

Crowdsourcing Digital Cultural Heritage

Goran Zlodi

Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
gzlodi@ffzg.hr

Tomislav Ivanjko

Department of Information and Communication Sciences,
Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, Zagreb, Croatia
tivanjko@ffzg.hr

Summary

With the turn towards the digital age, a growing number of institutions in the GLAM (Galleries, Archives, Museums and Libraries) sector started to identify a need for digitising their different collections placing them online with goals to preserve and exhibit them in the digital environment. After the initial efforts to develop policies, methodologies and best practices in transferring the collections into the online environment, researchers and practitioners have started to investigate possibilities of communicating those digitised collections with the public and seizing the opportunities that arise from digitisation. One of the approaches that cultural heritage institutions started to explore in order to involve the general public in their activities on the Web is crowdsourcing - taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined community in the form of an open call. In the heritage sector this means inviting members of the public, ("the crowd"), to tag and classify, transcribe, organize, and otherwise add value to digital cultural heritage collection content. In this paper we provide an overview of approaches in using the collective intelligence in the cultural heritage domain. Key terms, concepts and corresponding case studies are discussed, providing the framework for crowdsourcing projects within the heritage sector.

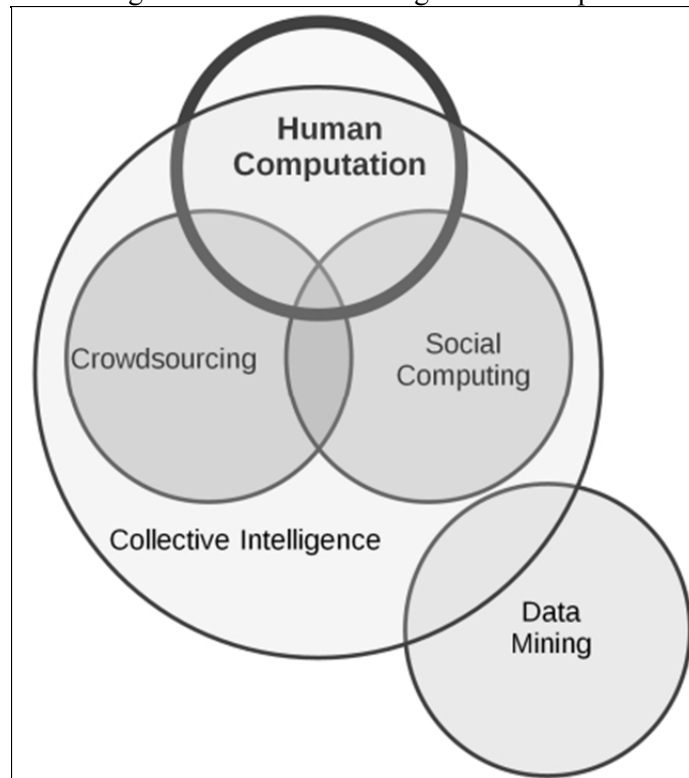
Key words: crowdsourcing, heritage, Web 2.0, archives, libraries, museums

Harnessing collective intelligence in the heritage sector – defining the field

When the notion of crowdsourcing is explored within the scientific literature related to the field, a number of related terms emerge. Crowdsourcing is often related with terms such as “collective intelligence”, “wisdom of the crowds” “human computation”, “social computing”, etc.

Figure 1 presents an attempt to classify the landscape of various systems harnessing the collective intelligence where the relations between the main concepts in the field are presented (Bederson and Quinn, 2011.)

Figure 1: Collective intelligence landscape



Authors identify three main notions that comprise the field of collective intelligence: human computation, crowdsourcing and social computing. All three notions fall in the field of collective intelligence, by having the same prerequisite for successful implementation – they all depend on a group of participants (Bederson and Quinn, 2011).

Within this framework, the most interesting is the differentiation between human computation and crowdsourcing. The modern usage of the term human computation is largely credited to the title of the thesis from of Luis von Ahn

and the related papers (von Ahn, 2005; Law and von Ahn, 2009; von Ahn and Dabbish, 2008). In his thesis (von Ahn, 2005), von Ahn defines the term as "...a paradigm for utilizing human processing power to solve problems that computers cannot yet solve." The main idea behind human computation is using the collective intelligence of users to solve problems that are hard or still impossible to do by using computer programs or algorithms. One of the most common examples of a human computation system is ReCAPTCHA (recaptcha.net), used for transcribing scanned texts for which OCR is not very effective. It takes advantage of the need for CAPTCHAs, the distorted images of text that are used by websites to prevent access by automated programs (von Ahn et al., 2008). On the other hand, term crowdsourcing is derived from the word outsourcing where a job traditionally performed by a designated agent (usually an employee) is outsourced it to an undefined, generally large group of people in the form of an open call (Howe, 2008). As Bederson and Quinn (2009) summarize, the difference between crowdsourcing and human computation is that "...Whereas human computation replaces computers with humans, crowdsourcing replaces traditional human workers with members of the public." Owens (2012) also notes that crucial difference between those two perspectives (using slightly different terminology), considering human computation and the wisdom of crowds as opposing poles of crowdsourcing activity and provides an overview of key differences between them (Table 1)

Table 1: Key differences between human computation and wisdom of crowds

	Human Computation	Wisdom of Crowds
Tools	Sophisticated	Simple
Task Nature	Highly structured	Open ended
Time Commitment	Quick & Discrete	Long & Ongoing
Social Interaction	Minimal	Extensive Community Building
Rules	Technically Implemented	Socially Negotiated

When thinking of human computation, one should imagine the example of the ESP game (link) which Owens (2012) describes as "a sophisticated little tool that prompts us to engage in a highly structured task for a very brief period of time...with almost no time commitment...practically no social interaction...and the rules of the game are strictly moderated by the technical system." On the other hand Wikipedia being the example of wisdom of the crowds: "While the tool is very simple the nature of our task is huge and open-ended...it's open-ended nature invites much more long-term commitment...an extensive community building process..." (Owens, 2012).

Following the rise of the Web 2.0, heritage institutions quickly realized the potential of the fundamental ideas underlying Web 2.0 - successful network applications are systems for harnessing collective intelligence (O'Reilly and Batelle, 2009). Many libraries, institutions or archives started using different Web 2.0 tools, such as blogs or social networks or to extend and enhance their communi-

cation with their users. According to ICOM, the museum "...acquires, conserves, researches, communicates and exhibits the tangible and intangible heritage of humanity and its environment for the purposes of education, study and enjoyment" (ICOM, 2013). That can be said for basically all institutions within the heritage sector so the idea of crowdsourcing fitted right in with the goals of researching and communicating heritage for different purposes and users. After the initial projects explored the field, some researchers found the term "crowdsourcing" isn't really appropriate for many cultural heritage projects. Both the appropriateness of the term "crowd" or the term "sourcing" is questioned, because such projects in the heritage sector mostly don't involve large and massive crowds and have very little to do with outsourcing labor. (Owens, 2013). One of the terms that can offer new perspectives is "nichesourcing", a "...specific type of crowdsourcing where complex tasks are distributed amongst a small crowd of amateur experts ... rather than the "faceless" crowd" (de Boer et al., 2012). As Owens (2012) argues, projects within the cultural heritage sector are mostly about just that: "...inviting participation from interested and engaged members of the public... [they] continue a long standing tradition of volunteerism and involvement of citizens in the creation and continued development of public goods (Owens, 2013). Following these arguments, we can see that the original Web 2.0 idea of harnessing collective intelligence found its application in the cultural heritage but denoting a specific field of application.

The initial idea offered a broad perspective with the idea that a large group of people can create a collective work whose value far exceeds that provided by any of the individual participants (O'Reilly and Batelle, 2009). This is closely related to the idea of social computing, and opened a door for different Web 2.0 applications such as Facebook, Youtube or Wikipedia where the collective power of the users was harnessed in its full strength. Along with that the field of human computation investigated how using the collective intelligence of users can solve problems that are hard or still impossible to do by using computer programs or algorithms (von Ahn, 2005). This produced many great applications such as the ESP game or reCaptcha where it was clearly shown how many users can overcome serious computing problems, such as image labelling on the Web. The third incentive came in the coining of the idea of crowdsourcing – outsourcing a job usually carried out by employees to the public (Howe, 2008). One example being the annotation of large collections of digitized materials that require a lot more human effort than the institution employees can handle. Although these goals can be accomplished by crowdsourcing, because of the need for specific knowledge on the subject matter often the "public" is not the right target if the quality of metadata gathered is vital. One approach can be found in the idea of "nichesourcing", a further development of the crowdsourcing idea focused on solving different complex knowledge-intensive tasks and providing quality results by involving amateur experts instead of the general "faceless crowd" (de Boer et al., 2012).

Taking all these perspectives in consideration we can see that the field of harnessing collective intelligence in the cultural heritage sector is not only based on the idea to *use* the public, but to engage them to contribute, collaborate and co-create (Bonney et al., 2008). It is not only about getting things done (e.g. describe or transcribe certain materials) but to communicate the collection to the users by shifting their focus from consuming digital collections to collaborating in its development. This most commonly leads to a group of amateur experts that have joined the project because they care about the cause and have intrinsic motivations to participate. Both of these aspects underline the paradox of using the term “crowdsourcing” when describing cultural heritage projects since they, in most cases, include engaged amateurs (not the “crowd”) that are intrinsically motivated and don’t consider their work as labour (thus nothing is really „outsourced”). Since the term is already in use, there is a need for refinement and distinction when using the term “crowdsourcing” in the cultural heritage domain by including all the perspectives from human computation to nichesourcing as parts of the field itself.

Setting the framework for applying crowdsourcing in the cultural heritage domain

Growing number of institutions in the GLAM (Galleries, Archives, Museums and Libraries) sector started to investigate possibilities of communicating their digitised collections with the public and seizing the opportunities that arise from digitisation by applying different crowdsourcing approaches. For that reason, it is important to classify the different types of crowdsourcing in the GLAM domain so potential new projects have an overview of opportunities and challenges in the domain. In this chapter we discuss the existing attempts to classify the field of crowdsourcing with special focus on the digital heritage sector.

One of the attempts to provide a general overview of the field from the *service perspective* is “Crowdsourcing Industry Landscape”, an infographic presenting crowdsourcing taxonomy in order to provide a framework for the industry. The infographic is constantly revised, with the latest version reflecting the third generation of the taxonomy, categorising the field in six different areas:

1. crowdfunding – financial contributions from online investors, sponsors or donors to found for-profit or non-profit initiatives or enterprises (e.g. kickstarter.com, gofundme.com)
2. crowd creativity – tapping of creative talent pools to design and develop original art, media or content (e.g. istockphoto.com, minted.c)
3. distributed knowledge – development of knowledge assets or information resources from a distributed pool of contributors (e.g. GalaxyZoo, openbuildings.com)
4. cloud labour – leveraging of a distributed virtual labour pool available on-demand to fulfil a range of tasks from simple to complex(e.g. AmazonMechanicalTurk, tagasauris.com)

5. open innovation – use of sources outside of the entity or group to generate, develop and implement ideas (e.g. challengepost.com, innocentive.com)
6. tools – applications, platforms and tools that support collaboration, communication and sharing among distributed groups of people (e.g. socialvibe.com, bigdoor.com)

When approaching these categories from a heritage perspective, the category of distributed knowledge stands out as a category where crowdsourcing digital cultural heritage fits right in. As we mentioned earlier, when defining the field, the idea is to engage motivated users to contribute, collaborate and co-create, which is closely related to the distributed knowledge category description and the projects listed there. Although these categories are mainly aimed at the industry and applications, this overview is a good place to start because it lists a large number of actual projects and ideas in the field and is very informative on the possible approaches when thinking about implementing crowdsourcing in the heritage sector. Since this categorisation is very broad there is a need to look at he attempts to classify the specific domain of crowdsourcing in the cultural heritage domain. Based on the approaches in the literature the field can be categorized through three aspects:

1. participation – categorising crowdsourcing projects based on the level of user engagement
2. activities – categorising crowdsourcing projects on the types of activities undertaken
3. crowdsourcing initiatives – categorising crowdsourcing projects based on the tangible outcomes

These aspects represent the three main questions every institution should answer when considering crowdsourcing their collection, namely who are the intended users, what activities should be implemented and what are the tangible outcomes.

In the book *The Participatory Museum* (Simone, chapter 5) author list four models based on public participation in cultural institutions: (1) *contributory* projects, where visitors are solicited to provide limited and specified objects, actions, or ideas to an institutionally controlled process; (2) *collaborative* projects, where visitors are invited to serve as active partners in the creation of institutional projects that are originated and ultimately controlled by the institution; (3) *co-creative* projects, community members work together with institutional staff members from the beginning to define the project's goals and to generate the program or exhibit based on community interests and (4) *hosted* project in which the institution turns over a portion of its facilities and/or resources to present programs developed and implemented by public groups or casual visitors.

This classification looks at the field through different model of participation, providing a framework on how an institution could engage its users. As the au-

thor puts it: “No one model is better than the others. Nor should they even be seen as progressive steps towards a model of “maximal participation”...The differences among participatory project types are highly correlated with the amount of ownership, control of process, and creative output given to institutional staff members and visitors.” As a starting point for every institution, a chart displaying the fundamental characteristics of each model is provided (Simone, 2010).

Complementing the models of participation, a decision on *crowdsourcing activities* is another important aspect of the process. Ridge (2011) looks at the types of various activities which can be applied to digitized objects. Although her work primarily deals with the context of things people can do with museums and games to improve museum collections, this categorisation can be applied to general crowdsourcing activities:

- tagging – applying unstructured labels to individual objects
- debunking – flagging content for review and/or researching and providing corrections
- linking – linking objects with other objects, objects to subject authorities, objects to related media or websites;
- categorising – applying structured labels to a group of objects, collecting sets of objects or guessing the label for or relationship between presented set of objects
- stating preferences – choosing between two objects or voting on or 'liking' content
- recording a personal story – contextualising details by providing subjective oral histories or eyewitness accounts
- creative responses – writing an interesting fake history for a known object or purpose of a mystery objects.

This overview briefly summarizes what types of crowdsourcing activities can be implemented when dealing with objects in the cultural heritage sector.

Table 2: Types of crowdsourcing initiatives

Crowdsourcing type	Sort definition
Correction and Transcription Task	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 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.

The final aspect of the decision on implementing crowdsourcing is looking at the field from the perspective of tangible outcomes, i.e. how can different crowdsourcing types contribute to the working practices and what can different initiatives offer as real outcomes. Oomen and Arroyo (2011) list the classification of crowdsourcing activities based on their outcomes, a result of a study gathering examples of crowdsourcing initiatives around the globe (Table 2).

Conclusion

There are many advantages in implementing crowdsourcing services and projects within cultural heritage institutions. Users tagging, annotating and adding contextual knowledge results in easier managing of huge collections and demonstrates obvious benefits for galleries, archives, museums and libraries. Important added value for institutions is strengthening the relations with end users and consequently getting more precise insight in user's needs. On the other side, persons involved in crowdsourcing also find value in contributing to cultural heritage research and enrichment of their cultural identity. In this new way of partnership in digital environment it is necessary to continuously improve mechanisms of collaboration to achieve desired level of trustworthiness and quality of added content. Such collaboration of heritage professionals and involved end-users will as a result bring data sets of different origins to exist in the same databases. However, in presenting of content through public web services, origin of data will have to be clearly labeled so that user created data is easily differentiated from data created by heritage professionals.

References

- Bonney, R. et al. Public participation in scientific research: defining the field and assessing its potential for informal science education. A CAISE Inquiry Group report. Washington, D.C., Center for Advancement of Informal Science Education (CAISE), 2009.
- Crowdsourcing industry landscape. 2011. <http://www.crowdsourcing.org/document/december-2011-crowdsourcing-industry-landscape-download-version/9664> (2013/09/11)
- de Boer, V. et al. Nichesourcing: harnessing the power of crowds of experts. // Knowledge Engineering and Knowledge Management. Proceedings of the 18th International Conference, EKAW 2012. Berlin Heidelberg, Springer, 2012, 16-20.
- Howe, J. Crowdsourcing: a definition. 2010. <http://crowdsourcing.typepad.com> (2013/09/11)
- Howe, J. Crowdsourcing: why the power of the crowd is driving the future of business. Crown, 2008.
- Howe, J. The rise of crowdsourcing. // Wired magazine. 14 (2006), 6; 1-4.
- Law, E.; von Ahn, L. Input-agreement: a new mechanism for collecting data using human computation games. // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09). New York: ACM, 2009. 1197-1206.
- Oomen, J.; Arroyo, L. Crowdsourcing in the cultural heritage domain: opportunities and challenges. // *Proceedings of the 5th International Conference on Communities and Technologies (C&T '11)*. New York: ACM, 2011, 138-149.
- O'Reilly, T.; Battelle, J. Web squared: Web 2.0 five years on. O'Reilly Media, 2009. http://gossgrrove.com/sites/default/files/web2009_websquared-whitepaper.pdf
- Owens, T. Digital cultural heritage and the crowd. // *Curator: The Museum Journal*. 56 (2013), 1; 121-130.

- Owens, T. Human Computation and Wisdom of Crowds in Cultural Heritage. 2012. <http://www.trevorowens.org/2012/06/human-computation-and-wisdom-of-crowds-in-cultural-heritage/> (2013/09/11)
- Quinn, A. J.; Bederson, B. B. Human computation: a survey and taxonomy of a growing field . // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11). New York: ACM, 2011, 1403-1412.
- Ridge, M. Playing with difficult objects – game designs to improve museum collections. // Museums and the Web 2011: Proceedings / Trant, J.; Bearman, D. (eds). Toronto: Archives & Museum Informatics. 2011. http://conference.archimuse.com/mw2011/papers/playing_with_difficult_objects_game_designs_improve_museum_collections
- Simon, N. The Participatory Museum. Santa Cruz, California: Museum 2.0, 2010.
- von Ahn, L. et al. ReCAPTCHA: human-based character recognition via web security measures. // Science. 321 (2008), 5895; 1465-1468.
- von Ahn, L. Games with a Purpose. // Computer. 39 (2006), 6; 92-96.
- von Ahn, L. Human Computation. Doctoral Thesis. Carnegie Mellon University, 2005.
- von Ahn, L.; Dabbish, L. Designing games with a purpose. // Communications of the ACM. 51 (2008), 8; 58-67.
- von Ahn, L.; Dabbish, L. Labeling images with a computer game. // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04). New York: ACM, 2004, 319-326.
- von Ahn, L.; Kedia, M.; Blum, M. Verbosity: a game for collecting common-sense facts. // Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06) / Grinter, R. et al. (eds.). New York: ACM, 2006., 75-78.