

Crowdsourced transcription of digital herbaria in scientific communication

Lisa Chupin, PhD student in communication science at the Dicoen-IDF Lab, Le Cnam, Paris, Supervised by Prof. Manuel Zacklad Contact : lisa.chupin @ cnam.fr Twitter : @Lisa_Chupin

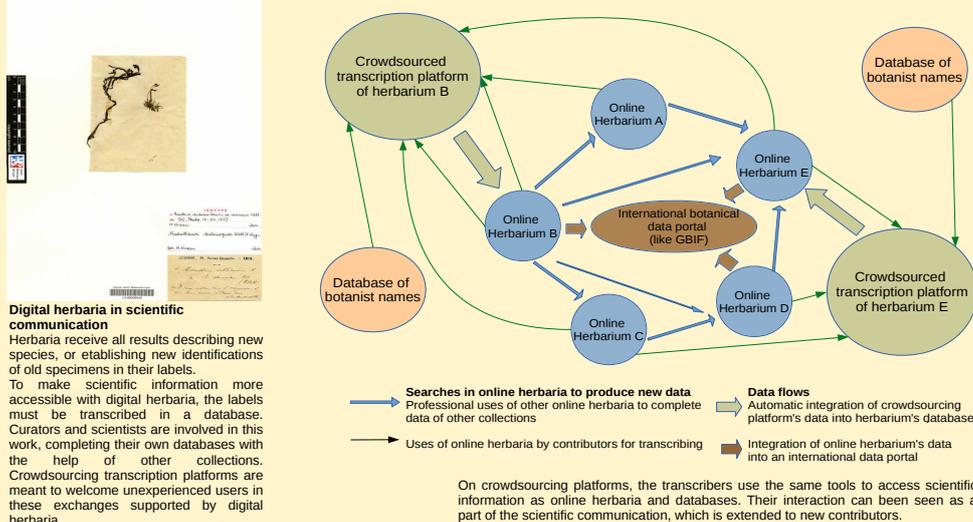
A theoretical framework in communication science about crowdsourcing for scientific heritage

« Academic crowdsourcing » (Dunn et Hedges, 2014) becomes a new way for scientific communities to create knowledge. Our approach focusses on the consequences of crowdsourcing for scientific communication, which we consider as the process of accessing the scientific information and interacting with other actors to create knowledge. Using the case of the crowdsourced transcription of herbaria, which are documents used in systematics and ecology to preserve past expertise and make it available to researchers, we aim at describing new forms of interactions between scientists, curators and neophytes about documents of digitalized collections.

Methodology

We first studied practices of scientists and curators involved in the digitalization infrastructure eReCoNat, who work with herbaria and databases of digitalized herbaria. We adopted an ethnographic method focussed on the way of using herbaria to access and exchange scientific results and information resources. Afterwards, we studied 4 transcription platforms of herbaria with the goal to picture these crowdsourcing practices in the informational ecosystem of scientists. We examined more precisely the information and communication tools and the validation process developed by the platforms to enable contributors to create data. We analysed more thoroughly the transcribing platform of the French National Museum of Natural History (MNHN), called "Les Herbonautes", for which our study aims at developing new tools. We studied the audience of "Les Herbonautes" using an online form about users' motivations and their involvement in other citizen sciences or botanical activities.

1 The integration of transcription platforms in scientific communication



Digital herbaria in scientific communication
Herbaria receive all results describing new species, or establishing new identifications of old specimens in their labels. To make scientific information more accessible with digital herbaria, the labels must be transcribed in a database. Curators and scientists are involved in this work, completing their own databases with the help of other collections. Crowdsourcing transcription platforms are meant to welcome unexperienced users in these exchanges supported by digital herbaria.

2 The transcription frame as learning tool to share interest for scientific documents

Transcription interfaces split and order the reading of labels, giving resources for each step. Auto-completion and drop-down menus are used to ease and control transcription.

In *Notes from Nature*, you first select the label and then transcribe each field of the label.



The French transcription platform *Les Herbonautes* offers quizzes and thematic missions :



Name and icon of the mission

First step of transcription open to all : the country

Following steps (in grey)

3 The role of communication tools to build communities of contributors

Involving crowds means light forms of collaboration (Haythorhwaite, 2009). Transcribers are often acting alone, and interactions are even sometimes feared as a source of error propagation (as we discovered studying *Les Herbonautes*) But heavier forms of collaboration are enhanced too : interactions are used to validate data : between moderators and contributors (as in the *Atlas of Living Australia*, or in *Herbaria@Home*), but even between contributors : in *Les Herbonautes*, when transcriptions are not concurring, the transcribers are warned with a pop-up window, and they have to find an agreement, modifying or confirming their transcription.

All these platforms enhance direct interactions between contributors in forums, which are necessary to solve questions raised by transcription tasks, as you see looking at the hashtag used in the platform *Notes From Nature* :

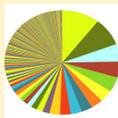
In the site *Notes from Nature*, you can comment each specimen using a hashtag :

Popular hashtags

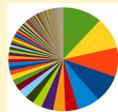
error	scientist	unclear
handwriting	date	collector
incomplete	unclear/location	cross
unclear/date	no/collector	illegible
location	error	ready/proof
multiple	no/collector	red/pine
genus	moderator	



Asking for help is not enough : organizing collected information is very needed. It is possible on the wiki of *Herbaria@home*.



The repartition of the contributions by contributor in the platform *Les Herbonautes* shows that a little group of contributors achieves most of the transcriptions.



The repartition of commentaries in forums between contributors shows that this little group of people heavy involved in transcription tasks have a lots of exchanges about transcription best practices in dedicated fora.

4 From interactions producing data to knowledge sharing

Communication tools on *Les Herbonautes* are forums related to each specimen. These forums are a place for scientists to exchange with people discovering the collections. Explaining rules of transcription are an important part of the discussions.

Word patterns analysed with Iramuteq of these forums of *Les Herbonautes* show an appropriation of technical vocabulary of herbaria. URLs cited as proofs for transcription or useful resource appear in the center of the cloud



Conclusion
The crowdsourced transcription of digital herbaria grants new publics with access to digital collections, and to their use to index new specimens. The transcription platforms allow a neophyte to discover collections as scientific information resources with enough precision to contribute to enrich them. Contributors adopt a behaviour similar to that of scientists, producing new transcriptions, after using different sources of informations to identify the exact context of the origin of the document. Discussion and communication tools on the platforms create new interactions between scientists, professional curators, and new audiences discovering collections or already interested in them. New forms of sharing knowledge appear for scientists in these platforms

Transcription platforms for herbaria
ALA (1999). *Atlas of Living Australia*, <http://www.ala.org.au/>.
Botanical Society of Britain & Ireland (2006). *Herbaria@home*, <http://herbariaunited.org/atHome/>.
Denslow, M. et al. (2013). *Notes from nature*, <http://www.notesfromnature.org>.
MNHN (2012). *Les Herbonautes*, <http://lesherbonautes.mnhn.fr/>.

Bibliography
Dunn, S. and Hedges, M. (2014), How the Crowd Can Surprise Us : Humanities Crowdsourcing and the Creation of Knowledge, in Ridge, M. (dir.) *Crowdsourcing our Cultural Heritage*, p.231-247.
Chupin, L. (2015), Enjeux de la numérisation des herbiers pour l'information et la communication scientifiques : de la transformation des matières documentaires à l'évolution des pratiques. *Enjeux de l'information et de la communication* (to be published).
Haythorhwaite, C. (2009). Crowds and Communities: Light and Heavyweight Models of Peer Production. In *Proceedings of the Hawaii International Conference On System Sciences*, ICCS, Big Island, Hawaii, janvier 2009.
Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K., Szalay, A. S. and Vandenberg, J. (2010), Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers. *Astronomy Education Review* 9 (http://aer.aas.org/resource/1/aerscz/v9/i1/p010103_s1).
Ridge, M. « From tagging to theorizing: deepening engagement with cultural heritage through crowdsourcing ». *Curator: The Museum Journal*, 56(4), 2013, p. 435-450.
Zacklad, Manuel and Chupin, Lisa (2015). Le crowdsourcing scientifique et patrimonial à la croisée de modèles de coordination et de coopération hétérogènes : le cas des herbiers numérisés. *Revue canadienne des sciences de l'information et de bibliothéconomie*.