Humanities Crowdsourcing

Laurence Favier
University of Lille 3, Department of Information and Document Sciences
GERiiCO laboratory

Abstract

Purpose/Thesis: The management literature provides a generic model of crowdsourcing for all applications. This paper is a discussion of selected features of crowdsourcing applied in science and humanities in comparison to citizen sciences. It emphasizes the relationship between the model of crowdsourcing used in humanities research and the debate on the research infrastructure.
Approach/Methods: The comparative analysis of humanities crowdsourcing models based on the review of academic literature and current crowdsourcing projects.
Results and conclusions: Crowdsourcing modeling attempts do not result in a unified approach to e-collaboration and single model of e-science. The humanities have their own issues and resources, deal with cultural diversity and are supported by appropriate methodologies. Moreover, the standardization of the description of issues studied by the humanities and modeling of humanities research practices cannot really reduce their own diversity. Further research must be conducted on the evolution of learning practices and methods of collaboration.
Originality/Value: The comparison of approaches to the phenomenon of crowdsourcing used in academic literature on management, the analysis of crowdsourcing practices in citizen sciences and crowdsourcing patterns in humanities literature.

Keywords

Humanities crowdsourcing. Citizen sciences. Research infrastructures.

Received: 18 July 2016. Reviewed: 9 August 2016. Revised: 17 October 2016. Accepted: 20 October 2016.

1. Introduction

Since Howe coined the term in 2006 many studies on crowdsourcing have been carried out. Indeed Howe defined crowdsourcing as

> the act of a(n) institution taking a function once performed by employees and outsourcing it to an undefined and generally large network of people in the form of an open call (Howe, 2006).

An *open call* is a call for participation to all connected people, without any conditions. Thus outsourcing towards the virtual world is a new perspective for production and creation as it was, in the real world, a few years ago, towards Asia:

> Remember outsourcing? Sending jobs to India and China is so 2003. The new pool of cheap labor: everyday people using their spare cycles to create content, solve problems, even do corporate R & D (Howe, 2006).

Instead of maintaining contractual and negotiated labor relations, online platforms gather the labor force of individuals and anonymous workers to perform some tasks. Internet is the infrastructure of this organization. Crowd work could concern not only unskilled labor but also expert one and intellectual creation. The main goal of this organization is to take advantage of free (or cheap) labor force availability.

Another approach presents crowdsourced labor as a large distributed system in which each worker is similar to a processor. It aims at performing human intensive activities that cannot be replaced by automatization. Thus Amazon Mechanical Turk is presented as an Internet marketplace that enables individuals and businesses to coordinate the use of human intelligence to perform tasks that computers are currently unable to do. These tasks include choosing the best shot along several photographs of a storefront, writing product descriptions or identifying performers on music CDs.

In this context, citizen sciences and especially humanities crowdsourcing (HC) can be seen as an application to mechanized labor. Citizen science is a global movement through which scientists and non-scientists alike make observations, collect data, and help answer some of our planet's most pressing questions. From GalaxyZoo designed to help classify galaxies according to their shapes in order to understand how they were formed, to the Trove project led by the National Library of Australia which lies in the correction and transcription of newspaper's archives there is an increasing number of examples that raise the question of crowdsourcing models. Indeed is there a generic model of crowdsourcing activities for any purpose? Can citizen sciences and humanities crowdsourcing be assimilated even though they do not involve the same actors in the process? Libraries role for instance is a specificity of HC.

This paper deals with the specificity of HC. It first tackles the generic models and definitions of crowdsourcing and their application to citizen sciences and then it presents HC modeling and the issues for research infrastructures design.

2. Crowdsourcing Definitions and Models

Between low cost industrial production and social intelligence creative power, crowdsourcing models attempt to unify this type of collective action. Two main approaches frame the debate on models are: management one and information science.

2.1. Management Approach

Open call is the main crowdsourcing specific feature. Crowd refers to an undefined group of people responding to a call for participation that does not result in short-listed candidates. It is equivalent to photos or to posters competitions in the real world but on a larger scale, since it involves every individual who might be connected to Internet. Crowdsourcing has, in fact, been presented as

> a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable

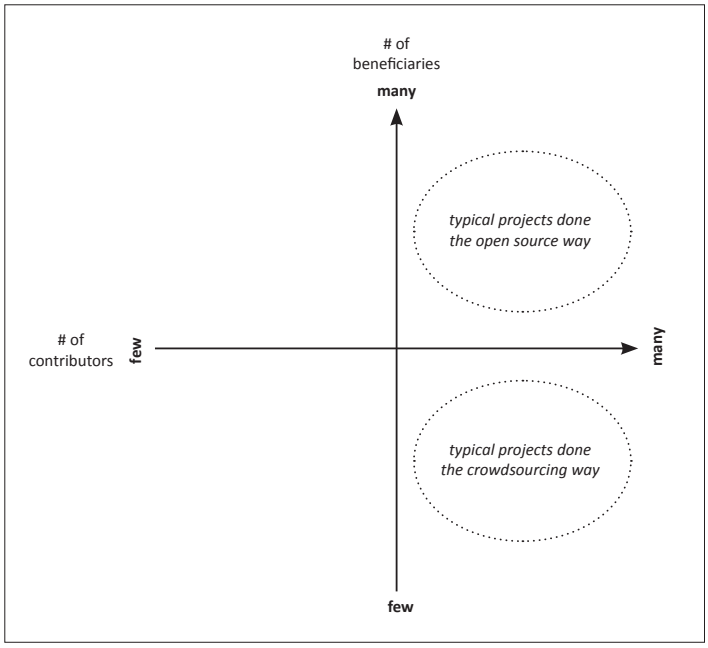
complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit (Estellés-Arolas & Gonzáles-Ladrón-de-Guevara, 2012).

Voluntary crowd contribution means that autonomous individuals converge to take part in a project without subordination link to an organization (what is known as "perverted crowdsourcing").

Thus in a management perspective, crowdsourcing is an open outsourcing that is basically different from close participation. Close participation is, for example, obtaining external service provider via a tendering process. In this case, consultants are selected for their specific expertise or a partnerships' strategy (consortium) is built in order to take advantage of respective expertise.

Moreover, for some authors, crowdsourcing is not a community-based participation model. In community-based participation, each one can contribute to define the problem as well as solve it. A typical example is open source software project. In crowdsourcing, an institution or a company publishes its needs (a problem or a task list) online and asks the crowd to respond – an approach called "top-down". Grams (2010) establishes a difference between open source and crowdsourcing (Fig.1): the first is an organization requesting many contributors for many beneficiaries (open source); the second seeks many contributors for a few beneficiaries (crowdsourcing).

Fig. 1. Crowdsourcing diagram by Grams (2010)



If crowdsourcing is always initiated by an organization (company or institution) taking advantage of the crowd work and deciding the tasks that should be performed, its goals nonetheless vary from one project to another. Burger-Helmchen and Pénin (2011) propose

a typology of crowdsourcing based on three basic forms: crowdsourcing for supporting daily activities, content crowdsourcing and inventive crowdsourcing.

In the first category, crowd brings time and process information abilities. Applications dedicated to traffic and navigation illustrate this category. Waze, for example,

> enables drivers to share real-time traffic and road info, saving other local drivers time and gas money on their commute. Like other GPS software, it learns from users' driving times to provide routing and real-time traffic updates (https://www.waze.com/).

Another example is Internet Eyes that rewards Internet users

> to watch live cameras from shops in order to warn shop owners about crimes in real time (https://www.interneteyes.com.br/ie/).

In the two cases, a large crowd is required. The motivations can be driven by micro-payment (Internet Eyes) or gamification (Waze) with a benefit for all.

In the second category, content crowdsourcing, crowd enriches with its own content a public online repository: it can be user-generated content shared through a platform (iStockPhoto) or participatory archives project (http://www.europeana1914–1918.eu/fr) launched by a library or a museum. Crowd diversity as much as its size is an important criterion for the success of the project.

The third kind of crowdsourcing is dedicated to inventive activities and lies in bringing solutions to problem solving. Open innovation platforms, like InnoCentive belong to this category. Thus

> working with InnoCentive eases the difficulty of solving complex issues and can provide actionable solutions and fresh ideas when organizations hit a roadblock in a project. Our offerings allow organizations to crowdsource solutions from invited audiences or to open their Challenges to our problem-solving network (https://www.innocentive.com/).

In this category crowd diversity is a much more important criterion than its size.

Many crowdsourcing taxonomies are available to characterize the complexity of the task (Nakatsu et al., 2014), the level of coordination in the crowd (to undertake independent or interdependent tasks), task structuration degree (if the solution to the problem is well defined or not) and motivations. In these models there are no basic differences between crowdsourcing for science, for innovation (consumer-driven innovation), for Human Intelligence Tasks (HITs) like Amazon's Mechanical Turk.

2.2. Information Science Models: Human Computation and Citizen Sciences

The purpose of the activity is not the main criteria of the crowdsourcing taxonomy. In information science approaches, design of online platforms plays a crucial role in harvesting collective intelligence. From this point of view, the main crowdsourcing genres are: Mechanized Labour (MLab) and GWAPs (Games With a Purpose). MLab allows requesters to post their micro-tasks in the form of HITs that are paid in return. Altruistic crowdsourcing is a variant that refers to projects realized by contributors only driven by intrinsic motivation because they are interested in a domain. GWAPs enable human contributors to carry out computation tasks as if they were playing online games.

When managed within a framework of scientific research, crowdsourcing becomes "citizen sciences". Citizen sciences also called "citizen cyberscience" are supposed to enable amateurs learning about science so that they take part in the most creative aspects of research, beyond helping scientists execute hard tasks. Thus, a large group of people can offer solutions to search questions and data analysis that would be unavailable to individual or a small group.

In knowledge acquisition for computing related fields (artificial intelligence, Semantic Web, machine learning, natural language processing, speech processing) Crowdflower, a MLab platform, has been used for defining semantic relations between concept pairs, useful for ontology learning or matching. "Human Computation" has been developed as a computing paradigm and human computation methods that

> can help to gather training data for intelligent algorithm, to perform tasks that are too difficult for the algorithms or to evaluate algorithms' output (Sabou et al., 2013).

Crowd then represents human processing power to solve problems that computers cannot solve.

Altruistic crowdsourcing projects are at the core of citizen sciences. Zoouniverse (https://www.zooniverse.org/projects) portal gathers many of them, including the famous Galaxy Zoo, which drives people to assist in the morphological classification of large numbers of galaxies. In the fourth and latest version of the Galaxy Zoo project, users are shown images of a galaxy and then asked a series of questions to classify its morphology. The current sample includes images of high-redshift galaxies taken by the Hubble Space Telescope and low-redshift galaxies from the Sloan Digital Sky Survey in New-Mexico. According to Wikipedia in the Zoouniverse chapter, the main users' tasks required, in all areas (astronomy, life sciences, humanities and others) are: decision tree, annotation, filtering, ranking, pattern matching (sound classification, comparison between original images and simulations) and transcription.

Citizen sciences also employ GWAPs. *Foldit* is an online puzzle video game that

> allows players to fold and design proteins, which are often implicated in human disease. In particular, Foldit players can have a huge impact on rare and neglected diseases (https://fold.it/portal/).

In tuberculosis challenge,

> *Foldit* can help through structure-based drug design (SBDD). The steps involved in SBDD are 1) identification of a target (a protein), 2) crystallization of the target, and 3) design of small-molecule drugs for the target (https://fold.it/portal/).

Many puzzles categories are suggested to meet these objectives.

Another example is *Phylo Game* (http://phylo.cs.mcgill.ca/) which helps genetic research in solving the Multiple Sequence Alignment problem, whereby a sequence alignment is a way of arranging the sequences of DNA, RNA or protein to identify regions of similarity. Traditionally, multiple sequence alignment algorithms use computationally complex heuristics to align the sequences. Given the size of the genome (consisting of three billion base pairs) it would be computationally too expensive to obtain an optimal alignment genome automatically. Transformed into a game based on recognizing patterns and solving visual problems, multiple sequence alignment can be optimized in ways that the computer algorithm can not.

Many comparisons between genres (MLab vs GWAPs) have been proposed by HC specialists (Sabou et al., 2013; Fraisse et al., 2015) related to the role of crowdsourcing for knowledge acquisition in order to build knowledge organization systems. The criteria for the evaluation of results are speed, costs and quality. Most of them showed that methods produced comparable results to those obtainable with experts or traditionally hired and trained contributors. When they compare a GWAP "Climate challenge" and a collaborative annotation site based on Crowdflower aiming at building ontology in climate change field, Fraisse et al. (2015) showed that the quality of results is similar to the one obtained by experts with reduced costs and time. But the question to be asked is whether we can extend these results to any scientific field? Zooniverse portal gives a unified vision of citizen sciences including humanities. Yet is crowdsourcing unambiguous? An answer to these questions will help to highlight the debate about research cyberinfrastructures (Favier, 2015).

3. Humanities Crowdsourcing

3.1. Specificity of Humanities Crowdsourcing

Humanties[1] crowdsourcing does not share the same properties as mechanized labor and citizen sciences. Hence task division cannot be similar because the understanding of the whole project is necessary to perform micro-tasks. Moreover, GWAPs are very rarely used in HC design. By the way, it is worth noting that HC initiatives are seldom presented in citizen sciences.

Indeed, data production and processing do not have the same meaning in sciences and in humanities. As Borgman (2009) indicated:

> The humanities and arts are the least likely of the disciplines to generate their own data in the forms of observations, models, or experiments. Humanities scholars rely most heavily on records, whether newspapers, photographs, letters, diaries, books, articles; records of birth, death, marriage; records found in churches, courts, schools, and colleges; or maps. Any record of human experience can be a data source to a humanities scholar. Many of those sources are public while others are private. Cultural records may be found in libraries, archives, museums, or government agencies, under a complex mix of access rules (...).

The consequence is that various kinds of actors are involved in humanities crowdsourcing, for instance, researchers, computing experts as well as many actors from the cultural heritage sector. Yet, the most important thing is that the "crowd" involved in such projects have intrinsic motivation. Open call finally becomes a selection of individuals who are interested and engaged in adding value to collections of records. According to Schreibman et al. (2016), they "are not about anonymous masses of people". The authors stress that in HC

> The work is not *labour* but a meaningful way in which individuals can interact with, explore, and understand the historical record. It is often highly motivated and skilled individuals that offer to help, rather than those who can be described with the derogatory term *amateurs*.

[1] The scope of the name "Humanities" used in the article is equivalent to the scope of the name "Human and social sciences" used in French official documents.

Neither "amateurs", nor "crowd", most of HC actors are not motivated by earning money or enjoying the game. What are they then doing in HC? According to Carletti et al. (2013), two main trends emerged: projects that require the "crowd" to integrate, enrich or reconfigure existing institutional resources and projects that ask the "crowd" to create or to contribute to new resources (participatory archives). Following a web survey carried out on thirty-six crowdsourcing projects, the authors classified them in four categories: contributory, collaborative, co-creative and hosted projects. In the first one, visitors were invited to provide limited and specified objects, actions or ideas to an institutionally controlled process; in the second, visitors were encouraged to serve as active partners in the creation of institutional projects that originated from and were ultimately controlled by the institution; in the third, community members worked together with institutional staff members from the beginning to define the project's goal and to generate the program or exhibit based on community interest; in the fourth, the institution turned over a portion of its facilities and/or resources to present programs developed and implemented by a public group or casual visitors (institutions share space and/or tools with groups).

HC can be initiated by research institutions as well as the GLAM (Galleries, Libraries, Archives, Museums) sector: *Transcribe Bentham* (http://blogs.ucl.ac.uk/transcribe-bentham) was launched by University College London, *Citizen Archivist* (http://www.archives.gov/citizen-archivist/) by The National Archives and Records Administration (NARA) of the United States. Regardless of initiators, the crowd's basic skills in demand are:

(a) providing document
Here are some examples about participatory archives:
– http://www.archives.gov/citizen-archivist/ (United States),
– http://www.europeana1914–1918.eu/fr (France).

(b) indexing/annotation/tagging
Two important projects illustrate such crowd's competencies:
– Library of Congress (United States), 2008: https://www.flickr.com/commons
Teamed up with Fickr, the popular social network and photo-sharing site, users are invited to apply tags (folksonomy) to more than 3 thousands photos collections of the Library of Congress. Librarians planned that Flickr's 23 million members told them something about the images they did not already know.
– Kuvatalkoot (Finland National Library) a service, launched to the public by end of 2013, for annotating newspaper articles: http://digi.kansalliskirjasto.fi/sanomalehti
The former initative, DigiTalkoot, was integrated in it. DigiTalkoot was a joint project run by the National Library of Finland and Microtask. The goal was to index the library's huge archives so that they were searchable on the Internet for easier access to the Finnish cultural heritage. DigiTalkoot was launched to the public on February 2011.

(c) transcription
Digital images of unpublished manuscripts are available through a platform. People are invited to perform a collaborative transcription, and further improve access to, and searchability of important collection of literary, historical and philosophical materials.
Examples of projects using crowdsoursing for this kind of needs are:
– Transcribe Bentham: http://blogs.ucl.ac.uk/transcribe-bentham/,

- ArchiveMOM-CA (Monasterium Collaborative) de Monasterium: http://monasterium.net/mom/home,
- Shakespeare World: https://www.shakespearesworld.org/#/,
- Trove by The National Library of Australia (2013): https://www.nla.gov.au/our-publications/staff-papers/trove-crowdsourcing-behaviour.
  Trove helps people to find and use resources relating to Australia. It's more than a search engine as it brings together content from libraries, museums, archives and other research organizations and gives them tools to explore and build. The National Library of Australia is using crowdsourcing to enable text correcting for its digitized newspaper collection.

(d) content creation

This approach can be illustrated by two examples:

- collaborative encyclopedia about museum collections – Rosalipédie project (France) enables everybody to write about the collections. The website is yet moderated (http://rosalis.bibliotheque.toulouse.fr/index.php?pages/rosalipedie#.V4uLgDWfZrw),
- Wiki on national archives, as *Your Archives* (United Kingdom *National Archives*) – a wiki, launched in 2007 and now integrated in the new catalogue, allowing registered people to contribute or update articles about the institution's archival resources (http://yourarchives.nationalarchives.gov.uk/index.php?title=Home_page).

Far from being reduced to mechanized micro-tasks, crowd's action must be understood as a part of the learning practices specific to humanities. The focus on "learning practices" and not only on the "history of sciences" is an approach centered on

> the experience of individuals and groups who were involved in production, handling of and circulation of knowledge, [if knowledge] is defined less by contents (...) than by the arrangements linking individual and social dimensions, combining hand gesture and thought process (Jacob, 2007[2]).

Attention to "learning practices" thus highlights knowledge production from an anthropological and historical point of view[3]. From this perspective, HC shows the digital conversion of traditional methods, what is presented as the "digital content life cycle" by Oomen and Aroyo (2011) who establish a correspondence between the HC type and digital content life (Fig. 2).

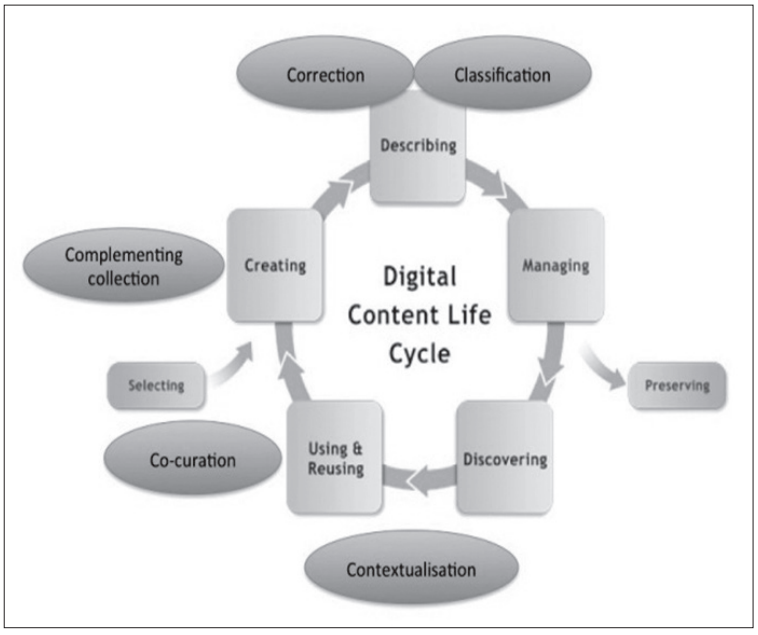
In this presentation of HC, learning practices remain unchanged: correction and transcription, contextualization, classification, co-curation. Neither the nature of tasks nor their division in a mechanized process, constitutes the element that changed humanities production of knowledge in the virtual world. Learning practices in digital humanities are still grounded in the "intellectual technologies" of writing defined by Goody (1977, 2000).

[2] Translation from Jacob (2007): l'expérience des individus et des groupes qui se sont attachés à la production, au maniement et à la circulation des savoirs – ceux-ci étant définis moins par des contenus (...) que par les modalités qui articulent l'individuel et le social, qui combinent les gestes de la main et els opérations de la pensée (*Avant-propos*, 7).

[3] In French: anthropologie des savoirs.

Fig. 2. Classification of crowdsourcing initiatives and digital content life cycle and crowdsourcing by Oomen and Aroyo (2011, 140–141)

Crowdsourcing type	Short definition
Correction and Transcription Tasks	Inviting users to correct and/or transcribe outputs of digitisation processes.
Contextualisation	Adding contextual knowledge to objects, e.g. by telling stories or writing articles/wiki pages with contextual data.
Complementing Collection	Active pursuit of additional objects to be included in a (Web) exhibit or collection.
Classification	Gathering descriptive metadata related to objects in a collection. Social tagging is a well-known example.
Co-curation	Using inspiration/expertise of non-professional curators to create (Web)exhibits.
Crowdfunding	Collective cooperation of people who pool their money and other resources together to support efforts initiated by others.



3.2. Research Infrastructures Debate

The main change enabled by digital humanities and, in particular, by crowdsourcing is the collaborative dimension of research projects. Infrastructures studies focus on this dimension in order to understand how a global science can be built in a networked environment. Infrastructure is a mix of social and technical conditions that allow large-scale collaboration and according to Bowker et al. (2010):

> Key to any new infrastructure is its ability to permit the distribution of action over space and time.

The model of such infrastructures comes from genetics or climate sciences requiring data gathering in a single archive. The result is a global database such as Worldwide *Protein Data Bank* (http://www.wwpdb.org/about/agreement) and the main challenge is to build standards and create convergence at a global level even if this objective is not easy to reach (Edwards et al., 2011). Digital humanities also require the need for infrastructures, some examples are: *Cultures of knowledge* (http://www.culturesofknowledge.org/), *ClarosNet* (http://explore.clarosnet.org/XDB/ASP/clarosHome/), *European Holocaust Research Infrastructure* (EHRI, http://www.ehri-project.eu). Thus

> EHRI provides online access to information about dispersed sources relating to the Holocaust through its Online Portal Online Portal, and tools and methods that enable researchers and archivists to collaboratively work with such sources (https://www.ehri-project.eu/about-ehri, accessed 14.01.2017)

However, the model is not the same in humanities. As van Peursen et al. (2010) stated:

> This is a fundamental difference between databases as they are used in the humanities and those that are used in the natural sciences. The way in which inscriptions are photographed or in which text corpora are transcribed and encoded, is crucial for the way in which these research objects will be studied in the future.

Adoption of standards like Text Encoding Initiative (TEI), for example, did not reduce the diversity of textual data encoding in humanities, according to each approach or each discipline. Collaboration in humanities is not based on a single data model (Favier, 2015).

The research cyberinfrastructure debate is at the core of the humanities crowdsourcing analysis: how to design digital tools for science and humanities so that a new way of discovering and thinking become available? Moreover, how could amateurs and professionals pool skills and resources to enhance research? The title of Schreibman et al. (2013) paper is an answer: *Beyond Infrastructure, Modelling Scholarly Research and Collaboration. Digital Humanities*. The authors state that:

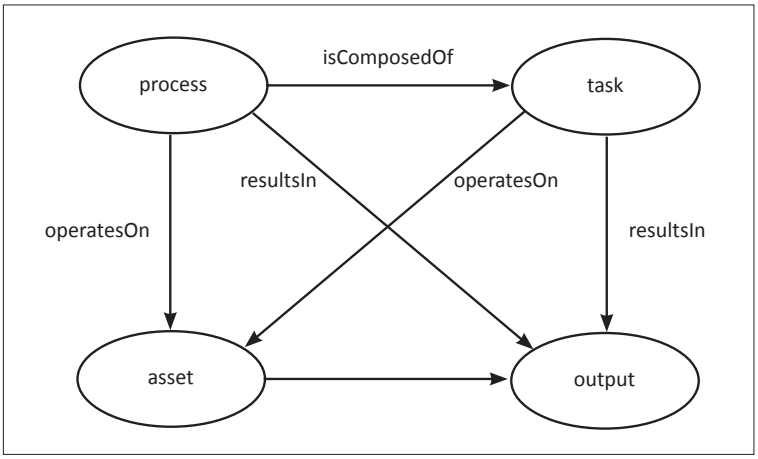
> It has become clear, however, that the focus on building infrastructure, while essential to support digital humanities scholarship, needs to be accompanied by a concomitant methodological emphasis (Schreibman et al, 2013, 2).

Because humanities cannot produce a single data repository as the result of a global collaboration, HC models are based rather on learning practices. Dunn and Hedges (2012) assert that:

> Despite the great diversity of humanities crowdsourcing, it is possible to observe patterns in which such communities thrive: these patterns are dependent on the correct combinations of asset type (the content or data forming the subject of the activity), process type (what is done with that content), task type (how it is done), and he output type (the thing produced) desired (Dunn & Hedges, 2012, 2).

Fig. 3 presents the framework of typology of humanities crowdsourcing proposed by Dunn and Hedges.

Fig. 3. Typology framework of crowdsourcing in humanities research
by Dunn & Hedges, 2013, 157



Process Type can be:
- Collaborative tagging;
- Linking Correcting/modifying content Transcribing;
- Recording and creating content;
- Commenting, critical responses and stating preferences;
- Categorising;
- Cataloguing, Contextualisation;
- Mapping;
- Georeferencing;
- Translating.

Asset Types are:
- Geospatial/ Text/ Numerical or statistical information/ Sound Image Video/ Ephemera and intangible cultural heritage.

Task Types can be:
- Task Mechanical/ Configurational/Editorial/ Synthetic/ Investigative/ Creative.

Output Types are:
- Original text/ Transcribed text/Corrected text/ Enhanced text/ Transcribed music/ Metadata/Structured data/Knowledge-awareness/ Funding Synthesis/ Composite digital collection with multiple meanings.

Dunn and Hedges representation of HC may be thought of as a framework of "primitives", in a manner that is analogous to that of "scholarly primitives". Scholarly primitives may

be defined as "basic functions common to scholarly activity across disciplines", and they provide a conceptual framework for classifying scholarly activities. John Unsworth (2000) conceptualized "scholarly primitives" as basic functions that are common to any scholarly activity in the humanities independent of discipline, theoretical orientation, or era. He suggested seven recursive and interrelated scholarly primitives – discovering, annotating, comparing, referring, sampling, illustrating, and representing – that he saw as the basis for tool-building enterprises for the Digital Humanities. Hennicke et al. (2015) suggest to extend the primitives beyond textual content and propose a more complex model (named "Scholarly Domain Model") than that which restricted to scholarly primitives. They justify the model explaining that:

> In the light of a recognizable deficit in conceptual work on the constituents of scholarship in the digital humanities and a predominance of infrastructure-oriented projects in the field, the SDM (Scholarly Domain Model) provides a framework for the systematic investigation of the relation between scholarly practices and the emergence of digital practices and methodology in continuously evolving Virtual Research Environments (VRE) (Hennicke et al., 2015).

E-science and digital humanities are often defined as "data driven" but the real focus is more on scholarly practices than on data.

4. Conclusion

This study on humanities crowdsourcing showed that the industrialization of science production through mechanized labor does not reflect reality and that digitalization does not unify scientific practices. Standardization efforts to build a common language to describe digitized objects do not represent the same challenge in science and humanities. The way of producing data and of collaborating doesn't lead to a unique e-science model: humanities have their own issues and stakes, dealing with cultural diversity and supported by appropriate methodologies. The fact that actors involved in HC projects are neither a "crowd" nor amateurs indicates new forms of public engagement in science, different from "wisdom of crowd" or mechanized labor. Thus humanities crowdsourcing emphasizes the specificity of humanities in e-science based on motivation of individuals and scholarly practices.

References

Borgman, C. (2009). A Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quaterly* [online] 3(4) [17.10.2016], http://www.digitalhumanities.org/dhq/vol/3/4/000077/000077.html

Bowker, G. C.; Baker, K. S.; Millerand, F.; Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: J. Hunsinger, L. Klastrup, J. M. Allen (eds.). *International Handbook of Internet Research*, Londres, Springer, 97–118.

Burger-Helmchen, T.; Pénin, J. (2011). Crowdsourcing: définition, enjeux, typologie. *Management & Avenir* (41), 254–269.

Carletti, L.; McAuley, D.; Price, D.; Giannachi, G.; Benford, S. (2013). Digital Humanities and Crowdsourcing: an exploration. *MW 2013: Museums and the Web. The Annual Conference of Museums and the Web* [online], April 17–20, 2013, Portland, USA [17.10.2016], http://mw2013.museumsandtheweb.com/proposals/digital-humanities-and-crowdsourcing-an-exploration/

Dunn, S.; Hedges, M. (2012). Crowd-Sourcing Scoping Study. Engaging the Crowd with Humanities Research [online]. Arts & Humanities Research Council, A project of AHRC Connected Communities Theme [17.10.2016], http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/12/Crowdsourcing-connected-communities.pdf

Dunn, S.; Hedges, M. (2013). Crowdsourcing as a Component of Humanities Research Infrastructures. *International Journal of Humanities and Arts Computing* 7(1–2), 147–169.

Edwards, P. N.; Mayernik, M. S.; Batcheller, A. L.; Bowker G. C.; Borgman, Ch. L. (2011). Science Friction: Data, Metadata, and Collaboration. *Social Studies of Science* 41(5), 667–690.

Estellés-Arolas, E.; Gonzáles-Ladrón-de-Guevara, F. (2012). Towards an Integrated Crowdsourcing Definition. *Journal of Information Science* 38(2), 1–14

Favier, L. (2015). Les Humanités numériques et l'évolution des infrastructures de recherche: quels enjeux pour l'organisation des connaissances? [online]. 1 Les Humanités numériques et l'évolution des infrastructures de recherche: quels enjeux pour l'organisation des connaissances [online]. The 10th ISKO-France conference, 5–6 novembre 2015, Systèmes d'organisation des connaissances et humanités numériques [17.10.2016], http://www.isko-france.asso.fr/isko2015/

Fraisse, A.; Paroubek, P. (2015). Vers des pratiques collaboratives pour les systèmes d'organisation de connaissances [online]. The10th ISKO-France Conference, 5–6 novembre 2015, Systèmes d'organisation des connaissances et humanités numériques [17.10.2016], http://www.isko-france.asso.fr/isko2015/

Fraisse, A.; Paroubek, P. (2014). Toward a Unifying Model for Opinion, Sentiment and Emotion Annotation and Information Extraction. In: *Proceedings of (LREC 2014) the 9th International Conference on Language Resources and Evaluation*, May 26–31, 2014. Reykjavik, Iceland, 3881–3886.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge : Cambridge University Press. *La raison graphique. La domestication de la pensée sauvage*. French translation: Minuit, coll. "Le sens commun", 1979.

Goody, J. (2000). The Power of the Written Tradition. London: Smithsonian Institution Press. French translation: *Pouvoirs et savoirs de l'écrit*. Paris: la Dispute, 2007.

Grams, C. (2010). Why the Open Source Way Trumps the Crowdsourcing Way [online]. *Opensource.com Blog* [17.10.2016], http://opensource.com/business/10/4/why-open-source-way-trumps--crowdsourcing-way

Hennicke, S.; Gradmann, S.; Dill, K.; Tschumpel, G.; Thoden, K.; Morbindoni, Ch.; Pichler, A. (2015). Research Report on DH Scholarly Primitives. Digitised Manuscripts to Europeana. Project co--funded by the European Commission within the ICT Policy Support Programme, http://dm2e.eu/files/D3.4_1.0_Research_Report_on_DH_Scholarly_Primitives_150210.pdf

Howe, J. (2006). The Rise of Crowsourcing [online]. *Wired Magazine* [6.03.2015], http://archive.wired.com/wired/archive/14.06/crowds_pr.html

Jacob, C. (ed.). (2007). *Lieux de savoir. Espaces et communautés*. Paris: Albin Michel.

Nakatsu, R.T.; Grossmann, E.B.; Iacovou, Ch.L. (2014). A Taxonomy of Crowdsourcing Based on Task Complexity. *Journal of Information Science*, 40(6), 823–834.

Oomen, J.; Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. In: *5th International Conference on Communities & Technologies* – C&T 29 June – 2 July 2011, The Edge, State Library of Queensland, Brisbane, Australia. ACM 978-1-4503-0824-3

Rouse, C. A. (2010). A Preliminary Taxonomy of Crowdsourcing. In: *ACIS 2010 Proceedings* [online], Paper 76, http://aisel.aisnet.org/acis2010/76

Sabou, M.; Bontcheva, K.; Scharl, A.; Föls, M. (2013). Games with a Purpose or Mechanised Labour? A Comparative Study. In: *I-Know'13. Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies*, Graz, Austria, Paper 19.

Schreibman, S.; Siemens, R.; Unsworth, J. (eds.). (2016). Crowdsourcing in the Digital Humanities. In: *A New Companion to Digital Humanities*. Wiley-Blackwel, 420–439.

Schreibman, S.; Gradmann, S.; Hennicke, S.; Blanke, T.; Chambers, S.; Dunning, A.; Gray, J.; Lauer, G.; Pichler, A.; Renn, J. et al. (2013). Beyond Infrastructure Modelling Scholarly Research and Collaboration. *Digital Humanities*, Jul 2013, Lincoln, United States. <hal-00801439>, https://hal.inria.fr/hal-00801439/document

Unsworth, J. (2000). Scholarly Primitives: What Methods Do Humanities Researchers Have in Common, and How Might Our Tools Reflect This? [online]. Symposium on "Humanities Computing: formal methods, experimental practice" sponsored by King's College, London, May 13, 2000 [17.10.2016], http://people.brandeis.edu/~unsworth/Kings.5–00/primitives.html

Van Peursen, W. T.; Thoutenhoofd, E.; van der Weel, A. (2010). *Text Comparison and Digital Creativity: The Production of Presence and Meaning in Digital Text Scholarship*. Leiden BRILL.

Crowdsourcing w humanistyce

Abstrakt

Cel/Teza: W literaturze z zakresu nauk o zarządzaniu wprowadzony został ogólny model crowdsourcingu dla wszelkich zastosowań. Niniejszy artykuł dotyczy specyficznych cech crowdsourcingu wykorzystywanego w nauce i humanistyce w porównaniu z nauką obywatelską. Podkreślony został w nim związek między modelowaniem crowdsourcingu wykorzystywanego w badaniach w humanistyce a debatą nad infrastrukturą badawczą.
Koncepcja/Metody badań: Analiza porównawcza modeli crowdsourcingu wykorzystywanych w humanistyce na podstawie analizy literatury akademickiej i istniejących projektów crowdsourcingowych.
Wyniki i wnioski: Próby modelowania crowdsourcingu nie prowadzą do wyodrębnienia jednego podejścia do e-współpracy i jednego modelu e-nauki. Nauki humanistyczne, które cechują specyficzne problemy i zasoby, mają do czynienia z różnorodnością kulturową i są wspierane przez właściwe im metody badawcze. Co więcej, standaryzacja opisu przedmiotów badań nauk humanistycznych oraz modelowanie ich praktyk badawczych nie zredukują ich różnorodności. potrzebne są dalsze badania w zakresie: praktyk uczenia się i sposobów współpracy.
Oryginalność/Wartość poznawcza: Porównanie podejść do zjawiska crowdsourcingu stosowanych w literaturze nauk o zarządzaniu, analiza praktyk crowdsourcingu w naukach obywatelskich i wzorów crowdsourcingu w literaturze humanistycznej.

Słowa kluczowe

Crowdsourcing w humanistyce. Nauki obywatelskie. Infrastruktura badawcza.

LAURENCE FAVIER is Professor of information and communication sciences at University of Lille3-Charles de Gaulle (France). She teaches library and information science. Her main research interests are: knowledge organization, history of documentation and e-government. In these fields she's focused on: social aspects of information organization, use and practices (crowdsourcing, collaborative information filtering, social influence, research infrastructures and digital humanities, interoperability issues to facilitate communication of information between groups and disciplines) and the European tradition of document and documentation. Selected publications: Favier, L.: Diderot et Otlet : quelques remarques sur deux contributions majeures à l'organisation des connaissances. AIDA Informazioni n°3–4 2016, (to be published); Favier, L.; Vinck, D.; Mustafa El Hadi, W. (eds.): L'interopérabilité culturelle. Communication 2016 vol. 34/1 (https://communication.revues.org/6606); Favier, L.: Les Humanités numériques et l'évolution des infrastructures de recherche : quels enjeux pour l'organisation des connaissances?; 10ème Colloque ISKO-France, Strasbourg 5–6 novembre 2015, Systèmes d'organisation des connaissances et humanités numériques; Favier, L.: The Library as Place in the Digital Age. In: B. Sosińska-Kalata et al. (eds.). Information science in the change:

conceptions, methods, research, practices. Miscellanea Informatologica Varsoviensia vol. VII. Warsaw: Wydaw. SBP 2014, 195–207; Favier, L.; Mustafa El Hadi, W.: L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir? In: Recherches ouvertes sur le numérique. Editions Hermès, Traité des sciences et techniques de l'information 2013, 23–38; Favier, L.; Mustafa El Hadi, W.: From text to image: the concept of universality in Knowledge Organization systems designed by Paul Otlet and the International Institute of Bibliography. In: Categories, Contexts and Relations in Knowledge Organization. Proceedings of the 12th International ISKO (International Society of Knowledge Society) Conference, 6–8 August 2012, Mysore, India. Würzburg: Ergon Verlag 2012, 220–229.

Contact to the Author:
e-mail: laurence.favier@univ-lille3.fr
University of Lille 3
Department of Information and Document Sciences, GERiiCO Laboratory
B.P. 60149
59 653 Villeneuve d'Ascq Cedex, France