

CROWDSOURCING IN HISTORY AND DIGITAL HUMANITIES

Aleksandra A. Fostikov

The Institute of History Belgrade, Belgrade

e-mail: aleksandra.fostikov@iib.ac.rs

Abstract: *Crowdsourcing data in history and digital humanities is the one of the newest technique that is applying in gathering of all sort of the historical sources essential for the researching work. Moreover, it is important for the one more process - the preservation of the cultural heritage. As this method still is not in wide use particularly because of the digital gap within geographical and cultural net, this paper focuses on the meaning and the significance of the crowdsourcing method and its usefulness in digital humanities, as history science itself. Bearing in mind all mentioned, this article deals also with the implementation of this method, as with the consideration of the new methodology of scientific research in this case, not only for its presentation to the researchers but also to the wide public. Additionally, here we have considered the value of this method through the potential outcomes, opportunities as the possible problems. For now, this new method of get-together data for historical science and other humanities need time and resource, especially human to be implement. In the future, the AI could support using of the crowdsourcing, if one will be created for this purpose.*

Keywords: *crowdsourcing, history, digital humanities, methodology, cultural heritage.*

Апстракт: *Кроудсорсовање података у историји и дигиталној хуманистици представља једну од најновијих техника које се примењују ради прикупљања свих могућих врста извора, неопходних за рад истраживача. Иста техника, штавише, важна је и за још један процес – очување културне баштине. Како дотични метод још није у широкој употреби, а углавном због дигиталног јаза који постоји у географској и културној умрежености, овај рад се примарно бави значењем самог термина кроудсорсинг и користима од његове примене у дигиталној хуманистици, као и у самој историјској науци. Имајући у виду све наведено, у чланку се такође разматра и примена ове методе, али и нова методологија научног истраживања у датом случају, не само ради презентације исте истраживачима, већ и широј публици. Додатно, размотрена је и вредност ове методе прикупљања података кроз потенцијална достигнућа, те могућности и проблеме њене примене. Иако за сада овај нови метод прикупљања података у историјској науци и другим хуманистичким наукама захтева време и ресурсе, те посебно примену људског рада, његову олакшану реализацију у будућности могла би да потпомогне и вештачка интелигенција, уколико би она била створена за ове намене.*

Кључне речи: *кроудсорсовање, историја, дигитална хуманистика, методологија, културна баштина.*

DIGITAL HUMANITIES, CULTURAL HERITAGE AND CROWDSOURCING – INTRODUCTION REMARKS

Not so long ago, just few decades in the past, at the almost end of the 20th century and at the first decade of this, 21st century, new age in the history science as in the humanities started parallel with the third and fourth industrial revolution. From the beginning of the alliance between history and other sciences, in mid-sixties of the 20th century, through the idea of the *total history* and introduction of computers to humanities, all lead to the new phase in historical research and methodology. Finally, with the appearance of the computers and then WWW (World Wide Web) in the late eighties and using its many possibilities, the new academic field formed firstly entitled *humanities computing*, and then in first years 21st century it was renamed in *digital humanities* (DH). In the last two decades, with fast steps forward humanities together with historical science established strong DH field within the net of institutions, project and scientists.¹

Today, DH have many branches and it still dividing in more and more sub disciplines, like as Digital History, than again Digital Diplomatic, HGIS (Historical geographic information system) etc. In addition, that brings new questions and possibilities not only to historiography, but also to methodology of humanities requesting finding of the new methods and technique, as gathering and using the multiplicity of data in research process.² In the same time, work on the preservation of the cultural heritage especially in its digital form became essential and new Alexandria library created in the form of net of the virtual GLAMs (galleries, libraries, archives, and museums).³

In the meantime, our world quickly transformed and adjusted to the new *Digital Age* (The Information Age, Computer Age or New Media Age).⁴ Then in the last decade, we are witness of the birth of the one more new era - *Imagination Age*, in which creativity and imagination become the primary creators⁵ opening the gate for the New Global Culture, engage communication, public history, citizen science, and so more.⁶

¹ Long 2005: 177-186; Fostikov and Mandić 2007: 40-46; Mandić 2008; Fostikov and Isailović 2014: 19 - 20. About third and fourth industrial revolution: Schwab 2016.

² Fostikov and Isailović 2014: 21.

³ Povroznik 2020: 125-136; Al Abdulla, Ames, Bray, Candela, Chambers, Derven, Dobрева et al. 2019.

⁴ Digital Age is primary the historic period that began in the mid-20th century characterized by the shift from traditional industry to an economy based on information technology. Nevertheless, “one of the most important characteristics of the age of information is a vast number of information, which were just until yesterday almost unknown to an average person”. Fostikov and Mandić 2007: 42. Already in 2004, it was pointed out that there is “800 MB of new informations per an average person which means that takes nine meters of books to get all of those informations printed.” Fostikov and Mandić 2007: 42 note no. 9.

⁵ Imagination Age is period that is followed the Digital Age. King 2007; King and Fouts 2012. Except the creativity and imagination, it is also characterized by the rise of collaboration and communication around the globe. In addition we could see the rise of the Virtual reality (VR) and Augmented reality (AR). For the terms VR and AR, see Encyclopædia Britannica online edition.

⁶ For the terms, “public history”, “citizen science”, see Wikipedia for short explanation. Also, see those tags at search engines like as Google etc. In addition, see note above no. 5.

Finally, with the spread of the Internet, the parallel universe formed in public space on the web portals and platforms. Through the net of the social and business networks, like as Facebook, or LinkedIn, or all those similar places like Discord⁷, let's mention only some of the existing, numerable private and public groups were made about infinite topics. And between them, in that extensive new virtual world we could find every day more and more resources and data significant for history and other humanities sciences. Moreover, there are lot of very valuable data for every historical period as for cultural heritage. Furthermore, internet users – crowd or public, uploaded almost all of those data as content in accordance with their own interests.⁸ Because of this interaction, we should have in front of our eyes that “digital humanities should not only be about the production of knowledge. It’s about challenging the ways that knowledge is represented and shared”.⁹

In that perspective, a new method for compiling data in digital humanities – crowdsourcing,¹⁰ easily found its way to the public that have interest for its own history and cultural heritage. Having that in mind, it is not only necessary but is also desirable to engage crowd to compile and share data in more than in one way – from gathering the simple historical data to sharing their own cultural heritage. In that way the cultural heritage itself is also have more chance to be better preserved. Finally, in this sense, it is obligatory not only to have exactly definition of the crowdsourcing meaning within DH field but also to know what its outcome is, opportunities as the possible problems.

CROWDSOURCING IN DIGITAL HUMANITIES

Today DH area is very wide and absolutely still novelty tree of the humanities that is almost newborn¹¹ and talking about the idea of implementing the crowdsource process itself in it is like a searching the way through forest. Moreover, the main question - what is the crowdsource in case of the DH, and how it should be used is still under consideration. And there are also more questions that are not simple as it appearances, like as: how to apply it, in what way it should be applied to get not the only positive response but

⁷ About Discord, see its homepage: <https://discord.com/company>.

⁸ Data uploaded for local history and cultural heritage is maybe most valuable in that context. While GLAMs and other cultural and research institutions are concentrated their efforts to upload primary their collections, crowd alone give us data about e.g. their families, settlements, history, and local geography. There so many pages just dedicated to some of those mentioned topics. Moreover, in many cases those data could be used for research some of questions that until global internet and public data could be only investigated by taking the terrain reconnaissance survey.

⁹ Sample 2013: 256; Fostikov and Isailović 2014: 22 note no. 16.

¹⁰ About crowdsourcing, see more detail below.

¹¹ Some of questions and answers could be find in *A Companion to Digital Humanities* from 2004 and in *A New Companion to Digital Humanities* from 2016. Schreibman, Siemens and Unsworth, John.Schreibman 2004; Siemens and Unsworth 2016. Those two *Companions* also show how our perception about DH is changing and how DH is changing itself by time.

exact required and verified data, which set is in the same time appropriate for researching the main topic (researcher question or these) are etc.

In the almost two last decade, from the time when the term was coined (2006) until today the idea of crowdsourcing change in its essence. From elementary formulation that originated in business and in connection to labor and capital or payment and less payment, today this term in short is meaning a using of the crowd to do wide range of online activities in short. Between varieties of the definitions of using the crowdsource in DH, the one that is could be taken as a good start, is that given by Melissa Terras in 2016. At that time, Terras described the crowdsourcing in the DH as the “practice of using contributions from a large online community to undertake a specific task, create content, or gather ideas” (Terras 2016: 420). In addition, Terras pointed out that it is “an easily accessible way in which we can harness computational platforms and methods to engage a wide audience to contribute to our understanding of society and our cultural inheritance” (Terras 2016: 436).¹²

Talking about the crowdsourcing in DH it also should be noted that just at present, there are around five thousand results of its mentions in the Google Scholar alone about it, and around 184,000 results in case of using the Google search engine. Moreover, many of them have data about the last or current projects. Parallel there is also a wide field of digital social sciences that is in some ways integrating with DH.¹³ Finally, with an escalation of new technologies and the high-speed internet, the crowd alone give us more and more data every day. So today, there is one more question: how to use all available data and how to include them in some form applicable not only for research purposes but also to preserve them too, no matter that in fact there are not the part of the projects itself.¹⁴

As it could be seen, except the definition of the crowdsourcing in DH that is partially resolve with Terras definition, there are few crucial questions more. Those are how properly to apply crowdsourcing, in which way tasks should be organized to get positive response and how to get exact required and verified data for researching topic? Moreover, there is next: it the crowd anonymous or not? Finally most important are how to use all available open source data and how to preserve all of data and dataset in proper manner?

SOME ANSWERS AND MORE QUESTIONS

Before the giving the best possible answers on those questions first we must see examples of applying the crowdsourcing in two area: cultural heritage and history as science, as main contributors in this case of interest.

¹² Terras also gave an extensive review of the evolving of the crowdsourcing in DH. Terras 2016: 420-439.

¹³ Tags: “crowdsourcing” and “digital humanities” on Google and Google Scholar on 19 November 2022.

¹⁴ For now, there is almost only site for preservation some data but only for the websites - The Internet Archive Wayback Machine is service that allows people to visit archived versions of Web sites. *The Internet Archive Wayback Machine*.

In the case of the crowdsourcing in cultural heritage, the crowd is not usually anonymous, and in some circumstances, researchers from variety institutions or GLAMs take a part in those projects. Two years ago, the “*final report*” about crowdsourcing in cultural heritage, by Robert Davis gave us big data about the status of numerable projects at that moment, as for open question, proposals and guidance.¹⁵ On the other side the question of the crowdsourcing in the digital history is in real everything what we can imagine in digital world as it is history and historical sources itself – from pure, simple facts about contemporary history to all data and databases that include any segment, theme or history field from word to the image and back. Furthermore, cultural heritage is one of main segments of history and every periods of it. Material remains or tangible culture as intangible culture and natural heritage give us more data about history and historical geography.¹⁶

In the core of humanities, history spreads in other related sciences and all of them have their time way alone. Even if the crowdsourcing in the digital history is in main connection with all we already said most of the digital history projects concentrated at one question in time, and usually have even more narrow so-called crowd, especially in some branches. In addition, all mentioned above raised one another question – is there still differences between history and other sciences in digital humanities and digital world in which the smallest branch is in connection and alone. In some cases, the history projects and those in other humanities, as in cultural heritage have the same goals, to see full picture of some object.¹⁷

Again, another question raised – who or what will manage all possible data about history in the net of other sciences. For now, this new method of get-together data for historical science and other humanities need time and resource, especially human to be implement. In this moment even with work of all researchers and whole crowd, there is enough data only on internet for working more than several generations.¹⁸ In the future, the AI (Artificial intelligence) could support using of the crowdsourcing, if one will be created for this purpose.¹⁹

¹⁵ Davies 2020. Also, see: Ridge, Blickhan, Ferriter et al. 2021.

¹⁶ The good example of crowdsourcing in that meaning is the virtual reconstruction of the stone bridge of Plaka, in which case the photographs from crowds was used. Stathopoulou, Georgopoulos, Panagiotopoulos and Kaliampakos 2015: 297-298.

¹⁷ See the note no. 16 above.

¹⁸ As it already pointed out “it is now quite clear that historians will have to grapple with abundance, not scarcity”. Berry 2012: 2.

¹⁹ The use of AI in DH has already begun, and as the previous examples show, the possibilities of using AI can be endlessly diverse. Pavlidis 2022: 026-034, with bibliography about this topic. About projects that include AI and crowdsourcing, see: Stančić, Seljan, and Ivanjko 2019: 380-390. The one of future possibilities could be to put one AI on one task like as bring together, mark and tag for e.g. data about one settlement or one custom from different forums, groups and sites through the internet, images of one church, house etc.

CONCLUSIONS AND RECOMMENDATIONS

At the end, there is more and more to be said. All mentioned earlier is just a scratch of the main topic stated above. Benefits of using the crowdsourcing in DH require solving some problems. However, systematically as all other problems in sciences those will be resolved by time. Moreover, in this new world where the virtual and digital can create the bridge between researchers and public, using the crowdsourcing to gathering data and preserve cultural heritage must be used. Starting even the smallest projects, for example taking data in this way about simple facts as the way of local customs or of celebrating “slava”, or wedding, or taking picture of local church and graveyard from all the angles, or giving data about local historical geography are not baby, but big steps in this moment. Therefore, it is important to point out that there is a plan for crowdsourcing apply on the territory of the Serbian ethnic space, as well on the own territory of the Serbia regardless of the ethnic origin of the population, accordingly to the assigned research questions.

REFERENCES

- Al Abdulla, Aisha, Sarah Ames, Paula Bray, Gustavo Candela, Sally Chambers, Caleb Derven, Milena Dobrova et al. Open a GLAMLab: Digital Cultural Heritage Innovation Labs, Book Sprint, Doha, Qatar, 23-27 September, 2019. University College London (UCL) Qatar, 2019. *Discord*. <https://discord.com/company>
- Berry, David M. “Introduction: Understanding the Digital Humanities”. *Understanding Digital Humanities: The Computational Turn and New Technology*, ed. David M. Berry, Palgrave Macmillan, 2012.
- Davies, Robert. *Crowdsourcing in cultural heritage. Final Report*. Europeana Pro, 2020. https://pro.europeana.eu/files/Europeana_Professional/Projectpartner/EuropeanaCommonCultureProjectFiles/Crowdsourcing%20study%20report.pdf (Accessed 19. 11. 2022.)
- Encyclopædia Britannica online edition. Encyclopædia Britannica, Inc., 2022. <https://www.britannica.com/>
- Fostikov, Aleksandra and Mandić, Slobodan. “New categories of historical sources: E-mail and Forum. Internet communication and History”. *Review of the National Center for Digitization* 10 [SEEDI Communications 3] (2007): 40-46.
- Fostikov, Aleksandra and Isailović, Neven. “Digital humanities or digital versus humanities”. *Review of the National Center for Digitization* 24 (2014): 19 – 23.
- King, Rita J. “The Emergence of a New Global Culture in the Imagination Age”. British Council Essays, 2007. <http://archive.is/20120722122948/www.britishcouncil.org/tn2020-research-findings-authors-king.htm> (Accessed 19. 11. 2022)

- King, Rita J. and Fouts, Joshua. "Global Communications and the Rise of Social Media: The Future of International Cultural Engagement". White Paper-Theme Four Distributed in preparation for discussion at Salzburg Global Seminar Session 490 Public and Private Cultural Exchange-Based Diplomacy: New Models for the 21st Century, Salzburg, Austria April 28–May 2, 2012; 2012. https://www.salzburgglobal.org/fileadmin/user_upload/Documents/2010-2019/2012/490/whitepaper2_490.pdf (Accessed 19. 11. 2022)
- Long, Pamela O. "The Annales and the History of Technology". *Technology and Culture*, 46 - 1 (Jan., 2005): 177-186. Published by: The Johns Hopkins University Press and the Society for the History of Technology.
- Mandić, Slobodan. *Kompijuterizacija i istoriografija 1995–2005*. Beograd: Istorijski arhiv, 2008.
- Pavlidis, George. "AI trends in digital humanities research". *Trends in Computer Science and Information Technology* 7-2 (2022): 026-034.
- Povroznik, Nadezhda. "Digital History of Virtual Museums: The transition from Analog to Internet Environment". *Proceedings of the Digital Humanities in the Nordic Countries 5th Conference, Riga, Latvia, 21-23 October 2020, Central EUROpe workshop proceedings* 2612 (2020): 125-136.
- Ridge, Mia, Blickhan, Samantha, Ferriter, Meghan, Mast, Austin, Brumfield, Ben, Wilkins, Brendon, Cybulska, Daria, Burgher, Denise, Casey, Jim, Luther, Kurt, Goldman, Michael Haley, White, Nick, Willcox, Pip, Carlstead, Sara, Brumfield, Coleman, Sonya J. and Prytz, Ylva Berglund. *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage*, British Library, 2021. <https://britishlibrary.pubpub.org/the-collective-wisdom-handbook-perspectives-on-crowdsourcing-in-cultural-heritage---community-review-version> (Accessed 19. 11. 2022.)
- Sample, Mark. "The Digital Humanities is not about Building, it's about Sharing". *Defining Digital Humanities: A Reader*. Edited by Melissa Terras, Julianne Nyhan, Edward Vanhoutte, Routledge, 2013: 255-257.
- Stančić, Hrvoje and Seljan, Sanja and Ivanjko, Tomislav. "Using AI and crowdsourcing in digitisation and processing of archival materials". *History of modernity: information resources, methods and research practices in Russia and abroad*. Ed. by. Malysheva, E. P., Moskva: Russian State University for the Humanities, 2019: 380-390.
- Stathopoulou, Elisavet (Ellie) Konstantina, Andreas Georgopoulos, George Panagiotopoulos and Dimitris Kaliampakos. "Crowdsourcing Lost Cultural Heritage." *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-5/W3 (2015): 295-300.
- Terras, Melissa. "Crowdsourcing in the Digital Humanities". *A New Companion to Digital Humanities*. Eds. by Schreibman, Susan, Siemens, Ray and Unsworth, John. Wiley-Blackwell, 2016: 420-439.
- Schreibman, Susan, Siemens, Raymond George and Unsworth, John. *A Companion to Digital Humanities*. Malden MA: Blackwell Pub, 2004.
- Schreibman, Susan, Siemens, Ray and Unsworth, John editors. *A New Companion to Digital Humanities*. Wiley-Blackwell, 2016.
- Schwab, Klaus. *The Fourth Industrial Revolution*, Geneva: World Economic Forum, 2016.
- Wikipedia. The Free Encyclopedia*. <https://www.wikipedia.org/>

Александра А. Фостиков, Историјски институт Београд, Београд

КРОУДСОРСОВАЊЕ У ИСТОРИЈИ И ДИГИТАЛНОЈ ХУМАНИСТИЦИ

Не тако давно, крајем 20. и почетком 21. столећа, започело је ново доба у историји и хуманистици, паралелно са трећом и четвртм индустријском револуцијом. Са рађањем сарадње између историје и других наука, преко појаве такозване „тоталне историје“, до формирања дигиталне хуманистике као новог академског поља, протекло је свега око осамдесет година. Истовремено, и свет је ступио у нову еру познату као „дигитално доба“, на коју се наставило и „доба имагинације“, а у ком је дошло и до успона комуникације, те појаве између осталог и такозване „јавне историје“. Паралелно, грана се и сама дигитална хуманистика развијајући свој пандан за скоро сваку поједину грану историје и других хуманистичких наука. Додатно, рад на очувању културне баштине, а посебно у дигиталној форми, доводи и до формирања мреже библиотеке, музеја, архива и галерија у виртуелном простору. Коначно, са развојем широко доступног интернета долази и до формирања новог виртуелног простора на интернету, у ком се окупљају безбројне групе формиране око појединих питања – од самих социјалних мрежа до појединих издвојених група унутар њих. Сваким даном у том бескрајном простору налази се све већи број података од значаја за историјску и остале науке, као и саму културну баштину. У таквим оквирима јавља се у историјској науци и дигиталној хуманистици и нова метода прикупљања података – позната у енглеском језику под називом *кродсорсинг*, који је под тим именом прихваћен и у највећој већини осталих језика.

Ипак, како би размотрили и разумели овај метод било је неопходно и дефинисати га у савременим оквирима историјске науке. Док је првобитно у општем значењу тај термин подразумевао везу рада и капитала, односно примену методе где се радом мноштва завршавају послови за мање новца, данас се под овим термином у историјској науци и дигиталној хуманистици у најширем смислу сматра практично онлајн добротворни рад јавности у стварању и подизању разних садржаја на интернет платформе сходно задатим критеријумима, почевши од таговања до учешћа корисника интернета у стварању дигиталне културне баштине. Поводом овога јављају се бројна питања - на који начин окупити жељену публику, односно јавност која има дигитална знања, какве задатке поставити и како креирати базе података које ће затим бити и широко доступне, те како развити саму методологију – да ли путем интервјуа, или пак давањем ситних задатака, а проблем представља и којим путем треба све то организовати како би се добио позитивни одговор од саме јавности. Такође, неопходно је и отворити питање да ли је у тим случајевима јавност анонимна и када, да ли, и у којим околностима имена информатора остају тајна или јавна, као и како организовати цео систем прикупљања података да би

прикупљени подаци били валидни и од значаја, те у којим случајевима могу то бити, а у којим не.

На основу досадашњих примера примене кроудсорса углавном у иностранству, искуства су и више него позитивна. То се види у случају више спроведених пројеката базираних на прикупљању чињеница из ранијих, али и савремених периода. Значај и могућности примене кроудсорсинга у историји и дигиталној хуманистици савременог доба јасно се могу видети и у два домаћа пројекта које је реализовала Народна библиотека Србије, у сарадњи са Еуропеаном, једном насталом поводом прикупљања сећања на први светски рат и другом поводом миграција. Ипак, могућности кроудсорсинга заправо су и далеко веће. Од прикупљања појединих факата преко фотографије или легенде, ова метода омогућава нам да завиримо и откријемо далеко више не само о савременим периодима, него и старијим и то више него икада пре. Ово је могуће не само захваљујући масовности потенцијаног одзива, већ и бројности података, било у речи или слици, који се већ налазе у незамисливим количинама у бескрајном виртуелном простору.

На крају, неопходно је отворити и нека питања, међу којима и питање где историја почиње а где се завршава, и да ли је заиста и даље тако велика разлика између историје и других сродних наука у дигиталној хуманистици, а посебно у гранама историје. Оне су уједно и гране осталих наука, које су сада тек испреплетане, а ако изумемо саму историјску причу и ток, у смислу костура времена и простора. Стога примена кроудсорса доноси низ предности над другим методама у савремено доба. Паралелно, примена кроудсорса омогућава и прикупљање података из културне баштине, као и њено очување. Ово је посебно важно у време када због низа измена у географском простору, као и начину живљења, било материјално или нематеријално наслеђе убрзано ишчезава. Истовремено је културна баштина, већ сама по себи, изузетно добар извор за саму историју, као и за друге сродне хуманистичке науке.

Стога, када се само погледа све наведено, увиђа се да је прикупљање података у виртуелном простору не само од посебног значаја, већ и методолошки значајно за потребе различитих истраживања. Зато, у овом новом свету, где виртуелно и дигитално могу чинити мост између истраживача и јавности, неопходно је наћи пут за што бољу и ширу примену кроудсорсинга. Примера ради, већ ангажовање јавности да макар фотографише и сними своје обичаје, попут славе или свадбе, или пак материјалне објекте или остатке у свом (рецимо) селу – попут цркве или гробља (из свих углова) – било би на много начина корисно, како за истраживаче тако и за очување културне баштине. Ипак, и поред огромне користи, неопходно је истаћи да постоје и проблеми који би требало да буду решени – од усклађивања јавности и задатог проблема, преко буђења интересовања.

Коначно, систем примене кроудсорсинга могао би бити проширен и делимично и прикупљањем већ расположивих података који су доступни на интернету.

Ипак, такав подухват захтевао би, у овом моменту, не само рад свих постојећих истраживача већ и наредних генерација. Систематичност у том смислу, као и брзина прикупљања и класификације података, можда би могли бити надокнађени и ангажовањем вештачке интелигенције, програмиране за те сврхе. Свакако, ма који мали корак којим би се овим путем кренуло и у домаћој науци, биће заправо не само зачетак примене једне нове методе у прикупљању података од значаја, већ ће омогућити и дубље сагледавање како прошлости тако и садашњости, као и боље очување домаће културне баштине. Стога је важно истаћи да постоји план за примену кроудсорсинга на територији српског етничког простора, као и на самој територији Србије без обзира на етничко порекло становништва, сходно задатим истраживачким питањима.