



**HAL**  
open science

## Le crowdsourcing appliqué aux bibliothèques numériques

Mathieu Andro, Imad Saleh

► **To cite this version:**

Mathieu Andro, Imad Saleh. Le crowdsourcing appliqué aux bibliothèques numériques. Bulletin d'informations (Association des bibliothécaires français), 2016, 83, pp.23-25. hal-01293615

**HAL Id: hal-01293615**

**<https://hal.science/hal-01293615>**

Submitted on 16 May 2020

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

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

## **Le crowdsourcing appliqué aux bibliothèques numériques**

Mathieu Andro, chef de projets de numérisation et de text mining à la Délégation à l'Information Scientifique et Technique de l'Institut National de la Recherche Agronomique

Imad Saleh, professeur des universités en sciences de l'information et de la communication et directeur du laboratoire Paragraphe de l'Université Paris 8

Dans le monde, les projets de bibliothèques numériques ayant recours aux internautes pour taguer, indexer, commenter, corriger l'OCR ou encore financer la numérisation des documents se sont multipliés.

Les bibliothèques bénéficient d'une bonne image auprès du public et apparaissent comme dignes de confiance, au service de l'intérêt général. Elles peuvent donc plus facilement recruter des bénévoles et avoir recours au crowdsourcing, c'est à dire à l'externalisation de micro-tâches auprès d'une foule d'internautes.

### **1/ le crowdsourcing : origines et représentations**

Historiquement, le crowdsourcing trouve son origine dans les concours et appels à contributions scientifiques du 18<sup>e</sup> siècle, dans les souscriptions publiques du 19<sup>e</sup> siècle, dans l'invention du libre service, du libre accès puis du « à la demande » qui ont fait faire au consommateur une partie du travail du producteur. Il trouve son origine technologique dans le web 2.0 qui a offert, au consommateur de contenus sur le web, la possibilité d'en produire. Ce modèle économique semble aujourd'hui dépasser les frontières entre production et consommation, entre travail sérieux et public, loisir ludique et privé, et enfin, entre professionnel expert et bénévole amateur.

Il semble donc être l'héritier :

- des idéologies socialistes et collectivistes avec la remise en cause du salariat et le partage « de chacun selon ses capacités à chacun selon ses besoins » dans un esprit d'émulation.

- des idées libertaires avec la contestation de l'autorité des experts, de l'humanisme et de l'altruisme avec la volonté de réhumaniser Internet en rendant à l'humain une place centrale comme origine irremplaçable des données et comme finalité.

- des idées libérales avec la confiance dans les libertés individuelles, dans la « main invisible » du marché et la remise en cause des monopoles d'ancien régime.

Mais le crowdsourcing pourrait aussi apparaître comme une possible « ubérisation » des bibliothèques. Les professionnels seraient remplacés par des amateurs sous-payés en dehors de tout cadre juridique ou par des bénévoles travaillant gratuitement, et produisant un travail de piètre qualité. Il y aurait aussi la privatisation et l'appropriation individuelle du patrimoine collectif par le lambda et le trivial, ou enfin, une simple publicité institutionnelle autour d'un sujet à la mode avec des projets qui ne réutiliseront finalement jamais les données produites par les internautes. Enfin, la question de la qualité et du vandalisme reste un argument central des opposants à l'externalisation de la production de données auprès d'internautes amateurs.

### **2/ Typologie du crowdsourcing**

Parmi les projets de crowdsourcing en bibliothèques, on peut distinguer les grands types suivants :

Types de crowdsourcing		Définition	Exemples
crowdsourcing explicite	crowdsourcing explicite gratuit	<i>recours au travail volontaire des internautes bénévoles</i>	TROVE <sup>1</sup>
	crowdsourcing explicite rémunéré	<i>recours au travail volontaire des internautes rémunérés</i>	Amazon Mechanical Turk Marketplace <sup>2</sup>
crowdsourcing implicite		<i>recours au travail involontaire des internautes</i>	reCAPTCHA <sup>3</sup>
Gamification “human computation” “games with a purpose”		<i>recours au travail des internautes sous la forme de jeux</i>	Digitalkoot <sup>4</sup>
Crowdfunding		<i>recours aux contributions financières des internautes</i>	Numalire <sup>5</sup>

**Tableau 1, Proposition d’une taxonomie du crowdsourcing pour les bibliothèques numériques**

### 3/ Crowdsourcing : participation et motivation

La majorité des données produites dans le cadre de projets de crowdsourcing scientifique a été réalisée par une minorité bien déterminée de participants et non par des foules d’anonymes comme le rapportent les projets Wikipédia, TROVE, MarineLives, Digitalkoot, les herbonautes, Rijksmuseum... Dans ces conditions, il serait parfois plus judicieux de parler de communitysourcing ou encore de nichesourcing et de chercher à recruter des personnes bien ciblées plutôt que de s’adresser à des foules sans visage.

Un projet de crowdsourcing implique des coûts en gestion de projets, en développement de plateformes, en communication pour recruter, et en community management pour maintenir la participation. En effet, les bénévoles devront être formés, soutenus, motivés. La qualité de leurs contributions devra être contrôlée avant d’être intégrées, la participation mesurée et suivie.

<sup>1</sup> <http://trove.nla.gov.au>

<sup>2</sup> <https://www.mturk.com>

<sup>3</sup> <https://www.google.com/recaptcha>

<sup>4</sup> <http://www.digitalkoot.fi>

<sup>5</sup> <http://www.numalire.com>

La participation des internautes doit en vérité toujours avoir une contrepartie. Les motivations suivantes ont été identifiées :

Motivations intrinsèques	Individuelles	Intérêt pour le patrimoine, pour un sujet intellectuel
		Développement personnel, pour se cultiver et apprendre, satisfaire sa soif de connaissance
		Activités variées réalisées de manière autonome
		Se distraire, s'amuser, jouer
		Tester une innovation
		Résoudre un problème
		Se prouver quelque chose, un challenge, avoir une bonne estime de soi
		Agir par esprit de compétition
		Avoir l'impression de posséder du pouvoir sur les choses, d'être auteur et acteur.
		Collectives
Rencontrer des personnes		
Agir par altruisme, philanthropie et redevabilité, de manière désintéressée		
Motivations extrinsèques	Rétribution immédiate	Motivations économiques et financières
		Récompenses symboliques (classements, grades, médailles...) ou réelles (avantages, citations, visites, évènements, formations, cadeaux, abonnements, bons d'achats...)
	Rétribution	Rechercher une réciprocité

	future	Améliorer son e-reputation, son CV, trouver un emploi
--	--------	-------------------------------------------------------

**Tableau 2, Proposition d'une taxonomie des motivations des contributeurs**

#### **4/ les résultats**

D'une manière générale, les projets de crowdsourcing visent à réduire les coûts des bibliothèques, à décupler leurs ressources humaines et à rendre possible des projets inimaginables auparavant.

Cependant les projets de crowdsourcing sont loin d'être tous rentables. Les projets qui semblent le mieux fonctionner sont ceux qui reposent sur des finalités claires, ont une communication efficace, des 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 et parfois, qui ont négligé la conduite du changement en interne.

Pour ce qui est des inquiétudes concernant la qualité et le vandalisme, il existe cependant de nombreux moyens 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, comme le prétendent les tenants de la « sagesse des foules », la diversité des profils constitués par une foule d'internautes aurait pour effet de neutraliser les erreurs individuelles et de donner des meilleurs résultats que ceux obtenus avec des experts.

Enfin, au-delà de l'efficacité en termes de coûts et de résultats, il ne faut pas sous-estimer que ces projets sont aussi des atouts pour démocratiser le patrimoine sous la forme d'un « empowerment » du public et d'un changement des relations des institutions avec la société.

#### **Conclusion**

On ne peut qu'encourager les bibliothèques quelles que soient leur taille et leurs collections à se pencher sur ce type de projets qui allient transfert de connaissances et complémentarité des compétences, sociabilité et solidarité, technologie et patrimoine, citoyens et institutions.

Andro, M., Saleh, I. (2014). Bibliothèques numériques et crowdsourcing : une synthèse de la littérature académique et professionnelle internationale sur le sujet. Actes du 17e colloque international sur le document électronique (CiDE.17).

Andro, M., Saleh, I. (2015). Bibliothèques numériques et gamification : panorama et état de l'art. I2D - Information, données & documents 52(4), p. 70-79.

Andro, M., Saleh, I. (2015). La correction participative de l'OCR par crowdsourcing au profit des bibliothèques numériques. Bulletin des Bibliothèques de France.