user ID '3371185511', and a unique identifier 'MNIIN/P/P023/U234'. There are buttons for 'Plus d'informations' and 'Spécimen suivi'. Below this is a yellow box with the text 'Pour résoudre le conflit, cliquez sur le symbole'. A 'Pays' section contains a dropdown menu for 'Sélectionnez un pays', checkboxes for 'Je l'ai déduit', 'Thésiste', and 'Pas d'information', and a 'Valider' button.

The right side of the page features a large image of a herbarium specimen, showing two dried plant branches on a light-colored background. Below the specimen are several labels. One label reads 'Leucadendron lewisii Willd.' and 'Détournement de l'Herbier de Paris le 17/10/14'. Another label is from the 'EX NATIONAL HERBARIUM, PRETORIA' and contains handwritten text: 'No. 1000', 'Spéc. Leucadendron, C. B.', 'Leucadendron tortuosum R. Br.', 'Leucadendron Parlati Montanari', 'Alpes - f. - Arbres à ft.', 'Alt. 2000-11-1930', 'coll. B. B. G. G. G.', and 'No. 1000'. A third label is a barcode with the text 'Herbier Muséum Paris' and 'Paris 1024'. A fourth label at the bottom right says 'HERB. MUS. PARIS' and 'Rec. le 28. Avril 1952'.

# Le crowdsourcing rémunéré

2005 : Amazon Mechanical Turk Marketplace.

- 500 000 travailleurs de 190 pays
- Jeunes américains + indiens
- En avril 2010 : 6 701 406 HITs proposés par 9436 sociétés pour une valeur marchande de 529 259 \$
- 700 000 HITs par semaine
- Travailler où on veut, quand on veut, pour qui on veut, être tantôt employeur tantôt employé



« Tuerkischer schachspieler windisch<sup>4</sup> » par  
Karl Gottlieb von Windisch. 1783.

Public domain via Wikimedia Common



# Le crowdsourcing implicite (involontaire)

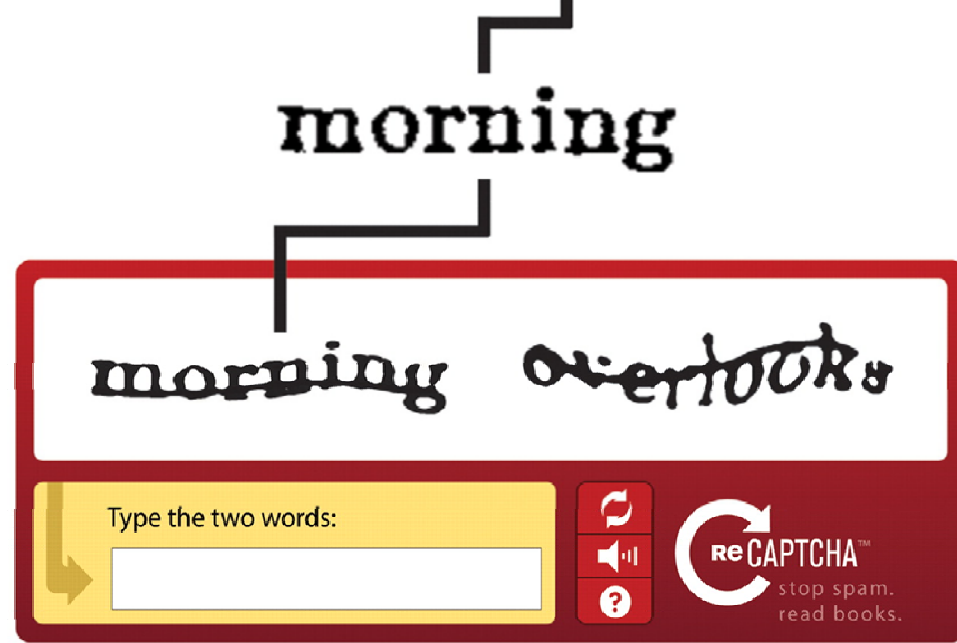
2007 : reCAPTCHA.

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.



# Le crowdsourcing implicite (involontaire)

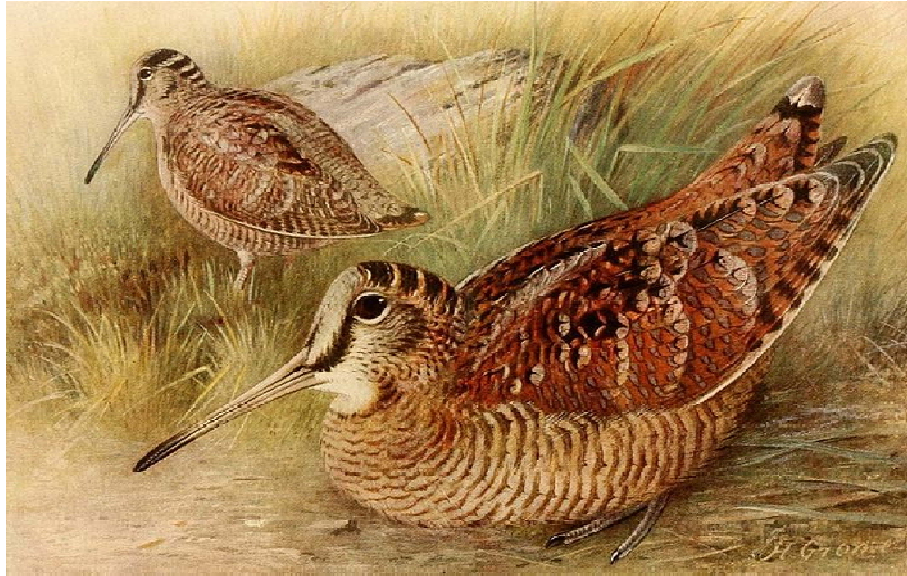


This image was created by user Bob (Bobzimmer) at Mushroom Observer, a source for mycological images. [CC BY-SA 3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons



Les recherches de morilles en fonction du jour de l'année d'après Google Trends et [www.identifier-les-champignons.com](http://www.identifier-les-champignons.com)

# Birdify, le shazam pour chants d'oiseaux

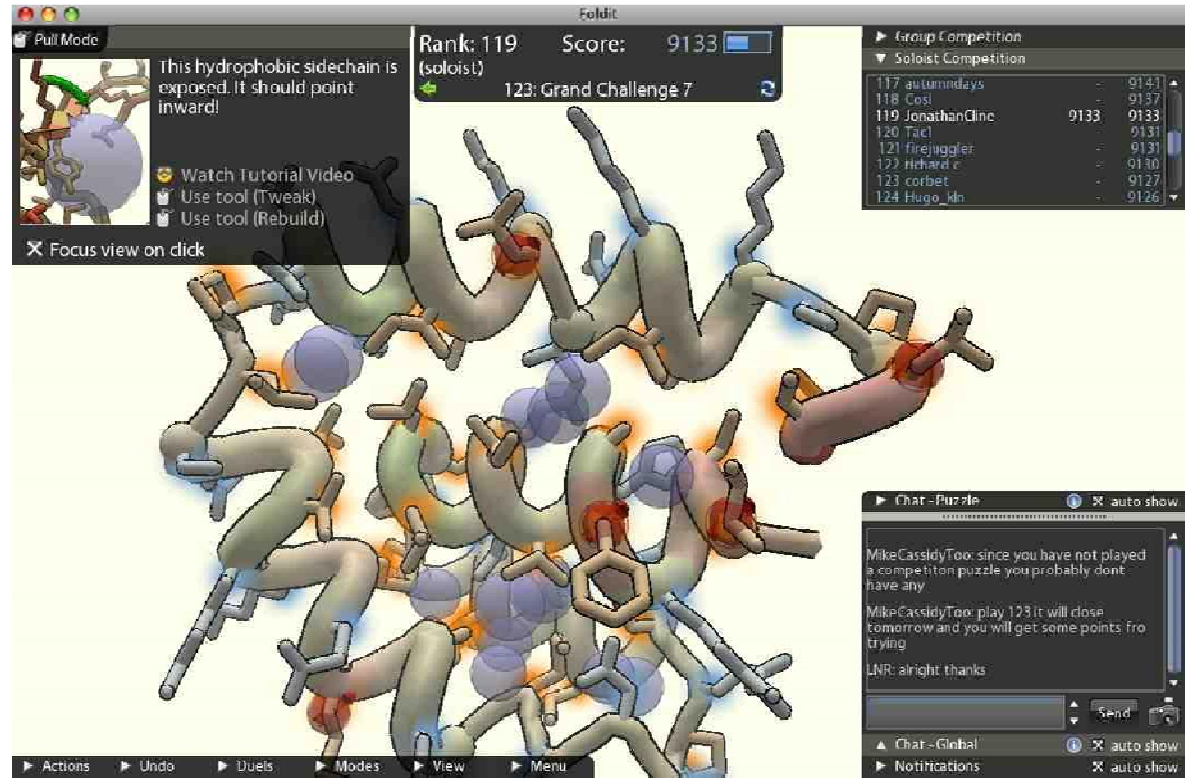


Henrik Grönvold [Public domain], via Wikimedia Commons

# La gamification (Games With A Purpose)

2008 : Fold.it

Jeux puzzles  
contribuant à  
une meilleure  
connaissance des  
protéines et du  
virus du SIDA.



# La sagesse des foules

2011 : Good judgement project.

*“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.” (Machiavel)*

En 1893, Francis Galton : concours pour deviner le poids d’un boeuf.

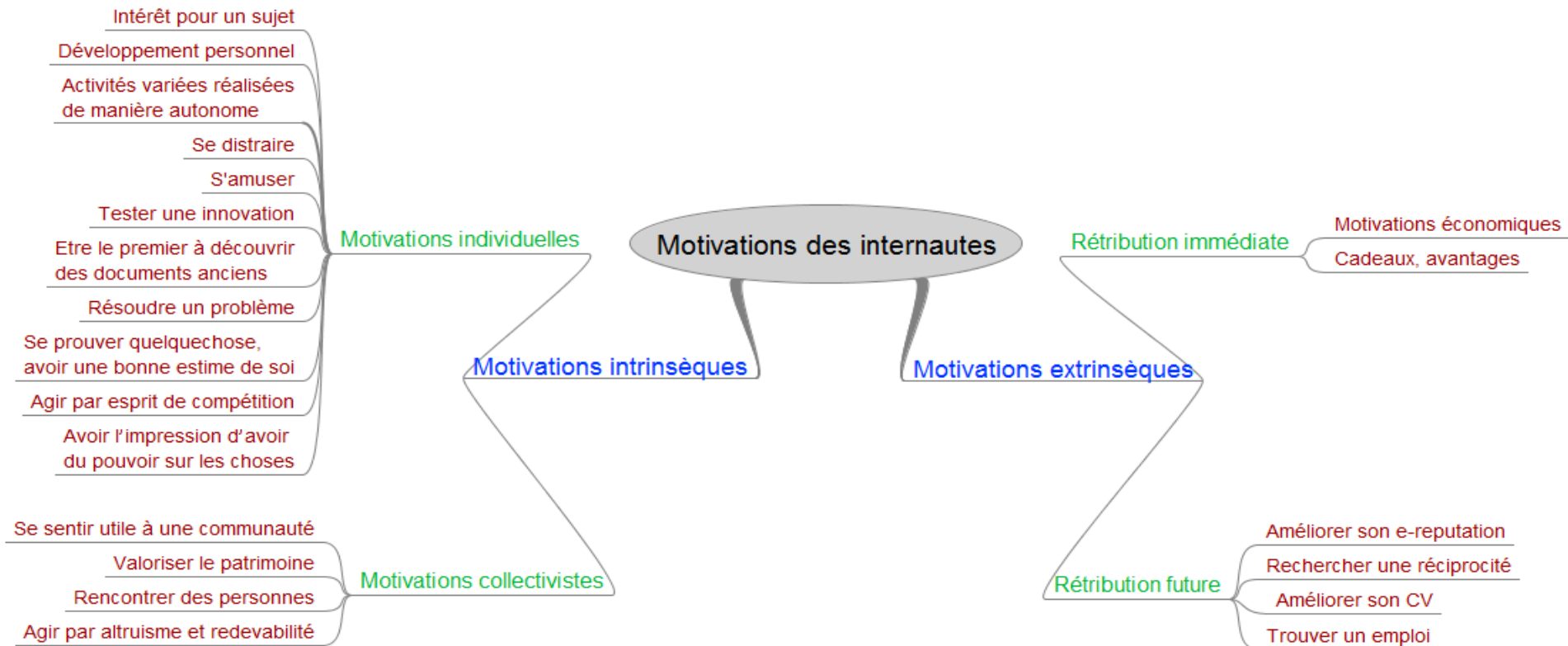


Santi di Tito [Public domain], via Wikimedia Commons

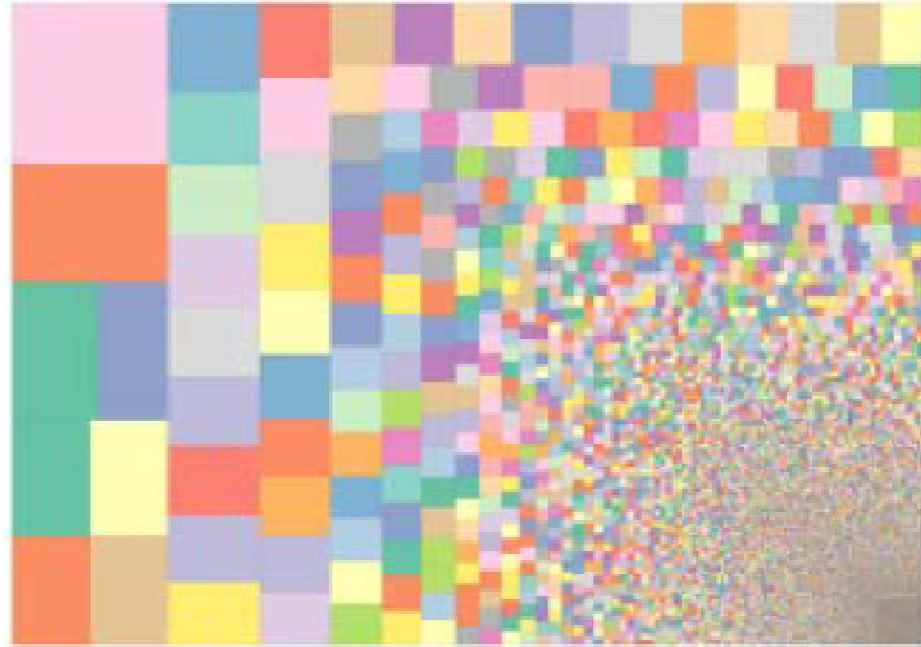
# Les clés du succès

- Communication et recrutement
- Motivations, récompenses
- Community Management
- Contrôle, évaluation de la qualité, formation
- Finalité : travail / participation ?
- Réintégration des données produites
- Statistiques, indicateurs, bénéfiques / coûts

# Pourquoi les internautes contribuent



# Crowdsourcing et communitysourcing



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



# Quelques premières expérimentations à la DIST de l'Inra

- Wikisource (correction de l'OCR)
- Numalire (crowdfunding)
- Guru.com (annotations pour text mining)
- Rapport sur les sciences participatives

→ une thèse en cours sur le crowdsourcing

## Bibliothèques numériques et crowdsourcing : concept, panorama, état de l'art et expérimentations

[Présentation de la thèse](#) | [Classement des contributeurs](#) | [Bibliographie utilisée dans la thèse](#) | [CV](#)

J'ai entrepris de creuser le sujet du crowdsourcing appliqué à la numérisation et aux bibliothèques numériques dans le cadre d'une thèse.

Cette thèse est rédigée sous la forme de Google Docs auxquels je vous donne accès en temps réel :

|   |   |
|---|---|
| 1- <b>Partie conceptuelle</b> : définitions, origines, philosophies politiques, économiques, managériales   | 93 pages déjà ouvertes à vos commentaires                           |
| 2 <b>Panorama des projets</b>   | 117 pages qui seront ouvertes à vos commentaires le 1er juin 2015   |
| 3- <b>Synthèse et état de l'art</b> : taxonomie des projets, motivations des internautes, community management, qualité des données produites, évaluation des projets, conduite du changement | 86 pages qui seront ouvertes à vos commentaires le 1er juillet 2015 |
| 4- <b>Expérimentations</b> conduites principalement autour d'un projet de numérisation à la demande par crowdfunding  | 143 pages qui seront ouvertes à vos commentaires le 1er août 2015   |

Bien que la thèse soit encore inachevée, vous pouvez néanmoins la citer :  
Andro, M. (2015). Bibliothèques numériques et crowdsourcing : concept, panorama, état de l'art et expérimentations autour de Numalire, projet de numérisation à la demande par crowdfunding. Thèse en cours en Sciences de l'Information et de la Communication. Version du 30/4/2015

En contrepartie de l'accès à ces informations en langue française, pouvez-vous m'aider dans mon travail ?

- en me signalant des erreurs ?
- en me suggérant des documents à lire ?
- en me suggérant des idées ?
- en partageant ainsi vos informations avec les autres personnes intéressées par le sujet ?

Vous pouvez écrire vos commentaires directement sur les Google Docs.

N'oubliez pas de signer si vous souhaitez être remercié dans la version finale de la thèse.

Des publications communes d'articles pourront également être proposées aux meilleurs contributeurs.

Contact : mathieuandro\_\_AROBASE\_\_yahoo.fr



# Vers une offre de services à la DIST ?

- Créer un groupe de travail (scientifiques + professionnels IST) afin d'identifier des projets possibles et de les appuyer
- Communiquer
- Libérer les imaginations

# Conclusion

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

# Merci pour votre attention

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

---

Diaporama : <http://tinyurl.com/pxnu3do>

Thèse : <http://www.bibliotheque-numerique.fr>

