

**COLLABORATIVE AND
CROWDSOURCING ARCHIVES
PARALLEL ARCHIVE (PA) AS A CASE STUDY AT
THE OPEN SOCIETY ARCHIVES**

ROBERT PARNICA

Open Society Archives, Central European University, Budapest, Hungary

ABSTRACT

The central problem for archives, libraries and museums is how to bring users together with the materials they need and want to research. The purpose of this paper is to address the issue of collaboration and user participation in the archival environment. It focuses on participatory archives that may also be referred to as collaborative or crowdsourced archives. In this paper we consider crowdsourcing as an archival technique or method on the one side, but also as broader social engagement on the other. The first part of the paper elaborates on the conceptual framework being applied and provides explanations of terms and their definitions. It continues with the description of some of the most important crowdsourcing projects today. The second part reflects on the collaborative and crowdsourcing praxis applied in the scholarly and collaborative project, Parallel Archive (PA), initiated by the Open Society Archives (OSA). In concluding, the author reflects on the users of PA and how to address problems of collaboration, and offers some thoughts on current trends in OSA.

KEYWORDS

collaborative archive, crowdsourcing, Parallel Archive, participatory archives, Open Society Archives (OSA), user studies

The purpose

The purpose of this paper is to address the issue of collaboration with and participation by users in the archival environment. It focuses on participatory archives that may also be referred to by terms such as collaborative or crowdsourced archives. In this paper we consider crowdsourcing primarily as an archival technique or method on the one side, but also as broader social engagement on the other.

The first part of the paper elaborates on the conceptual framework being applied and provides explanations of terms and their definitions. It continues with the description of some of the most important crowdsourcing projects today. The second part reflects on the collaborative and crowdsourcing praxis applied in one particular scholarly and collaborative project, a case study initiated by the Open Society Archives (OSA) called Parallel Archive (PA). The article concludes with reflections on the Parallel Archive's users by addressing problems of collaboration and offering some thoughts on current trends in OSA.

Introduction

The explosion of information technology and computers in the last decades of the 20th century, together with the increased expectations of researchers for efficient and immediate information/document retrieval, became a big challenge and at the same time a great opportunity for archives, archivists and their users. Before visiting archives and their physical research rooms to research primary sources, users nowadays strive to be familiar with the structure and taxonomy of a repository's archival fonds and collections together with the new granularities that are surfaced by digital catalogues or curated digital collections by visiting archives portals, consulting their online catalogues, researching finding aids, or using advanced search engines. Archives have become more inventive and innovative in publicizing their archival collections not only to preserve archival records for the future, but also to support open access by their users. Maximising the physical and intellectual accessibility of the archival materials and the information they contain is one of the most important principles and tasks of modern archives. While the traditional notion of archives has been changing, the end user entering an archive has also changed. However the concept of "user" has also evolved from being a specialist from in one single field (mostly historians and academics) to a wider public, (lawyers, journalists, economists, artists, students, the retired, enthusiasts, etc.) because everybody can be an end user.

Concurrent to the physical archives there is a parallel digital archive that is designed and constructed for the researchers in the virtual environment of web 2.0. Today, the internet offers a powerful tool to reach users in a way and mode that was hardly imaginable just a decade ago. Information technology has strongly contributed to the

opening and “democratization” of cultural heritage institutions, particularly archives, and resulted in more transparent and liberal procedures for their users. In the “Age of access” when the market economy has been competing with the “economy of networks” and where commodities and goods have been made available through online services, archival institutions have also been shifting their primary goal of preserving and granting access from analog (physical) data to providing access to digital records and collective memory.¹

Development and implementation of online finding aids and associated search engines are not enough by themselves to help stimulate archival research because of important technical, content and contextual obstacles. Recently, more and more archives are going beyond developing online catalogues or search engines and are involving themselves in digitizing entire archival fonds and collections. That process neither reflects nor supports the individual researcher’s particular interest and need for specific collections, or expectations of a digital system: instead it reflects the physical state or condition of those collections. It is neither a duty nor the goal of an archive to organize the taxonomy of its fonds, sub-fonds and collections to suit each and every user. However, archives have been inventing and implementing new ways of delivering and sharing knowledge about their collections. Other cultural heritage organizations (museums, libraries, etc.) face a similar situation in developing digital tools to manage and provide access to their paper and digital collections. This creates a hybrid environment for examining not only organizational practice but also to develop new models for providing broad access to digital objects/data through collaboration.² The fragmentation and individualization of organizational memory in most records-creating institutions has also caused blurring of borders between records and non-record materials and personal and institutional memory.³

- 1 Rifkin, Jeremy. *The age of access: the new culture of hypercapitalism, where all of life is a paid-for experience*. New York: Tarcher/Putnam, 2000.
- 2 Cruikshank, Kate; Caroline Daniels, Denis Meissner, Naomi, L. Nelson, Mark Shelstad. How do we show you what we got?: access to archival collections in Digital Age. // *Journal of the Association for History and Computing* 8, 2(September 2005), 1 [cited: 2013-04-11]. Available at: <http://quod.lib.umich.edu/j/jahc/3310410.0008.203?rgn=main;view=fulltext>
- 3 Ketelaar, Eric. *The Archives as a Time Machine, the ICT-industry and public sector partnership: to promote the preservation and accessibility of the European archival heritage*. Closing speech of the DLM-Forum 2002, Barcelona May 8, 2002 [cited: 2013-04-13]. Available at: <http://www.mybestdocs.com/ketelaar-e-dlm2002.htm>

The central problem for an archive as for a library is how to bring users together with the materials they need and want to research. Archives, museums and libraries are under pressure to make their materials understandable and accessible to many users who they might never see nor expect, because user habits and praxis have changed dramatically. A few years ago archives were counting physical visits to the research room, nowadays we count hits on the institution's web site as equally important information.⁴

Cultural and anthropological research and the concepts thereby introduced have also helped minimize the role of the archive as a "temple" where archival sources were guarded and preserved only for the privileged by the help of archival rules, restrictions, procedures and protocols.⁵ Such professionally constructed obstacles that prevented researchers from physically accessing the materials are losing their cultural and practical importance and thus also archives are losing their aura as institutions where knowledge power was practiced.

There is a trend to make archival collections more easily and widely accessible for research by using users' feedback on the usability and functionality of web sites, online finding aids, search engines, or the site's interface, and even actively involving users in participating in interface design or the processing (tagging, correcting, describing, etc.) of archival collections.⁶ We have definitively entered into the age of mass participation in archives, because not only sophisticated researchers but also enthusiastic volunteers with cognitive surplus and extra time can be devoted collaborators on various archival projects.

On terms and definitions

Participatory archives

"Friends, followers, taggers, fans, writers, editors, commenters, volunteers, collectors, scanners, sharers, transcribers, researchers, historians, students, users, collaborators, partners, re-users, re-mixers, masher-up-

4 Ibid.

5 Ketelaar, Eric. Archival temples, archival prisons: modes of power and protection. // *Archival science* 2(2002), 221-238 [cited: 2013-04-06]. Available at: <http://home.hccnet.nl/e.ketelaar/ArchivalTemples.pdf>

6 Altman, Burt; John B. Nemmers. The usability of on-line archival resources: the Polaris Project Finding Aid. // *Special collections and archives* 11 (2001) [cited: 2013-03-28]. Available at: <http://diginole.lib.fsu.edu/cgi/viewcontent.cgi?article=1010&context=sca>

pers, citizen archivists, enthusiasts, passionate amateurs, crowdsourcers, nerdsourcers—all are welcome in the participatory archives.”⁷

These opening sentences from the article of Kate Theimer from her 2010 article published in *ArchivesNext*, illustrate the variety of engagements and activities that can be invoked by the term “participatory archive.” The question is how to define a term which can be associated with so many activities and techniques, and which presupposes the massive involvement of end users or well-motivated nerdsourcers.

Despite the fact that it is difficult to settle on only one satisfactory definition of participatory archives, Isto Huvila suggests a new concept in which the archives could be radically integrated with its actual and future users. “In the post-modern sense, the notion of participation is built into any human interaction with information, which makes it and its implications also essential in the archival and records management contexts.”⁸ The basic characteristic of his approach centres around three themes: decentralized curation, radical user orientation and the contextualization of records and the entire archival process. The concept of decentralized curation, according to Huvila, refers to knowledgeable users of archival collections actively contributing to new and improved descriptions, translations and summaries of records. Following the principle of post-modern archival science, such radical user orientation refers to the functionality of an archival system to make the contents of an archive available for users. The third notion of contextualization refers to the attempt to capture much wider context than has been the case in traditional archival description.⁹

Critics of this approach complain that it is too radical for contemporary archival institutions.¹⁰ Kate Theimer suggested a new definition which has been seen as more neutral than Huvila’s and more applicable and acceptable for contemporary archives. She proposed that a participatory archive is “an organization, site or collection in which people other than archives professionals contribute knowledge or resources, resulting in increased understanding about archival ma-

7 Theimer, Kate. Building participatory archives. // *ArchivesNext* (August 22, 2010) [cited: 2013-03-22]. Available at: <http://www.archivesnext.com/?p=1536>

8 Huvila, Isto. Participatory archive: towards decentralized curation, radical user orientation and broader contextualization of records management. // *International journal on recorded information* 8, 1(2008), 15-36 [cited: 2013-04-09]. Available at: <http://link.springer.com/article/10.1007/s10502-008-9071-0#page-1>

9 Ibid.

10 Huvila, Isto. What is a participatory archive?: for real (?). // Isto Huvila’s blog (08/31/2011) [cited: 2013-04-08]. Available at: <http://istohuvila.eu/what-participatory-archive-real>

terials, usually in an online environment”¹¹ For this reason the introductory sentences quoting Theimer at the beginning of this section demonstrate a variety of activities and professions that might engage with participatory archives.

Today however, the participatory archive evokes different interpretations and meanings because the term is closely associated with broader concepts such as “citizen archives” and “citizen science” as well as efforts to include more voices and experiences through community archiving, and activism regarding social justice and human rights agendas.¹² In fact, authors today use these concepts in their works in a range of ways.¹³ What the different understandings have in common is that they exploit the fact that the internet has become very dynamic, that participation and collaboration has grown considerably, and that many major websites have been involving the broader public in the creation of their content. This is similarly applicable in archives and in other cultural heritage institutions. Even though a participatory archive is about broader collaboration and crowdsourcing, as Huvila states, it also “... focuses on deeper involvement and more complex semantics rather than on larger crowds and simple annotations.”¹⁴

Crowdsourcing

Crowdsourcing is a relatively new concept that emerged in the business world, rapidly spread into academia and the nonprofit sector, and is still undergoing constant evolution. The name derives from the word “crowd” relating to the people who participate in the project and “sourcing” relating to number of practices aimed at finding, evaluating and engaging suppliers of goods and services.¹⁵ Although we are not dealing with the economic aspects of crowdsourcing in this work, it is not always easy to draw a clear line between crowdsourcing

11 Theimer, Kate. Exploring the participatory archives. // ArchivesNext (August 30, 2011) [cited: 2013-04-09]. Available at: <http://www.archivesnext.com/?p=2319>

12 Gilliland, Anne and Sue McKemmish. The role of participatory archives in furthering human rights, reconciliation and recovery. // *Atlanti: review for modern archival theory and practice* 24(2014), 79-88.

13 Theimer, K. Building participatory archives. Op. cit.

14 Huvila, I. Participatory archive. Op. cit., p. 27.

15 Estellés-Arolas, Enrique; Fernando González-Ladrón-de-Guevara. Towards an integrated crowdsourcing definition. // *Journal of information science* 38, 2(2012), 189-200 [cited: 2013-02-14]. Available at: <http://jis.sagepub.com/content/38/2/189>

as successful alternative business model from crowdsourcing as social engagement as implemented in cultural heritage institutions.¹⁶

A review of recent works on crowdsourcing suggests a variety of definitions that derive from analysis of various examples (and their experiences), some of which are even contradicting each other. In some cases crowdsourcing is considered to be the ultimate open innovation while in others this role is contested and even questioned. In one case crowdsourcing is considered to be a distributed problem-solving model and not an open source practice at all, because the problems solved and products that are designed by the crowd become the property of profit-making firms and companies.¹⁷

The term goes back to 2006 when it was coined by Jeff Howe in his article entitled “The rise of Crowdsourcing” published in *Wired Magazine*. Howe defined crowdsourcing as “the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in a form of an open call”.¹⁸ In his “white paper definition”, he defines the term 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 the open call”. For Howe crowdsourcing and user innovations are simple tools through which open innovation can happen. Thus, crowdsourcing can result in innovation both in the development of new products as well as new practice, although some types of crowdsourcing are simply confusing because not all crowdsourcing leads to the creation of a product or innovation per se.

While trying to explain the confusion that occurs around the definition, Howe emphasized that crowdsourcing is not a single strategy, but rather an umbrella term for a very different group of approaches that share one obvious attribute – they all depend on some contribution from the crowd. We see that the adaptability of crowdsourcing

- 16 Vukovic, Maja; Claudio Bartolini. Towards a research agenda for enterprise crowdsourcing. // *Leveraging applications of formal methods, verification, and validation / 4th International Symposium on Leveraging Applications, ISoLA 2010, Heraklion, Crete, Greece, October 18-21, 2010, Proceedings, Part I*; editors T. Margaria, B. Steffen. Pp. 425-434 [cited: 2013-03-14]. Available at: http://link.springer.com/chapter/10.1007%2F978-3-642-16558-0_36#page-1
- 17 Brabham, Daren C. Crowdsourcing as a model for problem solving: an introduction and cases. // *Convergence: the international journal of research into new media technologies* 14, 1(2008), 75-90, 76 [cited: 2013-03-17]. Available at: http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf
- 18 Howe, Jeff. The rise of crowdsourcing. // *Wired magazine* 14.06(June 2006) [cited: 2013-02-14]. Available at: <http://www.wired.com/wired/archive/14.06/crowds.html>

makes it an accurate and successful practice but at the same time that adaptability makes it difficult to define or to create a feasible taxonomy because it is based on analysis of too many different examples. Today, there are varieties of definitions because different authors are looking at crowdsourcing from different angles where the kinds of problem resolution or innovation applied in the business sphere do not always correspond with the demands and requirements in cultural heritage institutions. The existence of such a range of definitions only illustrates that crowdsourcing cannot be coherently classified despite the characterization of crowdsourcing practices.¹⁹

In order to consolidate and harmonize various definitions on crowdsourcing, recent attempts have been made to analyze all existing definitions and suggest a synthetic or integrated definition which encompasses many various approaches and practices. The recent integrated definition suggested by Estellés-Arolas and González-Ladrón-de-Guevara consists of 8 elements or parameters, (1) clearly defined crowd, (2) task with clear goal, (3) reward received by the crowd is clear, (4) clearly identified crowdsourcer, (5) compensation received by the crowdsourcer is also clear, (6) it is an online task of participative type, (7) it uses an open call, and (8) it uses the internet.²⁰ Thus, the suggested integrated definition of crowdsourcing presupposes that: “Crowdsourcing is 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. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage that what the user has brought to the venture, whose form will depend on the type of activity undertaken.”²¹

Table 1 illustrates how some prominent crowdsourcing projects were assessed by the author by applying this integrated definition and

19 Schenk, Eric; Claude Guittard. Towards a characterization of crowdsourcing practices. // *Journal of innovation economics* 1, 7(2011), 93-107 [cited: 2013-03-14]. Available at: <http://www.cairn.info/revue-journal-of-innovation-economics-2011-1-page-93.htm>

20 Estellés-Arolas, E. [et al.] Op. cit., p. 197.

21 Ibid., pp. 196-197.

its parameters. As can be seen, only 2 out of 7 parameters have been represented in all 7 crowdsourcing projects – a clearly defined target group and use of the internet.

Crowdsourcing projects	1	2	3	4	5	6	7	8
Wikipedia	+	+	+	-	-	+	-	+
You Tube	+	-	-	-	-	-	-	+
Flicker	+	-	-	+	-	-	-	+
Star War Uncut (2011)	+	+	-	+	+	+	+	+
Australian Newspapers	+	+	+	+	+	+	+	+
What's on the Menu	+	+	-	+	+	+	+	+
OSA Parallel Archives	+	+	-	+	-	+	-	+

TABLE 1.

Crowdsourcing projects and integrated definition

Explanation: 1. clearly defined crowd, 2. task with clear goal, 3. award received by the crowd is clear, 4. clearly identified crowdsourcer, 5. compensation received by the crowdsourcer is also clear, 6. it is online task of participative type, 7. it uses an open call, and 8. it uses the internet.²²

Some of the crowdsourcing projects fail to meet a few of the parameters from the integrated definition. This does not necessarily mean that they fall outside the general notion of crowdsourcing because the suggested definition of crowdsourcing is based on a range of practices and results. Crowdsourcing projects that have been designed to produce certain value (where crowdsourcers obtained a compensation for their tasks) can be amply distinguished from those where the final result is knowledge. Because of this, crowdsourcing could be explained through the theory of “crowd wisdom”, an example of a collective intelligence model that is capable of aggregating scattered knowledge while reducing the time and money needed to solve the problem.²³

The emergence of the web did not invent crowdsourcing, but simply made it easier. Crowd participation was recorded earlier as a tool but not known in the form of this popular neologism. The first two examples of crowdsourcing listed below illustrate that they were perceived as primarily problem-solving initiatives motivated by financial gain, and only later did they become social engagement and considered in terms of the “wisdom of crowd”.

22 Ibid., p. 197.

23 Brabham, C. D. Op. cit., p. 79.

Canned Food – In the early 19th century, Napoleon’s army needed a large quantity of food especially in the winter period. The French government offered 12, 000 franc prize to anyone who could suggest a cheap and effective method of preserving large amounts of food. Peter Durand invented a technique which transferred the technology discovered for preserving food in glass jars into iron canisters.²⁴

Margarine – In 1869 Emperor Napoleon III had a slightly different problem from his famous predecessor with food because France could not meet the demands for butter which became very expensive and not affordable for the French army and the lower strata of society. The emperor offered a prize to anyone who could make a satisfactory substitute for butter, suitable for the army but also the poor. A French chemist Hippolyte Mege-Mouries produced a substance called oleomargarine from which the shortened name was “margarine”.²⁵

Oxford English Dictionary (OED) - Most of the English dictionaries prior to the OED were inconsistent and without historical context. OED was able to change that because it used hundreds of volunteers who were assigned to particular books, copying passages describing word usage onto citation slips. Despite errors, the project proved that engagement of the general public could be used to make tasks faster and cheaper.²⁶

The Sydney Opera House – The initiative for the opera house began in the 1940s but it was only in 1955 that the design competition was launched. The open call for a design solution was made public only after certain parameters concerning space, number of seats, etc., were announced. The competition committee received altogether 233 proposals from 23 countries. Two years later, it was announced that the Danish architect Jorn Utyon had won the first prize and thus, his building today defines the image of Sydney.²⁷

Many other similar projects over the past 200 years that used crowdsourcing as a problem-solving technique could be mentioned. Wikipedia is another example of the collaborative and knowledge sharing environment where the specific knowledge and thus, specific individual “crowd” is needed and favored. Let us mention a several

24 Canning [cited: 2013-02-02]. Available at: <http://en.wikipedia.org/wiki/Canning>

25 Margarine [cited: 2013-02-02]. Available at: <http://en.wikipedia.org/wiki/Margarine>

26 Thomas, Stuart. 9 examples of crowdsourcing, before ‘crowdsourcing’ existed. // Social networking (09.15.11) [cited: 2012-11-21]. Available at: <http://memeburn.com/2011/09/9-examples-of-crowdsourcing-before-%E2%80%98crowdsourcing%E2%80%99-existed/>

27 Ibid.

projects based on crowdsourcing that today offer important ideas and thoughts for cultural heritage institutions and archives in particular. Some of them are very innovative and unique initiatives while others solicit assistance in describing or otherwise exposing ephemeral collections when there are not enough human resources and money (Table 2).²⁸

Crowdsourcing	Description	Ideas for Archives
Star War Uncut, 2011	Star Wars Uncut is a crazy fan mashup remake of the original Star Wars movies. It is the brainchild of Casey Paugh a developer dedicated to creating interactive experiences on the web. In 2009, he was inspired to use the Internet and an ever-ready pool of passionate Star Wars fans to crowdsource the classic film <i>Star Wars IV: A New Hope</i> . This project turned into a labor of love and creativity on a large scale. Nearly a thousand fans came together to participate and the resulting movie is equal parts fun, and dearly nostalgic. ²⁹	Radio, TV, fans and consumers are subject experts. They possess in depth knowledge which is valuable for archives when we launch a discovery of audio or film items. Problem: what to do when dealing with the historical collections that do not have fan base motivation or, film items are not iconic.
What's on the Menu (NYPL), 2011	Since 1840 until the present The New York Public Library has collected more than 40 000 menus in their Restaurant Menu Collection. The menus were not searchable for specific information about particular dishes or their prices. NYPL invited the public to transcribe 9000 menus photographed for several years before for inclusion in Library's digital gallery. As of February 2012 there have been 758748 dishes from 12167 menus transcribed. ³⁰	When digitized, ephemera is digitized or scanned as an image file and the text is not indexed. Crowdsourcing is useful for transcribing, describing, annotating archival ephemeras such as tickets, pamphlets, posters, invitations, theatre programs or greeting cards. Tasks for which archivist do not have time but interested public certainly have.

28 Holley, Rose. Crowdsourcing: more cool sites to give libraries, archives and museums inspiration. // Rose Holley's Blog – views and news on digital libraries and archives (11 February 2012) [cited 2012-11-21]. Available at: <http://rose-holley.blogspot.hu/2012/02/crowdsourcing-more-cool-sites-to-give.html>

29 Star Wars Uncut [cited: 2013-04-30]. Available at: <http://www.starwarsuncut.com/about>

30 What's on the menu? [cited: 2013-04-30]. Available at: <http://menus.nypl.org>

Ancient Lives: Decoding Papyri, 2011	In 1897 two researchers from Oxford University started their archeological campaign in the Egyptian village Oxyrhynchus at the site which later turned out to be ancient garbage. Papyri were discovered containing fragments from Gospel of Thomas and St. Matthew, followed by 500 000 fragments of Greek lyric poetry, private letters, accounts, wills, marriage certificates etc., deposited in 700 boxes and brought into England. By the help of volunteers, in a project that started in 2011, the papyri have been relatively easily decoded. Before the emergence of the internet era, the task of decoding was reserved for scholars familiar with the code and for this reason, so little papyri have been decoded up until the recent call for the public to join the project. ³¹	A very difficult task has been simplified while retaining the challenge that is found in crosswords or code-breaking. The role of forums was positive since it fostered the volunteer community and made them aware how their work helps new discoveries and knowledge grow and develop.
Historypin 2011	Historypin, established in 2011, allows people to upload historic and contemporary photos, audio-visual recordings to a geolocation of their origin on a Google map which is the main technology partner. It is a collaborative form of crowdsourcing where individuals and organizations upload their digital content. However, it is also valuable educational site and provides services which libraries and archives can join to show their historical collections to new audiences. ³²	The crowd that Historypin wanted to attract is shifting toward library and archival communities which have massive quantities of digitized content in image, sound and video form. Their descriptions are more detailed, more exposed and more widely used. Beneficial for archives and libraries in that they do not need to support the services themselves while there are also no geographical boundaries. Thus, there is no reason why archives should not participate in this or similar collaborative projects that deliver social benefits and contribute to digital inclusion.

TABLE 2.

Crowdsourcing: recent trends and suggestions for cultural heritage institutions

31 Ancient lives: decoding Papyri [cited: 2013-04-30]. Available at: <http://ancientlives.org>32 Historypin [cited: 2013-04-30]. Available at: <http://ancientlives.org/story>

If we try to find the answer as to why crowdsourcing is good for archives, one must definitively mention its two substantial roles, one social and cultural, and the other practical and functional. On the social and cultural level, it stimulates open access to archival fonds but also promotes outreach. It also helps stimulate a positive sense of public ownership and responsibility toward common cultural heritage in their local and national surroundings. Crowdsourcing appears to be an important factor in helping to build trust and loyalty toward the domestic, i.e., the local and wider (national) community. Crowdsourcing also stimulates active interaction within the community in a way to support and boost users' spirit and enthusiasm especially *via* blogs and forums. Finally, crowdsourcing demonstrates the important social and cultural relevance of heritage for the community and, because of this, it defines crowdsourcing also as an important form of social engagement in which the value and importance of archives and their content is constantly increasing.³³

On the practical and functional level, crowdsourcing performs tasks for which archives have neither human resources nor money and time. Crowdsourcing utilizes expertise, skills and practices adopted from the wider community. In terms of data harvesting, crowdsourcing improves the quality of data and resources but also makes them more researchable and discoverable. It also adds new value to the archival items and collections either by adding new information or metadata or by their digital inclusion. Let us see what the crowdsourcers in archives can do. They can rate the reliability of the information or record, they can add new information about the record, or add user-created content. However, crowdsourcers can also make full text corrections, transcribe handwritten records, verify names in authority records, or describe not yet catalogued items.

At the same time as crowdsourcing is offering great opportunities to enhance the value of our collections, it can also present important problems which deserve to be mentioned. One is that of quality control. If not clearly defined from the beginning of the project, damage can be done to the success of the project if there is no quality control of the crowdsourced task and/or its product. Also, throughout the crowdsourcing process there are not many possibilities to control or influence

33 Holley, Rose. Crowdsourcing strategies for archives. Power Point Presentation, 8-12 November, 2010 [cited: 2012-12-02]. Available at: <http://www.slideshare.net/RHmarvelous/naa-archives-20-week-roseholleycrowdsourcingnov-2010>

the course of its development. Appropriate governance in the form of policies, procedures and oversight must thus be planned well ahead of its beginning. It is very important to have a clear goal for what the archives wants to achieve with the project because blurred goals and imprecisely defined results can bring project failure. Concurrent with this, it is also important to have clear decisions on which phases of the project are more important and thus have priority. One should not forget incentive, a very important motivating factor to encourage crowd-sourcers to start, continue and finish with the project's tasks. Finally, let us mention a burning issue nowadays which cannot be neglected either, which is the problem of intellectual property. Just as in the physical world, in the digital world contestations over intellectual property continue with great intensity.³⁴ All these should be resolved at the outset so that the final result of crowdsourcing can be a success.

In summary, crowdsourcing offers great opportunities for archives and other cultural heritage institutions but it can also provide us archives with substantial new challenges. In closing this section let us briefly review what the major tasks are that will lead us toward good results: (1) clear and reachable goal – communicated from the institution to the crowdsources via web 2.0, (2) quality check – regular reports on achieved tasks, (3) sustainability – project must be reliable and fun, (4) results that are clearly visible, (5) incentive – remarks and acknowledgements are communicated, (6) interesting task – exciting content for crowdsourcing, (7) volunteers to be visible – access on blogs and forums, (8) options for selection – to choose among multiple tasks, and (9) listen to your top volunteers or crowdsourcers – because they can provide valuable feedback.³⁵

Parallel Archive as a case study

Before elaborating on Parallel Archive (PA) in the context of user participation and crowdsourcing we should explain that the Open Society Archives (OSA) is not a national archive and so does not have strong institutional bonds with a country of origin. Rather it is an independ-

34 Trompette, Pascale; Valerie Chanal, Cedric Pellissier. Crowdsourcing as a way to access external knowledge for innovation: control, incentive and coordination in hybrid forms of innovation. // 24th EGOS Colloquium, Jul 2008, Amsterdam, France [cited: 2015-02-25]. Available at: <https://hal.archives-ouvertes.fr/halshs-00367373/document>

35 Compare to: Holley, Rose. Crowdsourcing: how and why should libraries do it? // D-Lib magazine 16, 3/4(March/April 2010) [cited: 2013-03-22]. Available at: <http://www.dlib.org/dlib/march10/holley/03holley.html>

ent academic institution with an international background. The nature of the collections it acquired strongly contributed to the notion of it as a non-Hungarian but rather an international institution in a form that is not characteristic for Eastern or Central Europe. Since its founding in 1995, OSA has emerged as an important international archives with users from all around the world. Early on in its development, OSA implemented liberal access procedures to its collections and embraced modern technologies in order to better serve the need of the academic community. The PA project or online repository was the logical outcome of the “archival laboratory,” a concept by which OSA gave precedence to the context of its documents rather than to the document *per se*. It might seem as if OSA has not been interested in the physical condition of its holdings, but once they are digitized and preserved more emphasis is placed on the acquired information and the knowledge deriving from it. Despite the fact that OSA improved the physical access to and intellectual accessibility of the fonds, through our experiences with the Research Room and its daily activity we realized that numerous problems, similar to those also faced in other archives, still persisted. As a consequence of our investigation, we have detected problems and obstacles that researchers still face.

In 2008, OSA initiated a project that sought to use academic crowdsourcing as a problem-solving tool for improving scholars’ and other users’ working environment. Initially it was intended that crowdsourcing would be used as a tool for digitizing archival fonds but soon OSA decided to implement solutions that would tackle much more substantial problems and issues.

One of the important obstacles has been locating archival collections of particular interest for scholars. PA provides information on the location and the nature of the physical source, thus helping researchers to locate relevant archival fonds and collections and go back to the physical source. The second important obstacle has been closely connected with the given taxonomy and rigid organization of archival fonds and collections. The hierarchical organization of traditional archives has been still difficult to use especially for those not accustomed for scholarly research, i.e. new users. With its library-type search, PA gives access to archival documents on the item level. Absence of online archival documents is the third important obstacle because archives have been pretty slow to digitize their fonds. PA proposes to capture the individual efforts of scholars to provide general access to archival

materials in digitized format. Finally, the lack of tools to study online documents directly also proved to be an important obstacle. While doing their own digitization of the documents they are consulting, scholars often end up with disorganized copies on their computers. In order to prevent this, PA creates PDF images, and offers tools to organize, describe and view them. It also provides optical character recognition (OCR) of images so that they can easily be converted into text documents which are easy to work with. The tools also facilitate making digital annotations on the page.³⁶

PA was thus designed to help find various types of archival sources such as documents of historical value, testimonies of the past, and other original documents that are somehow unique and generally non-published. PA helps users to search for individual documents by type, subject, name, language, country, date, source, etc.. It also helps locate fonds and collections of interest relevant to research through information on original sources, and allows a user to collect and store documents of interest in a personalized PA online digital repository.

The use of PA starts with a user's registration within the system. As a guest user, one will be able to browse, search and read all public documents in the repository. To gain access to personalized services and other features one needs to register with PA. Before assembling uploaded documents one can upload and store digital files in the PA shoebox. One can decide to make documents public or to keep them private for up to two years. By assembling files one can create documents and OSA guarantees that the original files constituting the document remain unaltered.³⁷ Their integrity will not differ over time. However it is highly encouraged to disclose one's documents to the public because that has been the major reason behind building an environment to support collaboration and sharing of digital files.

Contrary to hierarchical systems, tagging by attaching descriptive keywords to them is an easy way to categorize documents. These keywords are extremely useful if the same word does not appear elsewhere on or in the document. More specific instructions on how to register to PA, upload, assemble and described document, or how to cite documents and use tags, are described on PA web site.

Once researchers upload their digitized archival materials, PA becomes a useful tool to organize documents in thematic folders and

36 Parallel Archive [cited: 2013-04-20]. Available at: <http://www.parallelarchive.org>

37 Ibid.

sub-folders. Newly created documents can also be read optically and then annotated, edited and searched. As previously mentioned, the OCR program is a way of converting image files into a text form suitable for editing and full text searching. In theory the program can read 200 languages, and it automatically prepares image files for text conversion, and corrects contrasts, image quality orientation and positioning. At present the number of languages recognized by the OCR program includes all major European, Eastern and Central European languages.

Text versions of digital images files can only be edited by the owner of the document in the full description page. Any improvements made to the text version considerably improves the searchability of the document across the PA system, and the text version as PDF files can be downloaded by any registered user. The capability to download any document to any personal computer or to send it via email is another important characteristic of PA.

Scholars and researchers can benefit from using PA because it has a permanent Uniform Resource Identifier (URI) for access and citation. The URI creates direct and permanent access to stored documents. Thus, scholars can upload, store, study and manage their digital materials online. In the same manner URIs provide direct links to primary sources in publications and papers. Finally, PA exercises curatorial control over content to maintain a scholarly standard.

Having these characteristics, PA becomes a network for scholars to exchange their ideas. In practice it means that users can find scholars with similar research interest and collaborate. However, they can also form special research groups around specific documents or broader topics (an event, person or period). In many cases they can also share rare documents and offer different interpretations or comments. By uploading their documents to PA users enrich PA networking in many senses. They tag documents of interest to enhance access but also comment on others' documents to share advice or expertise.

To sum up, PA could be defined by its three important roles: (1) it is a digital archive, a repository made by users where they can upload and search for documents, (2) it is a personal workspace where users can store their documents, but also work with them directly online and access them anywhere through the internet, and (3) it serves as a collaborative network environment for scholars and specialists where they can discuss documents and other important research issues.

Still, PA has been facing considerable setbacks. This was visible from our first user testing report conducted at the beginning of the project in 2008. The diversity of users allowed OSA to test the system from the perspective of both the student and the scholar.³⁸ Although the *Report on PA User Testing* was generally positive it also brought constructive criticism.³⁹ Users reported not only on obstacles but also on more serious problems during the user testing. These problems were not only technical ones, and often concerned broader issues, linked to the role of PA as such, and the future direction in which it should develop. The results in the *Report* were presented in four sections and the Report ends with the analyses of some of the issues which came out of the user testing. In the first module there were minor obstacles that were easily fixed throughout the system. The second module contained more serious problems which were open for debate resulting in rethinking some of the final PA's modules. In the third module the users came up with proposals for further development of the system.⁴⁰ Users had a lot of new ideas and tools for future development relating to the interactive side of PA and forums. Finally, even though we insisted more on the issues that users had and less on their positive attitude towards the system, there was also mention of when users particularly liked something in a module.

Parallel Archive emerged thus as collaborative and crowdsourcing project, to contribute to better collaboration and cooperation between scholars by uploading their digital files online and in this way it is enhancing the scholarly debate. Although considerable improvements were made as a result of this user testing, however, PA today is confronting some issues that are still difficult to resolve, from copyright issues to collaboration.

Conclusion

Nowadays, Parallel Archive at the Open Society Archives is slowly developing but not to such extent as it was envisaged a couple of years ago. The recent feedbacks on usability and friendliness of the Parallel Ar-

38 OSA had 13 users, 10 MA students and 3 internal members of OSA. Although the students were well over represented it would have been better to test the system with more scholars and specialists.

39 Herman, Lise. Report on PA user testing. Budapest: Open Society Archives, November 2008. P. 2.

40 Ibid., p. 12.

chive have shown that some problems still remain unsolved.⁴¹ The great majority of users like it as a tool for storing digital files and other functions (assembling and editing documents, OCR, tagging, etc.). However, they were not very enthusiastic about document description and, even more importantly, about making the documents on which they were working public. There are several reasons for such an approach on the part of users. Some mentioned lack of time for uploading and especially for describing documents, while others remained hesitant to make their digital records public and as such accessible for other users. Although we are aware of the intellectual effort involved in source analysis and interpretation, some researchers still do not wish to share their work and believe that they should have privileged access to the sources. In this way they are challenging the whole principle of making archival documents accessible for the many others who are not able to physically visit the Open Society Archives. Thus, from the total number of uploaded documents almost 50 percent are not available for public, and we may conclude that the research strategies of end users have not changed in accordance with or as a result of the changes that affected archives and the role of archivists today. New user studies are needed in order to establish the relationship between modern archives that emphasize access to historical records, and end users who should be aware that possession of the historical source does not necessarily imply an intellectual monopoly of its context, because a historical record can be variously interpreted from many perspectives and disciplines. Until this perception is changed, it is difficult to expect collaboration and the emergence of forums in which scholars and others would discuss particular sources or debate on specific historical topic. In any case the Open Society Archives, with its Parallel Archive, set up a new approach toward historical documents in the digital world and offered its traditional users a new strategy *via* crowdsourcing for better and more diverse research but also for mutual collaboration. To sum up, the Parallel Archive still waits for the new type of researchers to fully use all of its potential and benefit from it.

41 In spring 2013 the author of this paper had several discussions with end users (mostly university professors) to critically reflect on benefits and problems concerning PA.

References

- Altman, Burt; John B. Nemmers. The usability of on-line archival resources: the Pólaris Project Finding Aid. // *Special collections and archives* 11(2001) [cited: 2013-03-28]. Available at: <http://diginole.lib.fsu.edu/cgi/viewcontent.cgi?article=1010&context=sca>
- Ancient lives: decoding papyri [cited: 2013-04-30]. Available at: <http://ancientlives.org>
- Brabham, Daren C. Crowdsourcing as a model for problem solving: an introduction and cases. // *Convergence: the international journal of research into new media technologies* 14, 1(2008), 75-90 [cited: 2013-03-17]. Available at: http://www.clickadvisor.com/downloads/Brabham_Crowdsourcing_Problem_Solving.pdf
- Cruikshank, Kate; Caroline Daniels, Denis Meissner, Naomi, L. Nelson, Mark Sheldstad. How do we show you what we got?: access to archival collections in Digital Age. // *Journal of the Association for History and Computing* 8, 2(September 2005) [cited: 2013-04-11]. Available at: <http://quod.lib.umich.edu/j/jahc/3310410.0008.203?rgn=main;view=fulltext>
- Estellés-Arolas, Enrique; Fernando González-Ladrón-de-Guevara. Towards an integrated crowdsourcing definition. // *Journal of information science* 38, 2(2012), 189-200 [cited: 2013-02-14]. Available at: <http://jis.sagepub.com/content/38/2/189>
- Gilliland, Anne and Sue McKemmish. The role of participatory archives in furthering human rights, reconciliation and recovery. // *Atlanti: review for modern archival theory and practice* 24(2014), 79-88.
- Herman, Lise. Report on PA user testing. Budapest: Open Society Archives, November 2008.
- Historypin [cited: 2013-04-30]. Available at: <http://ancientlives.org/story>
- Holley, Rose. Crowdsourcing: how and why should libraries do it? // *D-Lib magazine* 16, 3/4(March/April 2010) [cited: 2013-03-22]. Available at: <http://www.dlib.org/dlib/march10/holley/03holley.html>
- Holley, Rose. Crowdsourcing: more cool sites to give libraries, archives and museums inspiration. // *Rose Holley's Blog – views and news on digital libraries and archives* (11 February 2012) [cited 2012-11-21]. Available at: <http://rose-holley.blogspot.hu/2012/02/crowdsourcing-more-cool-sites-to-give.html>
- Holley, Rose. Crowdsourcing strategies for archives. Power Point Presentation, 8-12 November, 2010 [cited: 2012-12-02]. Available at: <http://www.slideshare.net/RHmarvellous/naa-archives-20-week-roseholleycrowdsourcingnov-2010>
- Howe, Jeff. The rise of crowdsourcing. // *Wired magazine* 14.06(June 2006) [cited: 2013-02-14]. Available at: <http://www.wired.com/wired/archive/14.06/crowds.html>
- Huvila, Isto. Participatory archive: towards decentralized curation, radical user orientation and broader contextualization of records management. // *International journal on recorded information* 8, 1(2008), 15-36 [cited: 2013-04-09]. Available at: <http://link.springer.com/article/10.1007/s10502-008-9071-0#page-1>
- Huvila, Isto. What is a participatory archive?: for real (?). // Isto Huvila's blog (08/31/2011) [cited: 2013-04-08]. Available at: <http://istohuvila.eu/what-participatory-archive-real>

- Ketelaar, Eric. Archival temples, archival prisons: modes of power and protection. // *Archival science* 2(2002), 221-238 [cited: 2013-04-06]. Available at: <http://home.hccnet.nl/e.ketelaar/ArchivalTemples.pdf>
- Katelaar, Eric. The Archives as a Time Machine, the ICT-industry and public sector partnership: to promote the preservation and accessibility of the European archival heritage. Closing speech of the DLM-Forum 2002, Barcelona May 8, 2002 [cited: 2013-04-13]. Available at: <http://www.mybestdocs.com/ketelaar-e-dlm2002.htm>
- Parallel Archive [cited: 2013-04-20]. Available at: <http://www.parallelarchive.org>
- Rifkin, Jeremy. *The age of access: the new culture of hypercapitalism, where all of life is a paid-for experience*. New York: Tarcher/Putnam, 2000.
- Schenk, Eric; Claude Guittard. Towards a characterization of crowdsourcing practices. // *Journal of innovation economics* 1, 7(2011), 93-107 [cited: 2013-03-14]. Available at: <http://www.cairn.info/revue-journal-of-innovation-economics-2011-1-page-93.htm>
- Star Wars Uncut [cited: 2013-04-30]. Available at: <http://www.starwarsuncut.com/about>
- Theimer, Kate. Building participatory archives. // *ArchivesNext* (August 22, 2010) [cited: 2013-03-22]. Available at: <http://www.archivesnext.com/?p=1536>
- Theimer, Kate. Exploring the Participatory archives. // *ArchivesNext* (August 30, 2011) [cited: 2013-04-09]. Available at: <http://www.archivesnext.com/?p=2319>
- Thomas, Stuart. 9 examples of crowdsourcing, before 'crowdsourcing' existed. // *Social networking* (09.15.11) [cited: 2012-11-21]. Available at: <http://memeburn.com/2011/09/9-examples-of-crowdsourcing-before-%E2%80%98crowdsourcing%E2%80%99-existed/>
- Trompette, Pascale; Valerie Chanal, Cedric Pellissier. Crowdsourcing as a way to access external knowledge for innovation: control, incentive and coordination in hybrid forms of innovation. // 24th EGOS Colloquium, Jul 2008, Amsterdam, France [cited: 2015-02-25]. Available at: <https://hal.archives-ouvertes.fr/halshs-00367373/document>
- Vukovic, Maja; Claudio Bartolini. Towards a research agenda for enterprise crowdsourcing. // *Leveraging applications of formal methods, verification, and validation / 4th International Symposium on Leveraging Applications, ISoLA 2010, Heraklion, Crete, Greece, October 18-21, 2010, Proceedings, Part I*; editors T. Margaria, B. Steffen. Pp. 425-434 [cited: 2013-03-14]. Available at: http://link.springer.com/chapter/10.1007%2F978-3-642-16558-0_36#page-1
- What's on the menu? [cited: 2013-04-30]. Available at: <http://menus.nypl.org>

SURADNIČKI I *CROWDSOURCING* ARHIVI PARALELNI ARHIV KAO STUDIJA SLUČAJA U ARHIVU OTVORENOG DRUŠTVA

Sažetak

Središnji je problem arhiva, knjižnica i muzeja kako korisnicima pružiti građu koju trebaju i žele istražiti. Svrha je ovog rada obraditi problem suradnje i sudjelovanja korisnika u arhivskom okruženju. Ponajprije se bavi participativnim arhivima o kojima se može govoriti kao o suradničkim arhivima ili o arhivima utemeljenim na tzv. *crowdsourcing*-u. U ovom radu *crowdsourcing* se smatra arhivskom tehnikom ili metodom, ali i širim društvenim angažmanom. U prvom dijelu rada pomno je razjašnjen korišteni konceptualni okvir te temeljni pojmovi i njihove definicije. Slijedi opis nekih od najznačajnijih suvremenih projekata koji koriste *crowdsourcing*. U drugom dijelu rada promišlja se o suradničkim praksama i onima koje koriste *crowdsourcing*, a koje se primjenjuju u znanstvenom i suradničkom projektu, uz studiju slučaja, Paralelni arhiv (PA) koji je pokrenuo Arhiv Otvorenog društva (OSA). Zaključno, autor promišlja o korisnicima Paralelnog arhiva i načinu prevladavanja problema suradnje te ukazuje na pojedine suvremene trendove u Arhivu Otvorenog društva.

Ključne riječi: suradnički arhiv, *crowdsourcing*, Paralelni arhiv, participacijski arhiv, OSA: Open Society Archives, istraživanje korisnika