Crowdsourcing digital history online

Despoina Valatsou

doi: 10.12681/historein.218

To cite this article:

Crowdsourcing digital history online

Digital history, among other things, offers the possibility for people to collaborate and work together on historical projects online. The notion of crowdsourcing is essential in this process, as well as in the theoretical study and approach of the field of digital humanities. Digital history online is no longer exclusively practiced by professional historians and scholars; it is increasingly performed by individuals with little or no academic training in the discipline of history, who become involved in this online historical practice out of personal interest and motivation and/or in response to open calls for participation. This contribution aims to highlight some key issues concerning the collaborative aspect in the current production and communication of history on the internet.

Crowdsourcing in general can be described as the process of using an open call to attract a crowd of people who contribute to the materialisation of a project. The actual word was first introduced in 2006 by author and journalist Jeff Howe in an article in *Wired* magazine, entitled “The Rise of Crowdsourcing”, where he discussed the concept in terms of the global economy. He described a new way of involving people who are willing to help or work on a project and offer their time and effort cheaply or even for free. “Remember outsourcing?” he wrote. “Sending jobs to India and China is so 2003. The new pool of cheap labour: everyday people using their spare cycles to create content, solve problems, even do corporate R & D.”

According to DesignCrowd (a website that links customers in search of logo, web, print and design services with professionals in the design industry from all over the world), crowdsourcing as a practice dates as far back as the early
eighteenth century. In 1714, the British government solicited the public’s help in solving what was called the "longitude problem". Indeed, the solution was found by a carpenter’s son, John Harrison, who invented the marine chronometer. Harrison was then given a financial award – something that is not so much part of the current practice of crowdsourcing when applied in the humanities.

Another early example of soliciting people to help accomplish a project is the Oxford English Dictionary (OED). Indeed, the OED is considered by some to be the first crowdsourced project in the humanities. In 1857 the London Philological Society announced its decision to create a new and comprehensive historical dictionary of the English language, from the Anglo-Saxon period onwards. The dictionary would not only offer detailed definitions and usages of English words, but it would also trace and present their etymology over time. In 1879 the same society announced its collaboration with Oxford University Press to publish the dictionary, entrusting lexicographer Sir James Murray with the mammoth task of overseeing its completion. Both Murray and previous editors of the project appealed to the public for help; many people volunteered to read numerous books in order to track down English words and submit quotation slips noting their early and contemporary uses.

Crowdsourcing comes in many forms and diverse contexts. In the software development business, for example, developers often release online beta versions of an application for users to test and improve. The DesignCrowd platform mentioned above serves as a good example of people with design interests and professional or amateur skills contributing with innovative ideas in the completion of projects. In the context of the new network technologies, the online encyclopaedia Wikipedia is the best example as it is written collaboratively by internet users, both amateurs and/or professionals, from all over the world.

Let us now focus on digital humanities, bearing in mind the digital turn in history, which is not so much of a new thing; rather, it is a reflection and reconfiguration process of the theories and of mostly the practices applied within this field of knowledge, both in the scholarly and the public domains. If we understand the term digital humanities to refer to the use of computing tools and processes to facilitate the work of professional humanities scholars, then digital humanities is quite old. Susan Hockey, emeritus professor of library and information studies at University College London, dates the beginning of humanities computing to 1949, when “an Italian Jesuit priest, Father Roberto Busa, began . . . to make an index verborum of all the words in the works of St Thomas Aquinas and related authors”, collaborating with IBM in order to find a computer system to facilitate his task. According to Matthew K. Gold, “after thirty years of work, Busa published the Index Thomisticus first in print and later on CD-ROM and the Web”. From the 1960s onwards there was a steady rise of humanities computing activities, including the foundation of Computers and the Humanities journal in 1966 (the first in its field), the organisation of related conferences, the introduction of humanities computing courses in academic courses, as well as the establishment of relevant academic centres. In the 1980s the personal computer and electronic mail appeared. Historians, to focus our
case, used databases and other software and computer systems to systematise their professional work in a mostly quantitative manner. In the following decade, as the humanities opened up to multimedia applications and the world wide web, orality and visuality vied to become as important as textuality and combine with it. Up to that point, the printed text was (and often still is) considered to be the standard and widely accepted medium for communicating knowledge in the humanities. The 1990s were a landmark decade for the historical discipline, with countless projects introducing historians to the potentials of the new media technologies and the internet. Important in this regard was the Roy Rosenzweig Centre for History and New Media (RRCHNM), founded in 1994 at George Mason University, and the pioneering work by its historians and directors Daniel J. Cohen and the late Roy Rosenzweig (its founder). In an attempt to condense their accumulated knowledge and experience on this field, Cohen and Rosenzweig published, among other works, a thorough introduction to all aspects of digital history entitled *Digital History: A Guide to Gathering, Preserving and Presenting the Past on the Web*.

Susan Schreibman, Ray Siemens and John Unsworth were among the first scholars to introduce the actual term “digital humanities”. They used it in the title of a 2004 volume they edited, *A Companion to Digital Humanities*. This choice of title highlighted a particular moment in the beginnings of this field where the focus was shifting from the technology applied in the study of humanities to the context of the humanities done with the help of that technology. Matthew Kirschenbaum argues for the following definition of digital humanities: “the digital humanities today is about a scholarship (and a pedagogy) that is publicly visible in ways to which we are generally unaccustomed, a scholarship and pedagogy that are bound up with infrastructure in ways that are deeper and more explicit than we are generally accustomed to, a scholarship and pedagogy that are collaborative and depend on networks of people and that live an active 24/7 life online.”

At the start of the 21st century, the humanities have found themselves moving in a constantly expanding and evolving digital environment, the internet, experimenting with a vast array of digital technologies that this environment fosters, mainly the networked technologies – network being the key concept here. That convergence seems to be the new element that is crucial to the digital turn in history today. Within that context, digital history is still an emerging field of knowledge, perhaps even in its early stages of development.

In 2009, Humanities Blast, a blog maintained by Todd Presner, chair of the Digital Humanities Program at UCLA, published the *Digital Humanities Manifesto*, which offers a grandiose but also quite eloquent description of digital humanities and of the reasons why it is of importance today:

Digital Humanities is not a unified field but an array of convergent practices that explore a universe in which: a) print is no longer the exclusive or the normative medium in which knowledge is produced and/or disseminated; instead, print finds itself absorbed into new, multimedia configurations; and b) digital tools, techniques, and media have altered the production and dissemination of knowledge in the arts, human and social sciences. The Digital Humanities seeks to play an inaugural role with respect to a world in which, no longer the sole producers, stewards, and disseminators of knowledge or culture, universities are called upon to shape natively digital models of scholarly discourse for the newly emergent public spheres of the present era.
(the www, the blogosphere, digital libraries, etc.), to model excellence and innovation in these domains, and to facilitate the formation of networks of knowledge production, exchange, and dissemination that are, at once, global and local.13

The RRCHNM offers a comprehensive definition of digital history, in particular: “Digital history is an approach to examining and representing the past that takes advantage of new communication technologies such as computers and the Web. It draws on essential features of the digital realm, such as databases, hypertextualization, and networks, to create and share historical knowledge.”14

It is important to mention that when it comes to understanding digital humanities as an expanding field within the humanities, based on the combination of theory and practice, the question of interdisciplinarity as an overarching and differentiating characteristic has no particular purpose, in the sense that digital humanities is per se interdisciplinary and institutionally diverse and dispersed. As such, it combines the humanities strand with all of its breadth and variety of disciplines, approaches and critical theories, along with the digital technology element and the digital media methods.15 It might be best to focus on the ways these interdisciplinary characteristics are accomplished within the developing field of digital humanities, as well as on the terms and the key notions that link everything together, whether they are in harmony or not. Instead of searching for interdisciplinarity (which is already present), perhaps it would be helpful to look for and examine the experimental and exploring element in the digital humanities,16 the hands-on experience and practice, all within the context of critical theory and thought. These elements, among others, differentiate and particularise digital humanities within the humanities.

Let us now visit the connection between crowdsourcing and history in the context of technologically and digitally networked humanities. What does the notion of crowdsourcing signify in the context of historical studies? The components of the term are two words: crowd and sourcing. Both components, taken in isolation and out of context, might be confusing when combined with history; this is especially so for the word sourcing since it is usually used in the business sector. A crowd is defined as “a large number of people gathered together usually without orderly arrangement”,17 while sourcing is the process of finding suppliers of goods and services (often outside one’s company in order to reduce costs).18 In reality, the majority of the digital collaborative and participatory online history projects being run by cultural institutions can be appealing to many, but not massive crowds; in addition, they are not associated with labour or any form of payment or economic compensation.19 Most often, they intrigue and engage people who show a particular and well-focused interest in a history topic. According to Alexandra Eveleigh, a researcher on the growth of collaborative projects in the archival domain, these projects invite participation from volunteers who “become self-selecting, dropping out when they find the undertaking too difficult . . . this is a leisure-time activity, so transcribers are unlikely to persevere unless they are up for the challenge and able to attain some sense of achievement from their participation.”20 Trevor Owens, a digital archivist at the Library of Congress, suggests “these projects can continue a long standing tradition of volunteerism and involvement of citizens in the creation and continued development of public goods”.21

Up to now, the crowdsourcing boom in the humanities and social sciences concerns mainly transcription projects.22 At its 126th annual meeting in January 2012, the American Historical Associa-
tion dedicated a whole session to crowdsourcing history, the call for which read: “This session seeks to explore the potential and pitfalls of crowdsourcing as a method for collecting transcriptions and for teaching wider audiences about reading and using historical manuscripts. We seek to bring together public historians, archivists, academic historians, technologists, and other scholars to learn about and discuss these projects and the future of crowdsourcing in the historical profession.”

An interesting transcription example is the Transcribe Bentham project, a joint effort of the law faculty and the Centre for Digital Humanities of University College London (UCL). A collaborative online initiative, it invites volunteers to transcribe unedited, unstudied and unpublished manuscripts of the English philosopher and jurist Jeremy Bentham (1748–1832), held at the Bentham Papers Special Collection of UCL’s library, in order to include them in a publication of Bentham’s collected works. Since its launch in 2010, “volunteers from around the world have transcribed several thousand Bentham manuscripts to an extremely high standard.”

The Flickr Commons Project offers another challenging example of how the crowdsourcing method can be applied in the field of digital history. The project was initially launched in January 2008 as a collaborative pilot effort between the Library of Congress and Flickr. The project addresses cultural heritage organisations, in particular libraries, archives and museums (commonly referred to as LAMs in the literature) and has secured the participation of some of the major LAMs worldwide, such as the Library of Congress, the US National Archives, the New York Public Library, Nasa, the Smithsonian Institution, Cornell University Library, the Centre for Jewish History (NYC), the National Archives UK, the National Library of Scotland, the National Archives of Norway, the Royal Library Denmark, LSE Library, the Australian War Memorial Collection, and many others. These participating institutions are considered as content providers. They use Flickr as an open space online to upload and present their image collections (in particular those with no copyright restrictions), along with all the information and metadata already available for them. Registered users, then, are invited to access, view and add information to the images either by leaving a comment or by tagging them. The overall purpose of the project is firstly to increase availability and access to institutional collections and secondly to mobilise participation from the public in an effort to enrich the description of the images with verifiable information as well as to attain a better understanding of the historical context within which these historical items were produced. According to the Library of Congress:

The wealth of interaction and engagement that has taken place within the comments section has resulted in immediate benefits both for the Library and users of the collections. By the end of October 2008, 2,562 unique Flickr accounts had added 7,166 comments on 2,873 photos (and the two set pages). A core group of about 20 “power commenters” returns regularly and provides corrected place names, more precise dates, event names, and fuller names for individuals previously identified only by surname. These Flickr “history detectives” contribute historical information, frequently supporting the information they offer by adding links to the New York Times archive, Wikipedia, and highly specialized Web sites devoted to specific relevant areas of interest such as military aviation, railroads, and sports history.
Apart from those mentioned above, there are other examples of crowdsourcing in digital history. Historical entries in Wikipedia are the most prominent of these examples. According to Robert S. Wolff, "Wikipedia is not simply an online encyclopaedia. Its historical entries serve as virtual ‘sites of memory’ (to borrow from Pierre Nora), places at which people attempt to codify the meaning of past events." While trying to accomplish what Wikipedia sets as a prerequisite, that is a "neutral point of view" (NPOV), people volunteer to write about the past and, in the process, they discuss, debate, argue over history, or at least they try to document, study and narrate it. Some users accept professional scholarship and expertise as valid, some others dismiss it or, to put it mildly, they do not welcome it as the sole and definitive source of knowledge.

As regards history, much has been written for and against Wikipedia. I will only refer to a landmark article by Rosenzweig, published in the Journal of American History in 2006, entitled "Can History be Open-Source? Wikipedia and the Future of the Past." In it, Rosenzweig attempts to address some critical concerns about history on Wikipedia, mostly in relation to the quality and the scholarly standards of the historical work produced there. Among other comparisons, he evaluates 25 biographies of historical figures in Wikipedia against the same entries in Microsoft’s Encarta and in the American National Biography Online. Most of these entries were equally well-researched, documented and written in all three encyclopaedias, apart from a few minor factual errors (wrong dates, for instance). Rosenzweig does not equate Wikipedia and other well-established encyclopaedias in terms of importance; he does, however, suggest that historians need to pay attention to this free and open-source encyclopaedia as well as to discuss the implications that the open and collaborative model of doing history online has for the historical discipline in the digital era.

The encounter of digital history with crowdsourcing practices has opened up a discussion on some theoretical issues and working hypotheses, which could stimulate critical thoughts in the field of digital humanities.

Historical information can be found in an endless abundance on the internet. A vast amount of primary as well as secondary sources has been digitised and is accessible online, removing many of the common difficulties users faced in accessing archival or library collections, such as physical distance, time constraints etc. Digitisation programmes organised by academic institutions, but also user-generated historical content published and circulated online, are changing the way historical information is preserved and disseminated through the networked technologies of the 21st century, and thus form a different and expanding community of knowledge. According to historian Paul Arthur, "Digital history comes in different sizes, from large institutional projects aimed at giving world-wide access to resources for the study of history, for example, right down to personal histories, self-published on YouTube." It is widely suggested that this digital and networked era favours a massive democratisation of historical knowledge and information through free access and open structures. At the same time, there are valid concerns that perhaps another hegemonic model of producing and distributing knowledge is being established, a model that understands the internet as a locus where novel power structures are shaped. As digital humanities evolves, the balance between democratic and hegemonic characteristics remains to be studied and analysed.
Lisa Spiro, director of the US-based National Institute for Technology in Liberal Education Labs, highlights the need to identify a set of basic values for digital humanities. In doing so, she contends that digital humanities is essentially a step towards reappropriating (and not negating) humanities in the digital era:

Even as the humanities continue to make a vital contribution to society, they must be “reassert[ed] and reinterpet[ed]” in a networked age. Whereas the traditional humanities typically value originality, authority, and authorship—an ethos based in part on the scarcity of information and the perceived need for gatekeepers— the Digital Humanities Manifesto instead promotes remixing, openness, and the wisdom of the crowd. For the digital humanities, information is not a commodity to be controlled but a social good to be shared and reused.

Within this context, there are several matters that concern professional historians in relation to participatory and collaborative history projects online, such as authority as well as knowledge expertise and consistency. Apart from the experienced professional scholar, there are countless numbers of people with no or little formal training producing and consuming history online. The disciplinary model of the academic structure of scholarship is challenged by the public engagement with history and by the different popular understandings of what history is and how it is experienced and performed. One could argue that in the digital and networked era of history, people feel the need to seek, discuss and express online their own perceptions of the past and of the way they remember it. This is not the first time that individuals have voiced their own story and memories. Cultural history and oral history in particular have foregrounded this approach, but the majority of relevant projects have been conducted within the framework of academia or other institutional contexts. This trend now seems to be vastly dispersed and performed without any institutional support, by large numbers of internet users of diverse and often unclear characteristics, credentials and affiliations. The strict standard of solid and all-inclusive scholarly expertise has, thus, been reappropriated. Public participation calls for new terms, such as “combined expertise”, to refer to collaboration and joint work by people with diverse skills and varied affiliations dispersed among different disciplines and knowledge fields, and “decentring expertise” addressing issues such as the concepts of openness and endlessness attributed in the process of knowledge and information production and diffusion online.

Does this turn reveal a popular demand for self-historicisation? Does it constitute an alternative historiographical model? Does it indicate another set of mechanisms of cultural knowledge and memory production and consumption in the transition from the twentieth to twenty-first century? In an attempt to think through these questions, it is important to consider the products of digital history as well as study its processes and practices.

Similar debates concerning the diffusion of historical authority and the sharing of historical experience and practice among professionals and amateurs have featured in the past as well, in one way or another. To note just a few examples: in 1998, Rosenzweig spoke about the importance of engaging the audience in making history. Within the theoretical framework of oral and public history, in 1990 Michael Frisch referred to “a shared authority” that is the distribution and sharing characteristics inherent in the process of making history and engaging with the past. One could
argue that this discussion is, in a sense, related to the intellectual environment formed by the new cultural history in the 1990s. In the words of Thomas Gallant:

Social historians believed that we could recapture past peoples’ lived realities, whereas cultural historians argued that all we could do was glimpse at the representations of their lived experiences. The essential positivism that social history shared with the other social sciences was replaced by the idea of contingency. While social history brought understudied groups on to the historical stage, new cultural historians wanted to give them their own. A multivocal chorus of groups telling their own stories replaced the authoritative voice of the historian.

One could also find traces of this discussion in the earlier rise of new social history during the 1960s and 1970s, the historiographical turn towards the study of the experiences of ordinary people, the tendency to rewrite history from below. Much earlier, in 1931, historian Carl Becker, president of the American Historical Association, devoted his annual address to the professionalisation of history. His speech, entitled “Everyman His Own Historian”, challenged the academic conventions and scholarly boundaries that shape the way people think about history and understand the past: “I tried to reduce history to its lowest terms, first by defining it as the memory of things said and done, second by showing concretely how the memory of things said and done is essential to the performance of the simplest acts of daily life.”

And he continued:

It must then be obvious that living history, the ideal series of events that we affirm and hold in memory, since it is so intimately associated with what we are doing and with what we hope to do, can not be precisely the same for all at any given time, or the same for one generation as for another. History in this sense can not be reduced to a verifiable set of statistics or formulated in terms of universally valid mathematical formulas. It is rather an imaginative creation, a personal possession which each one of us, Mr. Everyman, fashions out of his individual experience, adapts to his practical or emotional needs, and adorns as well as may be to suit his aesthetic tastes.

The core argument of Becker’s speech is that the past is never fully known as an objective truth and that history is practiced and experienced by us all, in all aspects of everyday life.

The current rise of public engagement and the multiplication of voices, alongside the single, authoritative professional voice (as a way of doing and experiencing history emerging from an already existing history of historiography and from previous multiple theoretical sequences as well as ruptures), are opening up many possibilities for the theory and practice of history. The overabundance of digitised and particularly digitally born, online primary sources, the emergence of personal histories and memory practices on the internet, as well as the changing and evolving condition of historical information within digital and networked technologies (to mention only a few factors), call for volunteering, participatory and collaborative projects. This process of distributing history in various forms of digital environments and projects provides alternative and diverse ways for the public to explore, interact with, and make sense of the past. In that sense, digital history is no long-
er based on a linear and teleological notion of time; instead it foregrounds a diffused sense of time constantly in flow and reappropriation. It no longer seeks a grand narrative, a whole history, an end and a closure; instead, all the more often it focuses on personal stories, on everyday history, and it promotes openness, unfixity, multiplicity and adaptation. Public history in the era of digital networked technologies is not “about” a topic, a period, a known fact or event; it is “about” us. It is not “for” the people, for a passive and receiving audience; rather, it is “with” the people, acting as dynamic and participating subjects.

Within this context, the notion of crowdsourcing needs to be further studied, rethought and appropriated when used in the context of digital humanities and in particular of digital history, in order to fully realise the possible potential of this convergence. User collaboration and participation are key concepts here. Public engagement and motivation follow along. It seems that, in the digital era, history is becoming all the more public and feels all the more informal, intimate and familiar.

Acknowledgement
This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.
NOTES


7. All information and arguments provided here come from Hockey “The History of Humanities Computing.”


Crowdsourcing digital history online


Owens, “The Crowd and the Library.”

According to Ben W. Brumfield, a software developer specialising in crowdsourced transcription tools, “crowdsourced transcription has been going on for a while, and it’s been happening in four different traditions that developed this all independently. You have genealogists who are doing this, primarily with things like census records. The 1940 census is the most prominent example: they have volunteers transcribing as many as ten million records a day. The natural sciences are doing something similar, particularly GalaxyZoo, the OldWeather people are looking at climate change data, where you have to look at old, handwritten records to figure out how the climate has changed, because you need to know how the climate used to be. And then there are also some projects going on in the Open Source/Creative Commons world: the Wikisource people – particularly the German language Wikisource community – and libraries, archives, and museums have jumped into this recently.” Ben W. Brumfield, “What does it mean to ‘support TEI’ for manuscript transcription?” Collaborative Manuscript Transcription, accessed 19 November 2012, http://manuscripttranscription.blogspot.gr/2012/11/what-does-it-mean-


30 Arthur, “History in Motion,” 3.


33 For this notion see the book Adair et al., Letting Go? Sharing Historical Authority in a User-Generated World, in particular the contributions by Michael Frisch, John Kuo Wei Tsenchen and Liz Ševčenko, Fred Wilson, Paula Marincola and Marjorie Schwarzer, Steve Zeitlin, and Benjamin Filene.


38 For Becker’s work, see Robert A. Hatch, “Historiography – Carl Becker,” The Scientific Revolution Homepage,
Crowdsourcing digital history online


41 For a similar argument in the field of museum studies, see Nina Simon, “Participatory Design and the Future of Museums,” in Adair et al., *Letting Go? Sharing Historical Authority in a User-Generated World*, 18–33. In particular, see p. 21 and the following citation: “Rather than delivering the same content to everyone, a participatory institution collects and shares diverse, personalized, and changing content coproduced with visitors. It invites visitors to respond and add to information about cultural artifacts, scientific evidence, and historical records on display. It showcases the diverse creations and opinions of nonexperts. People use the institution as meeting grounds for dialogue around the content presented. Instead of being ‘about’ something or ‘for’ someone, participatory institutions are created and managed ‘with’ visitors.”