From crowdsourcing cultural heritage to citizen science: how the Danish National Archives 25-year-old transcription project is meeting digital historians

Bárbara Revuelta-Eugercios[1,2], bre@sa.dk
Nanna Floor Clausen[1], nc@sa.dk
Katrine Tovgaard-Olsen[1], ket@sa.dk

[1] Rigsarkivet (Danish National Archives)
[2] Saxo Institute (University of Copenhagen)

Keywords: crowdsourcing, digital history, citizen science, historical records, archive

Extended abstract

The Danish National Archives have the oldest crowdsourcing project in Denmark, with more than 25 million records transcribed that illuminate the lives and deaths of Danes since the early 18th century. Until now, the main group interested in creating and using these resources has been amateur historians and genealogists. However, it has become clear that the material also holds immense value to historians, armed with the new digital methods. The rise of citizen science projects show, likewise, an alternative way, with clear research purposes, of using the crowdsourcing of cultural heritage material. How to reconcile the traditional crowd-centered approach of the existing projects, to the extent that we can talk about co-creation, with the narrowly-defined research questions and methodological decisions researchers required? How to increase the use of these materials by digital historians without losing the projects' core users?

This article articulates how the Danish National Archives (*Rigsarkivet*) are answering these questions. In the first section, we discuss the tensions and problems of combining crowdsourcing digital heritage and citizen science; in the second, the implications of the crowd-centered nature of the project in the incorporation of research interests; and in the third one, we present some strategies adopted to successfully attract digital historians to work on this material.

Crowdsourcing cultural heritage: for the public and for the humanists

In the last decades, GLAMs (galleries, libraries, archives and museums) have been embarked in digitalization projects to broaden the access, dissemination and appeal of their collections, as well as enriching them in different ways (tagging, transcribing, etc.), as part of their institutional missions. Many of these efforts have included audience or community participation, which can be loosely defined as either crowdsourcing or activities that predate or conform to its standard definition. Howe's (2006) first business-related definition describes it as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (Ridge 2014). However, the key feature that differentiates these crowdsourcing cultural heritage projects is that the work the crowd performs has never been undertaken by employees. Instead, they co-create new ways for the collections to be made available, disseminated, interpreted, enriched and enjoyed that could never had been paid for within their budgets.

These projects often feature "the crowd" at both ends of the process: volunteers contribute to improve access to and availability of the collections, which in turn will benefit the general public from which volunteers are drawn. In the process, access to the digital cultural heritage material is democratized and facilitated, transcribing records, letters, menus, tagging images, digitizing new material, etc. As a knock-on effect, the research community can also benefit, as the new materials open up possibilities for researchers in the digital humanities, which would never have achieved the transcription of millions of records within their financially limited projects.

At the same time, there has been a strand of academic applications of crowdsourcing in Humanities projects (Dunn and Hedges 2014). These initiatives fall within the so-called citizen science projects, which are driven by researchers and narrowly defined to answer a research question, so the tasks performed by the volunteers are lined up to a research purpose. Citizen science or public participation on scientific research, that emerged out of natural sciences projects in the mid-1990s (Bonney et al 2009), has branched out to meet the Humanities, building on a similar utilization of the crowd, i.e. institutional digitalization projects of cultural heritage material. In particular, archival material has been a rich source for such endeavours: weather observations from ship logs in *Old Weather* (Blaser 2014), Benthan's works in *Transcribe Bentham* (Causer & Terras 2014) or restaurant menus on *What's on the menu* (2014). While some of them have been carried out in cooperation with the GLAMs responsible for those collections, the new opportunities opened up for the digital humanities allow these projects to be carried out by researchers independently from the institutions that host the collections, missing a great opportunity to combine interests and avoid duplicating work.

Successfully bringing a given project to contribute to crowdsourcing cultural heritage material and citizen science faces many challenges. First, a collaboration needs to be established across at least two institutional settings – a GLAMs and a research institution- that have very different institutional aims, funding, culture and legal frameworks. GLAMs have foundational missions, which aim at the general public, so the research community is only a tiny percentage of its users. Any institutional research they undertake on the collections is restricted to particular areas or aspects of the collections and institutional interest which, on the other hand, is less dependent on external funding. The world of Academia, on the other hand, has a freer approach to formulating research questions but is often staffed with short-term positions and projects, time-constraints and a need of immediacy of publication and the ever-present demand for proving originality and innovation.

Additionally, when moving from cultural heritage dissemination to research applications, a wide set of issues also come into view: the boundaries between professional and lay expertise, the balance of power in the collaboration between the public, institutions and researchers, ethical concerns in relation to data quality and data property, etc. (Riesh 2014, Shirk et al 2012).

The Danish National Archives crowd-centered 25-year-old project

The Danish National Archives are dealing with the challenge of how to incorporate a more citizen-science oriented approach and attract historians (and digital humanists) to work with the existing digitized sources while maintaining its commitment to the volunteers. This challenge is of a

particular difficulty in this case because not only the interests of the archives and researchers need to align, but also those of the "crowd" itself, as volunteers have played a major role in co-creating crowdsourcing for 25 years.

The original project, now the Danish Demographic Database, DDD, (www.ddd.dda.dk), is the oldest "crowdsourcing project" in the country. It started in 1992 thanks to the interest of the genealogical communities in coordinating the transcription of historical censuses and church books. (Clausen & Jørgensen 2000). From its beginning, the volunteers were actively involved in the decision-making process of what was to be done and how, while the Danish National Archives were in charge of coordination and dissemination functions. Thus, there has been a dual government of the project and a continuous negotiation of priorities, in the form of, a coordination committee, which combines members of the public and genealogical societies as well as DNA's staff.

This tradition of co-creation has shaped the current state of the project and its relationship to research. The subsequent Crowdsourcing portal, CS, (https://cs.sa.dk/), which started in 2014 with an online interface, broadened the sources under transcription and the engagement with volunteers (in photographing, counselling, etc.), and maintains a strong philosophy of serving the volunteers' wishes and interests, rather than imposing particular lines. Crowdsourcing is seen as more than a framework for creating content: it is also a form of engagement with the collections that benefits both audiences and archive. However, it has also introduced some citizen-science projects, in which the transcriptions are intended to be used for research (e.g. the Criminality History project).

Digital history from the crowdsourced material: present and future

While *Arkivalieronline*, the collection of scanned images freely available online, is widely used among amateur historians, genealogists, historians and the public alike, the crowdsourcing projects are only widely used in genealogist and amateur historian circles. Some of the ways in which we are trying to reach professional historians and students are the following:

1. Disseminating the collections to different academic communities. On the one hand, in specialized fields as family history, demography or economic history (European Social History Conference) but also in larger digital humanities. From the beginning DDD, like other projects featuring individual-named tabulated material, participated in the Associate for History and Computing meetings. Lately, these efforts have been renewed by re-joining the field of digital humanities through participation in national events (Dighumlab) and international conferences such as DHN 2018.
2. Providing free extractions for individual users, through our own webpage but also through the North Atlantic Population Project (NAPP.org) at IPUMS, University of Minnesota, which has resulted in more than 50 articles being produced (see http://www.ddd.dda.dk/publikationer.html)
3. Participating in large-scale projects with university partners (particularly, with the University of Copenhagen) to actively pursue research with the collections. As part of the SHiP (Studies of Health in Port Cities) network, researchers at the DNA and the university collaborate in studying 19th century epidemiology patterns. The Link-Lives project is a partnership with the

university and Copenhagen City Archive to link people through the different transcribed sources available for research purposes.

4. Supporting research on the collections by providing training on how to use them. The lack of technical abilities within History faculties or student bodies in Denmark largely explains their under-utilization. The *Rigsarkivet* Digital History Methods Labs is a pilot project targeted at University students to disseminate the collections by providing a basic methodological training on how to use them for historical research. It consists on a series of small workshops where students learn the methods to address a research question using an extraction from our collections.

5. Expanding the focus of some of the traditional genealogist-driven crowdsourcing projects to incorporate a citizen-science approach. For example, the Death certificate project, initiated and run still today by volunteers, to whom DNA mostly facilitates their work (i.e., providing equipment for photographing, uploading and setting up the transcription project). However, the interest that the project has arisen among some historians and epidemiologists makes it advisable to try to bring some research considerations into play. However, this cannot be result in a researcher take-over, being paramount for the survival of the project as well as the preservation of the community to respect the citizens's ownership of the project and invite them to collaborate to make interests align.

The challenges described are not necessarily novel, as GLAM institutions (and in particularly archives) in the Nordic countries have similar collections, are involved in similar crowdsourcing projects and have similar communities that could be attracted. Thus, while some of the actions respond to specific Danish aspects (such as the reduced presence of the fields of digital history and historical demography in Denmark), most of them are already being implemented or could be implemented in other countries. Thus, the existence of partnerships university-archives in many Nordic countries around transcribed historical records (National archives of Norway and Sweden and universities of Tromsø and Umeå, for example) could very well also be used to form international inter-institutional alliances that could bring together efforts, reach a wider community and minimize investments in order to boost/liberate the potential of both the crowd and the collections.

References

Blaser, L., 2014 "*Old Weather*: approaching collections from a different angle" in Ridge (ed) *Crowdsourcing our Cultural Heritage*, Ashgate, 45-56.

Bonney et al. 2009. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. Center for Advancement of Informal Science Education (CAISE), Washington, DC

Clausen, N.C and Marker, H.J., 2000, "The Danish Data Archive" in Hall, McCall, Thorvaldsen *International historical microdata for population research*, Minnesota Population Center . Minneapolis, Minnesota, 79-92,

Causer, T. and Terras, M. 2014, "'Many hands make light work. Many hands together make merry work': *Transcribe Bentham* and crowdsourcing manuscript collections", in Ridge (ed) *Crowdsourcing our Cultural Heritage*, Ashgate, 57-88.

Dunn, S. and Hedges, M. 2014"How the crowd can surprise us: Humanities crowd-sourcing and the creation of knowledge", in Ridge (ed) *Crowdsourcing our Cultural Heritage*, Ashgate, 231-246.

Howe, J. 2006, "The rise of crowdsourcing", Wired, June.

Ridge, M. 2014, "Crowdsourcing our cultural heritage: Introduction", in Ridge (ed) *Crowdsourcing our Cultural Heritage*, Ashgate, 1-16.

Riesch, H., Potter, C., 2014. Citizen science as seen by scientists: methodological, epistemological and ethical dimensions. Public Understanding of Science 23 (1), 107–120

Shirk, J.L et al, 2012. Public participation in scientific research: a framework for deliberate design. Ecology and Society 17 (2),