

Crowdsourcing in the Digital Humanities

In Schreibman, S., Siemens, R., and Unsworth, J. (eds), (2016) "A New Companion to Digital Humanities", (p. 420 – 439). Wiley-Blackwell.

<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-1118680596.html>

© Wiley-Blackwell, January 2016. Author's last version provided here with permission.

As Web 2.0 technologies changed the World Wide Web from a read-only to a co-creative digital experience, a range of commercial and non-commercial platforms emerged to allow online users to contribute to discussions and use their knowledge, experience, and time to build online content. Alongside the widespread success of collaboratively produced resources such as Wikipedia came a movement in the cultural and heritage sectors to trial crowdsourcing - the harnessing of online activities and behaviour to aid in large-scale ventures such as tagging, commenting, rating, reviewing, text correcting, and the creation and uploading of content in a methodical, task-based fashion (Holley 2010) - to improve the quality of, and widen access to, online collections. Building on this, within Digital Humanities there have been attempts to crowdsource more complex tasks traditionally assumed to be carried out by academic scholars: such as the accurate transcription of manuscript material.

This chapter aims to survey the growth and uptake of crowdsourcing for culture and heritage, and more specifically, within Digital Humanities. It raises issues of public engagement and asks how the use of technology to involve and engage a wider audience with tasks that have been the traditional purview of academics can broaden the scope and appreciation of humanistic enquiry. Finally, it asks what this increasingly common public-facing activity means for Digital Humanities itself, as the success of these projects demonstrates the effectiveness of building projects for, and involving, a wide online audience.

Crowdsourcing: an introduction

Crowdsourcing – the practice of using contributions from a large online community to undertake a specific task, create content, or gather ideas – is a product of a critical cultural shift in Internet technologies. The first generation of the World Wide Web had been dominated by static websites, facilitated by search engines which only allowed information-seeking behaviour. However, the development of online platforms which allowed and encouraged a two-way dialogue rather than a broadcast mentality fostered public participation, the co-creation of knowledge, and community-building, in a phase which is commonly referred to as “Web 2.0” (O’Reilly 2005, Flew 2008). In 2005, an article in *Wired* Magazine discussed how businesses were beginning to use these new platforms to outsource work to individuals, coining the term “crowdsourcing” as a neologistic portmanteau of “outsourcing” and “crowd” (Howe 2006b):

Technological advances in everything from product design software to digital video cameras are breaking down the cost barriers that once separated amateurs from professionals. Hobbyists, part-timers, and dabblers suddenly have a market for their efforts, as smart companies in industries as disparate as pharmaceuticals and television discover ways to tap the latent talent of the crowd. The labor isn’t always free, but it costs a lot less than paying traditional employees. It’s not outsourcing; it’s crowdsourcing...” (ibid).

The term was quickly adopted online to refer to

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 the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by

sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers (Howe 2006c).

Within a week of the term being coined, 182,000 other websites were using it (Howe 2006a) and it rapidly became the word used to describe a wide range of online activities from contributing to online encyclopedias such as Wikipedia, to tagging images on image sharing websites such as Flickr, to writing on blogs, to proofreading out of copyright texts on Project Gutenberg, or contributing to open-source software. (An analagous term to crowdsourcing, Citizen Science, has also been used where the small-scale tasks carried out online contribute to scientific projects (Silvertown 2009)).

It is important to note here that the use of distributed (generally volunteer) labour to undertake small portions of much larger tasks, gather information, contribute to a larger project, or solve problems, is not new. There is a long history of scientific prizes, architectural competitions, genealogical research, scientific observation and recording, and linguistic study (to name but a few applications) that have relied on the contribution of large numbers of individuals to undertake a centrally managed task, or solve a complex problem (see Finnegan 2005 for an overview). For example, the Mass-Observation Project was a social research organisation in the United Kingdom between 1937 and the 1960s, which relied on a network of 500 volunteer correspondents to record every day life in Britain, including conversation, culture, and behaviour (Hubble, 2006). The difference between these projects and the modern phenomenon of crowdsourcing identified by Howe is, of course, the use of the Internet, the World Wide Web, and interactive web platforms as the mechanism for distributing information, collecting responses, building solutions, and communicating

around a specified task or topic. There was an intermediary phase, however, between offline volunteer labour, and the post-2006 “crowdsourcing” swell, where volunteer labour was used in conjunction with computers and online mechanisms to collect data. Brumfield (2013a) identifies at least seven genealogy projects, such as Free Births, Marriages and Deaths (FreeBMD, <http://freebmd.org.uk/>), Free Registers (FreeREG, <http://www.freereg.org.uk/>) and Free Census (FreeCEN, <http://www.freecen.org.uk/>), that emerged in the 1990s,

out of an (at least) one hundred year old tradition of creating print indexes to manuscript sources which were then published. Once the web came online, the idea of publishing these on the web [instead] became obvious. But the tools that were used to create these were spreadsheets that people would use on their home computers. Then they would put CD ROMs or floppy disks in the posts and send them off to be published online (Brumfield 2013a).

The recent phenomenon of crowdsourcing, or citizen science, can then be seen as a continuation of the use of available platforms and communications networks to distribute tasks amongst large numbers of interested individuals, working towards a common goal.

What types of web-related activities are now described as “crowdsourcing”? Daren Brabham (2013, p.45) proposes a useful typology, looking at the mostly commercial projects which exist in the crowdsourcing space, suggesting that there are two types of problems which can be best solved using this approach: information management issues and ideation problems. Information management issues occur where information needs to be located, created, assembled sorted, or analysed. Brabham suggests that knowledge discovery and management techniques can be used for

crowdsourced information management, as they are ideal for gathering sources or reporting problems: an example of this would be SeeClickFix (<http://en.seeclickfix.com/>) which encourages people to “*Report* neighborhood issues and see them get *fixed*” (SeeClickFix 2013). An alternative crowdsourcing approach to information management is what Brabham calls “distributed human intelligence tasking”: when “a corpus of data is known and the problem is not to produce designs, find information, or develop solutions, but to process data” (Brabham 2013, p.50). The least creative and intellectually demanding of the crowdsourcing techniques, users can be encouraged to undertake repetitive “micro-tasks”, often for monetary compensation, if the task is for a commercial entity. An example of this would be Amazon’s Mechanical Turk (<https://www.mturk.com/>), which “gives businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it’s convenient” (Amazon Mechanical Turk, 2014) – although Amazon Turk has been criticised for its “unethical” business model, with a large proportion of its workers living in third world countries, working on tasks for very little payment (Cushing 2013).

The second type of task that Brabham identified that is suited to crowdsourcing are ideation problems: where creative solutions need to be proposed, that are either empirically true, or a matter of taste or market support (Brabham 2013, p. 48-51). Brabham suggests that crowdsourcing is commonly used as a form of “broadcast search” to locate individuals who can provide the answer to specific problems, or provide the solution to a challenge, sometimes with pecuniary rewards. An example of an online platform using this approach is InnoCentive.com, which is predominantly geared towards the scientific community to generate ideas or reach solutions, for

research and development, sometimes with very large financial prizes: at time of writing, there were three awards worth \$100,000 on offer. Brabham suggests that an alternative crowdsourcing solution to ideation problems is “peer-vetted creative production” (ibid, p.49) where a creative phase is opened up to an online audience, who submit a large number of submissions, and voting mechanisms are then put in place to help sort through the proposals, hoping to identify superior suggestions. An example of this approach would be Threadless.com, a creative community that designs, sorts, creates, and provides a mechanism to purchase various fashion items (the website started with t-shirts, but has since expanded to offer other products).

Since being coined in 2006, the term “crowdsourcing” is now used to cover a wide variety of activities across a large number of sectors: “Businesses, non-profit organizations, and government agencies regularly integrate the creative energies of online communities into day-to-day operations, and many organizations have been built entirely from these arrangements” (Brabham 2013, xv). Brabham’s overall typology is a useful tool as it provides a framework in which to think about both the type of problem that is being addressed by the online platform, and the specific crowdsourcing mechanism that is being used to propose a solution. Given the prevalence of the use of crowdsourcing in online communities for a range of both commercial and not for profit tasks, it is hardly surprising that various implementations of crowdsourcing activities have emerged in the cultural and heritage sector at large, and the Digital Humanities in particular.

The growth of crowdsourcing in cultural and heritage applications

There are many aspects of crowdsourcing that are useful to those working in history, cultural and heritage, particularly within Galleries, Libraries, Archives and Museums (GLAMs), which have a long history of participation with members of the public and generally have institutional aims to promote their collections and engage with as wide an audience as possible. However, “Crowdsourcing is a concept that was invented and defined in the business world and it is important that we recast it and think through what changes when we bring it into cultural heritage” (Owens 2012b). The most obvious difference is that payment to those who undertake tasks is generally not an option for host institutions, but also that “a clearly ethical approach to inviting the public to help in the collection, description, presentation, and use of the cultural record” needs to be identified and pursued (ibid). Owens (2012b) sketches out a range of differences between the mass crowdsourcing model harnessed by the commercial sector and the use of online volunteer labour in cultural and heritage organisations, stressing that “many of the projects that end up falling under the heading of crowdsourcing in libraries, archives and museums have not involved large and massive crowds and they have very little to do with outsourcing labor.” Heritage crowdsourcing projects are not about anonymous masses of people, they are about inviting participation from those who are interested and engaged, and generally involve a small cohort of enthusiasts to use digital tools to contribute (in the same way as they may have volunteered offline to organize and add value to collections in the past). The work is not “labour” but a meaningful way in which individuals can interact with, explore, and understand the historical record. It is often highly motivated and skilled individuals that offer to help, rather than those who can be described with the derogatory term “amateurs”. Owens suggests that crowdsourcing within this sector is then a complex interplay between understanding the potentials for

human computation, adopting tools and software as scaffolding to aid this process, and understanding human motivation (ibid).

No chronological history of the growth of crowdsourcing in culture and heritage exists, but the earliest, large scale project which adopted this model of interaction with users was the Australian Newspaper Digitisation Program (<http://www.nla.gov.au/content/newspaper-digitisation-program>), which in August 2008 asked the general public to correct the OCR (Optical Character Recognition) text of 8.4 million articles generated from their digitised historic Australian Newspapers. This has been a phenomenally successful project, and at time of writing (July 2015), over 166 million individual lines of newspaper articles had been proof read and corrected by volunteer labour. The resulting transcriptions can both aid others in reading, but also in finding, text in the digitised archive. After the success of this project, and the rise of commercial crowdsourcing, other projects began to adopt crowdsourcing techniques to help digitise, sort, and correct heritage materials. In 2009 One of the earliest citizen science projects that is based on historical data, the North American Bird Phenology Program (<http://www.pwrc.usgs.gov/bpp/>) was launched to transcribe 6 million migration card observations collected by a network of volunteers “who recorded information of first arrival dates, maximum abundance, and departure dates of migratory birds across North America” between 1880 and 1970 (North American Bird Phenology Program, n. d). At time of writing, over a million cards have been transcribed by volunteers since launch, allowing a range of scientific research to be carried out on the resulting data (ibid).

Crowdsourcing in the heritage sector began to gather speed around 2010 with a range of projects being launched that asked the general public for various types of help via an online interface. One of the most successful of these is another combination of historical crowdsourcing, and citizen science, called Old Weather (<http://www.oldweather.org/>) which invites the general public to transcribe weather observations that were noted in ship's log books dating from the mid 19th Century to the present day in order to "contribute to climate model projections and ...improve our knowledge of past environmental conditions" (Old Weather 2013a). Old Weather launched in October 2010 as part of the Zooniverse (<http://www.zooniverse.org/>) portal of fifteen different citizen science projects (which had started with the popular gallery classification tool, Galaxy Zoo (<http://www.galaxyzoo.org/>), in 2009). The Old Weather project is a collaboration of a diverse range of archival and scientific institutions and museums and universities in both the UK and the USA (Old Weather 2013b), showing how a common digital platform can bring together physically dispersed information for analysis by users. At time of writing, over 34,000 logs and seven voyages have been transcribed (three times, by different users to insure quality control, meaning that over 1,000,000 individual pages have been transcribed by users (Brohan, P. 2012)), and the resulting data is now being used by both scientists and historians to understand both climate patterns and naval history (with their blog regularly updated with findings: <http://blog.oldweather.org/>).

A range of other notable crowdsourcing projects launched in the 2010 to 2011 period, showing the breadth and scope of the application of online effort to cultural heritage. These include (but are not limited to): Transcribe Bentham, which aims to transcribe the writings of the philosopher and jurist Jeremy Bentham

(<http://blogs.ucl.ac.uk/transcribe-bentham/>); the Victoria and Albert Museum's tool to get users to improve the cropping of their photos in the collection (<http://collections.vam.ac.uk/crowdsourcing/>); The United States Holocaust Museum's "Remember Me" project which aims to identify children in photographs taken by relief workers during the immediate aftermath of the second World War, to facilitate connections amongst survivors (<http://rememberme.ushmm.org/>); New York Public Library's "What's on the menu?" project (<http://menus.nypl.org/>), in which users can transcribe their collection of historical restaurant menus; and the National Library of Finland's DigitalKoot project (http://www.digitalkoot.fi/index_en.html) which allowed users to play games which helped improve the metadata of their Historical Newspaper Library. The range and spread of websites that come under the crowdsourcing umbrella in the cultural and heritage sector continues to increase, and it is now a relatively established, if evolving, method used for galleries, libraries, archives and museums. A list of non-profit crowdsourcing projects in GLAM institutions is maintained at <http://www.digitalglam.org/crowdsourcing/projects/>. Considering this activity in light of Brabham's typology, above, it is clear that most projects fall into the "information management" category (Brabham 2013), where an organisation (or collaborative project between a range of organisations) tasks the crowd with helping to gather, organise, and collect information into a common source or format.

What is the relationship of these projects to those working in Digital Humanities? Obviously, many crowd-sourcing projects depend on having information – or things – to comment on, transcribe, analyse, or sort, and therefore GLAM institutions, who are custodians of such historical material, often partner with University researchers who

have an interest in using digital techniques to answer their Humanities or Heritage based research question. There is often much sharing of expertise and technical infrastructure between different projects and institutions: for example, the Galaxy Zoo platform which underpins Old Weather also is used by Ancient Lives (<http://ancientlives.org/>) to help crowdsource transcription of papyri, and Operation War Diary (<http://www.operationwardiary.org/>) to help transcribe First World War Unit Diaries. Furthermore, those working in Digital Humanities can often advise and assist colleagues in partner institutions and scholarly departments: the Transcribe Bentham project is a collaboration between University College London's Library Services (including UCL's Special Collections), The Bentham Project (based in the Faculty of Laws), UCL Centre for Digital Humanities, The British Library, and The University of London Computing Centre, with the role of the Digital Humanities centre being to provide guidance and advice with online activities, best practice, and public engagement. Another example of collaboration can be seen in events such as the CITSCribe Hackathon in December 2013, which "brought together over 30 programmers and researchers from the areas of biodiversity research and digital humanities for a week to further enable public participation in the transcription of biodiversity specimen labels" (iDigBio 2013). Crowdsourcing in the Digital Humanities can also be used to sort and improve incomplete data sets, such as a corpus of 493 non-Shakespearean plays written between 1576 and 1642 in which 32,000 partially transcribed words were corrected by students over the course of an eight week period using an online tool (<http://annolex.at.northwestern.edu>, see Mueller 2014), indicating how we can use crowdsourcing to involve Humanities students in the gathering and curating of corpora relevant to the wider Humanities community. Scholars in the Digital Humanities are well placed to research, scope and

theorise crowdsourcing activities across a wider sector: for example, the “Modeling Crowdsourcing for Cultural Heritage” project (<http://cdh.uva.nl/projects-2013-2014/m.o.c.c.a.html>) based at the Centre for Digital Humanities and the Creative Research Industries Amsterdam, both at the University of Amsterdam, is aiming to determine a comprehensive model for “determining which types and methods of crowdsourcing are relevant for which specific purposes” (Amsterdam Centre for Digital Humanities 2013). As we shall see, below, Digital Humanities scholars and centres are investigating and building new platforms for crowdsourcing activities – particularly in the transcription of historical texts. In addition, Digital Humanities academics can help with suggestions on what we can do with crowdsourced information once collected: we are now moving into a next phase of crowdsourcing, where understanding data mining and visualisation techniques to query the volume of data collected by volunteer labour is necessary. Finally, there is the beginnings of a body of literature on the wider area of crowdsourcing, both across the Digital Humanities and the GLAM sector, and taken together these can inform those who are contemplating undertaking a crowdsourcing project for a related area. It should be stressed that it is often hard to make a distinction between what should be labelled a “GLAM sector” project and what should be labelled “Digital Humanities” in the area of crowdsourcing, as many projects are using crowdsourcing not only to sort or label or format historical information, but to provide the raw materials and methodologies for creating and understanding novel information about our past, our cultural inheritance, or our society.

Following on from the success of the Australian Newspapers Digitisation Program which she managed, Holley (2010) brought issues of “Crowdsourcing: How and Why

Should Libraries Do it” to light, in a seminal discussion that much subsequent research and project implementation has benefited from. Holley proposes that there are a variety of potential benefits in using crowdsourcing within a library context (which we can also extrapolate to cover those working across the GLAM sector, and Digital Humanities). The benefits of crowdsourcing noted are that it can help to: achieve goals the institution would not have the resources (temporal, financial, or staffing) to accomplish itself; achieve these goals quicker than if working alone; build new user groups and communities; actively engage the community with the institution and its systems and collections; utilise external knowledge, expertise and interest; improve the quality of data which improves subsequent user search experiences; add value to data; improving and expanding the ways in which data can be discovered; gain an insight into user opinions and desires by building up a relationship with the crowd; show the relevance and importance of the institution (and its collections) by the high level of public interest in the project; build trust and encouraging loyalty to the institution; and encourage a sense of public ownership and responsibility towards cultural heritage collections (ibid).

Holley also asks what the normal profile of a crowdsourcing volunteer in the cultural, heritage, and humanities sector is, stressing that from even early pilot projects the same makeup emerges: although there may be a large number of volunteers who originally sign up, the majority of the work is done by a small cohort of super users, which achieve significantly larger amounts of work than anyone else. They tend to be committed to the project for the long term, appreciate that it is a learning experience, which gives them purpose and is personally rewarding, perhaps because they are interested in it, or see it as a good cause. Volunteers often talk of becoming addicted

to the activities, and the amount of work undertaken often exceeds the expectations of the project (ibid). Holley argues that “the factors that motivate digital volunteers are really no different to factors that motivate anyone to do anything” (ibid), saying that interest, passion, a worthy cause, giving back to the community, helping to achieve a group goal, and contributing to the discovery of new information in an important area, are often reasons that volunteers contribute. Observations and surveys of volunteers by site managers noted various techniques that can improve user motivation, such as adding more content regularly, increasing challenges, creating a camaraderie, building relationships with the project, acknowledging the volunteer’s help, providing rewards, and making goals and progress transparent. The reward and acknowledgement process is often linked to progress reports, with volunteers being named, high achievers being ranked in publicly available tables, and promotional gifts (ibid).

Holley provides various tips that have provided guidance for a variety of crowdsourcing projects, and are worth following by those considering using this method. The project should have a clear goal, a big challenge, report regularly on progress, and showcase results. The system should be easy and fun, reliable and quick, intuitive, and provide options to the user so they can choose what they work on (to a certain extent). The volunteers should be acknowledged, be rewarded, be supported by the project team, and be trusted. The content should be interesting, novel, focussed on history or science, and there should be lots of it (ibid).

Holley’s paper was written just before many of the projects outlined above came on-stream, stressing the potential for institutions, and challenging institutional structures to be brave enough to attempt to engage individuals in this manner. By 2012, with

various projects in full swing, reports and papers began to appear about the nuances of crowdsourcing in this area, although “there is relatively little academic literature dealing with its application and outcomes to allow any firm judgements to be made about its potential to produce academically credible knowledge” (Hedges and Dunn 2012, p.4).

Ridge (2012) explores the “Frequently Asked Questions about Crowdsourcing in Cultural Heritage”, noting various misconceptions and apprehensions surrounding the topic. As with Owens (2012a), Ridge agrees that the industry definition of crowdsourcing is problematic, suggesting instead that it should be defined as

an emerging form of engagement with cultural heritage that contributes towards a shared, significant goal or research area by asking the public to undertake tasks that cannot be done automatically, in an environment where the tasks, goals (or both) provide inherent rewards for participation” (Ridge 2012).

Ridge draws attention to the importance of the relationships built between individuals and organisations, and that projects should be mindful of the motivations for participating. Institutional nervousness around crowdsourcing is caused by worries that malicious or deliberately bad information will be provided by difficult, obstructive users, although Ridge maintains this is seldom the case, and that a good crowdsourcing project should have inbuilt mechanisms to highlight problematic data or users, and validate the content created by its users. Ridge returns again to the ethics of using volunteer labour, allaying fears about the type of exploitation seen in the commercial sector exploitation by explaining that

Museums, galleries, libraries, archives and academic projects are in the fortunate position of having interesting work that involves an element of social good, and they also have hugely varied work, from microtasks to co-curated research projects. Crowdsourcing is part of a long tradition of volunteering and altruistic participation (Ridge 2012).

In a further 2013 post, Ridge also highlights the advantages of digital engagement via crowdsourcing, suggesting that digital platforms can allow smaller institutions to engage with users just as well as large institutions can, can generate new relationships with different organisations in order to work together around a similar topic in a collaborative project, and can provide great potential for audience participation and engagement (Ridge 2013). In fact, Owens (2012a) suggests that our thinking around crowdsourcing in cultural and heritage is the wrong way round: rather than thinking of the end product and the better data that volunteers are helping us create, institutions should focus on the fact that crowdsourcing marks a fulfilment of the mission of putting digital collections online:

What crowdsourcing does, that most digital collection platforms fail to do, is offers an opportunity for someone to do something more than consume information... Far from being an instrument which enables us to ultimately better deliver content to end users, crowdsourcing is the best way to actually engage our users in the fundamental reason that these digital collections exist in the first place... At its best, crowdsourcing is not about getting someone to do work for you, it is about offering your users the opportunity to participate in public memory (ibid).

The lessons learned from these museum and library based projects are important starting points for those in the Digital Humanities who wish to undertake crowdsourcing themselves.

Crowdsourcing and Digital Humanities

In a 2012 scoping study of the use of crowdsourcing particularly applied to Humanities research, 54 academic publications were identified that were of direct relevance to the field, and a further 51 individual projects, activities or websites were found which documented or presented some aspect, application, or use of crowdsourcing within humanities scholarship (Hedges and Dunn 2012). Many of these projects have crossovers with libraries, archives, museums, and galleries, as partners who provide content, expertise, or host project themselves, and many of them are yet to produce a tangible academic outcome. As Hedges and Dunn point out, at

a time when the web is simultaneously transforming the way in which people collaborate and communicate, and merging the spaces which the academic and non-academic communities inhabit, it has never been more important to consider the role which public communities -connected or otherwise - have come to play in academic humanities research (ibid, p. 3).

Hedges and Dunn (ibid, p.7) identify four factors that define crowd-sourcing used within humanities research. These are: a clearly defined core research question and direction within the humanities; the potential for an online group to add to, transform, or interpret data that is important to the humanities; a definable task which is broken down into an achievable workflow; and the setting up of a scalable activity which can be undertaken with different levels of participation. Very similar to the work done in the GLAM sector, the theme and research question of the project are therefore the

main distinguishing factors from other types of crowdsourcing, with Digital Humanities projects learning from other domains such as successful projects in citizen science, or industry.

An example of such a project fitting into this Humanities Crowdsourcing definition, given its purview, is Transcribe Bentham (<http://blogs.ucl.ac.uk/transcribe-bentham/>), a manuscript transcription initiative that intends to engage students, researchers, and the general public with the thought and life of the philosopher and reformer, Jeremy Bentham (1748–1832), by making available digital images of his manuscripts for anyone, anywhere in the world, to transcribe. The fundamental research question driving this project is to understand the thought and writings of Bentham more completely – a topic of fundamental importance to those engaged in eighteenth or nineteenth century studies – given that 40,000 folios of his writings remain untranscribed “and their contents largely unknown, rendering our understanding of Bentham’s thought—together with its historical significance and continuing philosophical importance—at best provisional, and at worst a caricature.” (Causer and Terras, Forthcoming 2014). The objectives of the project are clear, with the benefit to humanities (and law, and social science) research evident from the research objectives.

Hedges and Dunn (2012, p.18 -19) list the types of knowledge that may be usefully created in Digital Humanities crowdsourcing activities, resulting in new understanding of Humanities research questions. These Digital Humanities crowdsourcing projects are involved in: making ephemera available that would otherwise not be; opening up information that would normally be accessible to distinct groups, giving a wider audience to specific information held in little known

written documentation, circulation of personal histories and diaries, giving personal links to historical processes and events, identifying links between objects, summarising and circulating datasets, synthesizing new data from existing sources, and recording ephemeral knowledge before it dissipates. Hedges and Dunn stress that an important point in these crowdsourcing projects is that they enable the building up of knowledge of the process of how to conduct collaborative research in this area, whilst creating communities with a shared purpose, which often carry out research work that go beyond the expectations of the project (p.19). However, they are keen to also point out that

most humanities scholars who have used crowd-sourcing in its various forms now agree that it is not simply a form of cheap labour for the creation or digitization of content; indeed in a cost-benefit sense it does not always compare well with more conventional means of digitization and processing.

In this sense, it has truly left its roots, as defined by Howe (2006) behind. The creativity, enthusiasm and alternative foci that communities outside that academy can bring to academic projects is a resource which is now ripe for tapping in to (ibid, p. 40).

As with Owens' thoughts on crowdsourcing in the GLAM sector (2012), we can see that crowdsourcing in the humanities is about engagement, and encouraging a wide, and different audience to engage in processes of humanistic enquiry, rather than merely being a cheap way to encourage people to get a necessary job done.

Crowdsourcing and Document Transcription

The most high profile area of crowdsourcing carried out within the humanities is in the area of document transcription. Although commercial optical character

recognition (OCR) technology has been available for over 50 years (Schantz 1982), it still cannot generate high quality transcripts of handwritten material. Work with texts and textual data is still the major topic of most Digital Humanities research (see the analysis by Scott Weingart of submissions to the Digital Humanities Conference 2014, which showed that of the 600 abstracts, 21.5% dealt with some form of Text Analysis, 19% were about literary studies, and 19% were about Text Mining (Weingart, S. 2013)). It is therefore no surprise that most Digital Humanities crowdsourcing activities – or at least, those emanating from Digital Humanities centres and or associated in some sense with the Digital Humanities community - have been involved in the creation of tools in which to help transcribe important handwritten documents into machine processable form.

Ben Brumfield, in a talk presented in 2013, demonstrated that there were thirty collaborative transcription tools developed since 2005 (Brumfield 2013a), situating the genealogical sites, and those such as Old Weather and Transcribe Bentham, in a trajectory which leads to the creation of tools and platforms which people can use to upload their own documents, and manage their own crowdsourcing projects (reviews of these different platforms are available on Brumfield's blog, at <http://manuscripttranscription.blogspot.co.uk/>, and at time of writing there are now forty collaborative tools for crowdsourcing document transcription). The first of these customizable tools was Scripto (<http://scripto.org/>), a freely available, open source platform for community transcription, developed in 2011 by the Center for History and New Media at George Mason University alongside their Papers of the United States War Department project (<http://wardepartmentpapers.org/>). Another web based tool, specifically designed for Transcription for Paleographical and Editorial Notation

(T-PEN) (<http://t-pen.org/TPEN/>), coordinated by the Center for Digital Theology at Saint-Louis University, provides a web based interface for working with images of manuscripts. Transcribe Bentham has also released a customizable, open source version of its Mediawiki based platform (<https://github.com/onthimagen/cbp-transcription-desk>), which has since been used by the Public Record Office of Victoria, Australia (http://wiki.prov.vic.gov.au/index.php/Category:PROV_Transcription_Pilot_Project). The toolbar developed for Transcribe Bentham, which helps people encode various aspects of transcription such as dates, people, deletions, etc, has been integrated into the Letters of 1916 project at Trinity College Dublin (<http://dh.tcd.ie/letters1916/>). The platform the Letters of 1916 project uses is the DIYHistory suite, built by the University of Iowa, which itself is based on CHNM's Scripto tool. Links between crowdsourcing projects are common.

There are now a range of transcription projects online ranging from those created, hosted, and managed by scholarly or memory institutions, to those entirely organised by amateurs with no scholarly training or association. A prime example of the latter would be Soldier Studies, (<http://www.soldierstudies.org/>), a website dedicated to preserving the content of American Civil War correspondence bought and sold on eBay, to allow access to the contents of this ephemera before it resides in private collections, which, although laudable, uses no transcription conventions at all in cataloguing or transcribing the documents it finds (Brumfield 2013a).

The movement towards collaborative online document transcription by volunteers not only uncovers new, important historical primary source material, but it also “can open

up activities that were traditionally viewed as academic endeavours to a wider audience interested in history” (Causser and Terras, Forthcoming 2014). Brumfield (2013) points out that there are issues which come with this:

There's an institutional tension, in that editing of documents has historically been done by professionals, and amateur editions have very bad reputations. Well now we're asking volunteers to transcribe. And there's a big tension between, well how do volunteers deal with this [process], do we trust volunteers? Wouldn't it be better just to give us more money to hire more professionals? So there's a tension there.

Brumfield further explores this in another blog post (2013b) where he asks

what is the qualitative difference between the activities we ask amateurs to do and the activities performed by scholars... we're not asking "citizen scholars" to do real scholarly work, and then labeling their activity scholarship -- a concern I share with regard to editing. If most crowdsourcing projects ask amateurs to do little more than wash test tubes, where are the projects that solicit scholarly interpretation?

There is therefore a fear that without adequate guidance and moderation, the products of crowdsourced transcription will be what Shillingsburg referred to as “a dank cellar of electronic texts” where “the world is overwhelmed by texts of unknown provenance, with unknown corruptions, representing unidentified or misidentified versions” (2006, p.139). Brumfield (2013c) points out that Peter Robinson describes both the utopia and the dystopia of crowdsourcing transcription: utopia in which textual scholars train the world in how to read documents, and a dystopia in which hordes of “well-meaning but ill-informed enthusiasts will strew the web willy-nilly

with error-filled transcripts and annotations, burying good scholarship in rubbish.” (Robinson, quoted in Brumfield 2013c). To avoid this, Brumfield (2013c) suggests that partnerships and dialogue between volunteers and professionals is essential, to make methodologies for approaching texts visible, and to allow volunteers to become advocates “not just for the material and the materials they are working on through crowdsourcing project, but for editing as a discipline” (ibid).

Care needs to be taken, then, when setting up a crowdsourcing transcription project, to ensure that the quality of the resulting transcription is suitable to be used as the basis for further scholarly humanistic enquiry, if the project is to be useful over a longer term and for a variety of research. The methods and approaches in assuring transcription quality of *content* need to be ascertained: whether the project uses double-keying (where two or more people enter the same text to ensure its veracity), or moderation (where an expert in the field signs off the text into a database, agreeing that its content meets benchmarked standards). However, in addition to this the *format* that the data is stored in needs to be structured to ensure that complex representational issues are preserved, and that any resulting data created can be easily reused and textual models can be understood, repurposed, or integrated with other collections. As Brumfield points out (2013a) Digital Humanities already has a standard for documentary scholarly editing in the Text Encoding Initiative guidelines (TEI 2014), which have been available since 1990 and provide a flexible but robust framework within which to model, analyse, and present textual data. However, only seven of the crowdsourcing manuscript transcription tools (out of the thirty then available) attempt to integrate TEI compliant XML encoding into their workflow (Brumfield 2013a). Projects which have used TEI markup as part of the manuscript transcription process,

such as Transcribe Bentham, have demonstrated that users can easily learn the processes of encoding texts with XML if clear guidance and instruction is given to them, and it is explained why they should make the effort to do it (Brumfield 2013a, Causer and Terras Forthcoming 2014). Brumfield (2013a) stresses that it is the responsibility of those involved in academic scholarly editing within the Digital Humanities to ensure that their work on establishing methods and guidelines for academic transcription is felt within the development of public facing transcription tools, and if we are engaging users so that they can build their own skillsets, we need to use our digital platforms to train them according to pedagogical and scholarly standards: “Crowdsourcing is a school. Programs are the teachers. We have to get it right” (Brumfield 2013d). Brumfield (2013c) also highlights that it is the responsibility of those working in document editing, and the Digital Humanities, to release guides to editing and transcribing that are accessible to those with no academic training in this area, such as computer programmers building transcription tools, if we wish for the resulting interfaces to allow community-led transcription to result in high quality textual material.

Future Issues in Digital Humanities Crowdsourcing

We are now at a stage where crowdsourcing has joined the ranks of established digital methods for gathering and classifying data for use in answering the types of questions of interest to Humanities scholars, although there is much research that still needs to be done about user response to crowdsourcing requests, and how best to build and deliver projects. There are also issues about data management, given that crowdsourcing is now reaching a mature phase where a variety of successful projects have amassed large amounts of data, often from different sources within individual

projects: the million pages from Old Weather from different archives; over 3 million words transcribed by volunteer labour in the Transcribe Bentham project (Grint 2013) from both UCL and the British Library; approximately one and a half thousand letters transcribed in Soldier Studies (Soldier Studies 2014), which at a conservative estimate must give at least half a million words of correspondence from the American Civil War, culled from images of letters sold on eBay which are now in private hands. Issues are therefore arising about sustainability: what will happen to all this data, particularly with regard to projects that do not have institutional resources or affiliation for long term backup or storage? There are also future research avenues to investigate cross-project sharing and amalgamation of data: one can easily imagine either centrally managed or federated repositories of crowdsourced information that contain all the personal diaries that have been transcribed, searchable by date, place, person, etc; or all letters and correspondence that have been sent over time, or all newspapers that were issued on a certain date worldwide. Both legal and technical issues will come in to play with this, as questions of licensing (who owns the volunteer created data? Who does the copyright belong to?) and cross-repository searching will have to be negotiated, with related costs for delivering mechanisms and platforms covered. The question of the ethics of crowdsourcing is one that also underlies much of this effort in the Humanities and the cultural and heritage sector, and projects have to be careful to work with volunteers, rather than exploit them, when building up these repositories and reusing and repurposing data in the future. Ethical issues come sharply into focus when projects start to pay (usually very little) for the labour involved, particularly when using online crowdsourcing labour brokers such as Amazon's Mechanical Turk (<https://www.mturk.com/mturk/welcome>), which has been criticised as a

“digital sweatshop... critics have emerged from all corners of the labor, law, and tech communities. Labor activists have decried it as an unconscionable abuse of workers’ rights, lawyers have questioned its legal validity, and academics and other observers have probed its implications for the future of work and of technology” (Cushing 2013).

The relationship between commerce and volunteers, payment and cultural heritage, resources and outputs, online culture and the online workforce, is complex. A project such as “Emoji-Dick” (<https://www.kickstarter.com/projects/fred/emoji-dick>) - which translated Moby Dick into Japanese Emoji Icons using Amazon's Mechanical Turk – is a prime example of what emerges when the lines of public engagement, culture, art, fun, low-paid crowdsourced labour, crowdfunding, and an internet meme, collide. Institutions and scholars planning on tapping into the potential labour force crowdsourcing offers have to be aware of the problems in outsourcing such labour, often very cheaply, to low paid workers, often in third world countries (Cushing 2013).

Returning to Brabham’s typology on crowdsourcing projects, we can also see that although most projects that have used crowdsourcing in the Humanities are information management tasks in that they ask volunteers to help enter, collate, sort, organise, and format information, there is also the possibility that crowdsourcing can be used within the Humanities for ideation tasks: asking big questions, and proposing solutions. This area is undocumented within Digital Humanities, although the Association for Computers and the Humanities (ACH), and the 4Humanities.org initiative, have both used an open source platform, All Our Ideas (<http://www.allourideas.org/>) to help scope out future initiatives (ACH 2012,

4Humanities 2012). ACH also host and support DH Questions and Answers (<http://digitalhumanities.org/answers/>), a successful community based questions and answers board for Digital Humanities issues, which falls within the ideation category of crowdsourcing. There is much scope within the Humanities in general to explore this methodology and ideation mechanism further, and to engage the crowd in both proposing, and solving, questions about the Humanities, rather than only using it to self organise Digital Humanities initiatives.

Crowdfunding is another relatively new area allied to crowdsourcing, which could be of great future benefit to Digital Humanities, and Humanities projects in general. Only a few projects have been started to date within the GLAM sector, both for traditional collections acquisition and for digital projects: The British Library is attempting to crowdfund for the digitisation of historical London maps (British Library 2014); The Naturalis Biodiversity Centre in Leiden is raising funds via crowdfunding to purchase a Tyrannosaurus Rex skeleton (<http://tientjevoortrex.naturalis.nl/>), The Archiefbank or the Stadsarchief Amsterdam has raised 30,000 euros to digitise and catalogue the Amsterdam death registers between 1892 and 1920 (Stadsarchief Amsterdam 2012), and a campaign to crowdfund the £520,000 needed to buy the cottage on the Sussex coast where William Blake wrote “England’s green and pleasant land” was launched at time of writing (Flood 2014). A project, Micropasts (<http://micropasts.org/>) recently funded by the UK’s Arts and Humanities Research Council based at University College London and the British Museum, will be developing a community platform for conducting, designing and funding research into the human past: over the next few years this will

be an area which has much potential for involving those outside the academy with core issues within Humanities scholarship.

Crowdsourcing also offers a relatively agile mechanism for those working in Digital Humanities to respond immediately to important contemporary events, preserving and collating evidence, ephemera, and archive material for future scholarship, and community use. For example the September 11th Digital Archive (<http://911digitalarchive.org>) which “uses electronic media to collect, preserve, and present the history of the September 11, 2001 attacks in New York, Virginia, and Pennsylvania and the public responses to them” (September 11 Digital Archive) began as a collaboration American Social History Project at the City University of New York Graduate Center, and the Center for History and New Media at George Mason University, immediately after the terrorist attacks. Likewise, the Our Marathon Archive (<http://marathon.neu.edu/>), led by Northeastern University, provides an archival and community space to crowdsource an archive of “pictures, videos, stories, and even social media related to the Boston Marathon; the bombing on April 15, 2013; the subsequent search, capture, and trial of the individuals who planted the bombs; and the city’s healing process” (Our Marathon 2013). There is clearly a role here for those within the Digital Humanities with technical and archival expertise to respond to contemporary events by building digital platforms that will both keep records for the future, whilst engaging with a community – and often a society - in need of sustained dialogue to process the ramifications of such events.

There is also potential for more sustained and careful use of crowdsourcing within both the university and school classroom, to promote and integrate on going

Humanities research aims, but also to “meet essential learning outcomes of liberal education like gaining knowledge of culture, global engagement, and applied learning” (Frost Davis 2012). There are opportunities for motivated students to become more involved and engaged with projects that digitize, preserve, study, and analyse resources, encouraging them to gain first hand knowledge of humanities issues and methods, but also to understand the role that digital methods can play in public engagement:

Essential learning outcomes aim at producing students with transferrable skills; in the globally networked world, being able to produce knowledge in and with the network is a vital skill for students. Students also benefit from exposure to how experts approach a project. While these tasks may seem basic, they lay the groundwork for developing deeper expertise with practice so that participation in crowdsourcing projects may be the beginning of a pipeline that leads students on to more sophisticated digital humanities research projects. Even if students don't go on to become digital humanists, crowdsourced projects can help them develop a habit of engagement with the (digital) humanities, something that is just as important for the survival of the humanities. Indeed, a major motivation for humanities crowdsourcing is that involving the public in a project increases public support for that project (Frost Davis 2012).

Crowdsourcing within the Humanities will then continue to evolve, and offers much scope for using public interest in the past to bring together data and build projects which can benefit Humanities research:

Public involvement in the humanities can take many forms – transcribing handwritten text into digital form; tagging photographs to facilitate discovery and preservation; entering structured or semi-structured data; commenting on content or participating in discussions, or recording one’s own experiences and memories in the form of oral history – and the relationship between the public and the humanities is convoluted and poorly understood (Hedges and Dunn 2012, p.4).

By systematically applying, building, evaluating, and understanding the uses of crowdsourcing within culture, heritage and the humanities, by helping develop the standards and mechanisms to do so, and by ensuring that the data created will be useable for future scholarship, the Digital Humanities can aid in creating stronger links with the public and humanities research, which, in turn, means that crowdsourcing becomes a method of advocacy for the importance of humanities scholarship, involving and integrating non-academic sectors of society into areas of humanistic endeavour.

Conclusion

This chapter has surveyed the phenomenon of using digital crowdsourcing activities to further our understanding of culture, heritage and history, rather than simply identifying the activities of digital humanities centres, or self identified digital humanities scholars, which do so. This in itself is an important discussion to have about the nature of Digital Humanities research, its home, and its purview. Much of the crowdsourcing activity identified in the GLAM sector comfortably fits under the Digital Humanities umbrella, even if those involved did not self-identify with that classification: there is a distinction to be made between projects which operate within

the type of area which is of interest to Digital Humanities, and those run by Digital Humanities centres and scholars.

With that in mind, this chapter has highlighted various ways in which those working in Digital Humanities can help advise, create, build, and steer crowdsourcing projects working in the area of culture and heritage to both add to our understanding of crowdsourcing as a methodology for humanities research, and to build up resulting datasets which will allow further humanities research questions to be answered. Given the current pace of development in the area of crowdsourcing within this sector, there is much that can be contributed from the Digital Humanities community to ensure that the resulting methods and datasets are useful, and reusable, particularly within the arena of document transcription and encoding. In addition, crowdsourcing affords vast opportunities for those working within the Digital Humanities to provide accessible demonstrators of the kind of digital tools and projects which are able to forward our understanding of culture and history, and also offers outreach and public engagement opportunities to show that Humanities research, in its widest sense, is a relevant and important part of the scholarly canon to as wide an audience as possible. In many ways, crowdsourcing within the cultural and heritage sectors is Digital Humanities writ large: indicating 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.

Short Biographical Note

Melissa Terras is Director of UCL Centre for Digital Humanities, Professor of Digital Humanities in UCL's Department of Information Studies and Co-Investigator of the

award winning Transcribe Bentham crowdsourcing project (www.ucl.ac.uk/transcribe-bentham). Her research spans various aspects of digitisation and public engagement. You can generally find her on twitter @melissaterras.

Abstract

A recent movement in the cultural and heritage industries has been to trial crowdsourcing (the harnessing of online activities and behaviour to aid in large-scale ventures such as tagging, commenting, rating, reviewing, text correcting, and the creation and uploading of content in a methodical, task-based fashion) to improve the quality of, and widen access to, online collections. Building on this, within Digital Humanities there have been attempts to crowdsource more complex tasks traditionally assumed to be carried out by academic scholars, such as the accurate transcription of manuscript material. This chapter surveys the growth and uptake of crowdsourcing within Digital Humanities, raising issues which emerge when building projects for and with a wide online audience.

Keywords

Crowdsourcing, public engagement, digitisation, online participation, citizen science.

Further Reading

Brabham, D. C. (2013). *Crowdsourcing*. MIT Press Essential Knowledge Series. London, England, MIT Press.

Brumfield, B. (2013a). *Itinera Nova in the World(s) of Crowdsourcing and TEI*. Collaborative Manuscript Transcription Blog. <http://manuscripttranscription.blogspot.co.uk/2013/04/itinera-nova-in-worlds-of-crowdsourcing.html>. Accessed 17th January 2014.

Brumfield, B. (2013c). The Collaborative Future of Amateur Editions. Collaborative Manuscript Transcription Blog, <http://manuscripttranscription.blogspot.co.uk/2013/07/the-collaborative-future-of-amateur.html>. Accessed 28th January 2014.

Causer, T. and Terras, M. (Forthcoming 2014) "Crowdsourcing Bentham: beyond the traditional boundaries of academic history". Accepted, International Journal of Humanities and Arts Computing.

Flood, A. (2014). "Crowdfunding campaign hopes to save William Blake's cottage for nation". Guardian, 11st September 2014, <http://www.theguardian.com/culture/2014/sep/11/crowdfunding-campaign-william-blake-cottage>

Frost Davis, R. (2012). "Crowdsourcing, Undergraduates, and Digital Humanities Projects". <http://rebeccafrostdavis.wordpress.com/2012/09/03/crowdsourcing-undergraduates-and-digital-humanities-projects/>. Accessed 29th January 2014.

Hedges, M. and Dunn, S. (2012). Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research. Arts and Humanities Research Council. <http://crowds.cerch.kcl.ac.uk/>, Accessed 16th January 2014.

Holley, R. (2010). Crowdsourcing: How and Why Should Libraries Do It?, *D-Lib Magazine*, 16 (2010), <http://www.dlib.org/dlib/march10/holley/03holley.html>. Accessed 17th January 2014.

Owens, T. (2012b). The Crowd and The Library. <http://www.trevorowens.org/2012/05/the-crowd-and-the-library/>. Accessed 16th January 2014.

Ridge, M. (2012). Frequently Asked Questions about crowdsourcing in cultural heritage. Open Objects blog. <http://openobjects.blogspot.co.uk/2012/06/frequently-asked-questions-about.html>. Accessed 18th January 2014.

Bibliography

ACH (2014). ACH Agenda Setting: Next Steps. Association for Computers and the Humanities Blog, <http://ach.org/2012/06/04/ach-agenda-setting-next-steps/>. Accessed 29th January 2014.

Amazon Mechanical Turk, (2014). Amazon Mechanical Turk, Welcome. <https://www.mturk.com/mturk/welcome>. Accessed 16th January 2014.

Amsterdam Centre for Digital Humanities (2013). Modeling Crowdsourcing for Cultural Heritage. <http://cdh.uva.nl/projects-2013-2014/m.o.c.c.a.html>. Accessed 17th January 2013.

- Brabham, D. C., (2013). *Crowdsourcing*. MIT Press Essential Knowledge Series. London, England, MIT Press.
- British Library (2014). *Unlock London Maps and Views*. <http://support.bl.uk/Page/Unlock-London-Maps>, Accessed 29th January 2014.
- Brohan, P. (2012). *New Uses for Old Weather*. Position Paper, AHRC Crowdsourcing StudyWorkshop, May 2012. <http://crowds.cerch.kcl.ac.uk/wp-content/uploads/2012/04/Brohan.pdf>. Accessed 29th January 2014.
- Brumfield, B. (2013a). *Itinera Nova in the World(s) of Crowdsourcing and TEI*. Collaborative Manuscript Transcription Blog. <http://manuscripttranscription.blogspot.co.uk/2013/04/itinera-nova-in-worlds-of-crowdsourcing.html>. Accessed 17th January 2014.
- Brumfield, B. (2013b). *A Gresham's Law for Crowdsourcing and Scholarship*, Collaborative Manuscript Transcription Blog. <http://manuscripttranscription.blogspot.co.uk/2013/10/a-greshams-law-for-crowdsourcing-and.html>. Accessed 28th January 2014.
- Brumfield, B. (2013c). *The Collaborative Future of Amateur Editions*. Collaborative Manuscript Transcription Blog, <http://manuscripttranscription.blogspot.co.uk/2013/07/the-collaborative-future-of-amateur.html>. Accessed 28th January 2014.
- Brumfield, B. (2013d). In Van Zundert, J. J., Van den Heuvel, C., Brumfield, B., Van Dalen-Oskam, K., Franzini, G., Sahle, P., Shaw, R., Terras, M. (2013). *Text Theory, Digital Documents, and the Practice of Digital Editions*. Panel session, Digital Humanities 2013, University of Nebraska, Lincoln. July 2013.
- Causser, T. and Terras, M. (Forthcoming 2014) "Crowdsourcing Bentham: beyond the traditional boundaries of academic history". Accepted, *International Journal of Humanities and Arts Computing*.
- Cushing, E. (2013). "Amazon Mechanical Turk: The Digital Sweatshop." *UTNE*, <http://www.utne.com/science-and-technology/amazon-mechanical-turk-zm0z13jflin.aspx#axzz3DNzILSHI>, January/February 2013.
- Finnegan, R. (2005). *Participating in the Knowledge Society: Research beyond University Walls*. Houndmills-Basingstoke: Palgrave Macmillan.
- Flew, T. (2008) *New Media: An Introduction* (3rd ed.). Melbourne: Oxford University Press.
- Frost Davis, R. (2012). "Crowdsourcing, Undergraduates, and Digital Humanities Projects". <http://rebeccafrostdavis.wordpress.com/2012/09/03/crowdsourcing-undergraduates-and-digital-humanities-projects/>. Accessed 29th January 2014.

Grint, K. (2013). Progress Update, 24 to 30 August 2013, Transcribe Bentham Blog. <http://blogs.ucl.ac.uk/transcribe-bentham/2013/08/>. Accessed 29th January 2014.

Hedges, M. and Dunn, S. (2012). Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research. Arts and Humanities Research Council. <http://crowds.cerch.kcl.ac.uk/>, Accessed 16th January 2014.

Holley, R. (2010). Crowdsourcing: How and Why Should Libraries Do It?, *D-Lib Magazine*, 16 (2010), <http://www.dlib.org/dlib/march10/holley/03holley.html>. Accessed 17th January 2014.

Howe, J. (2006a). Birth of a Meme. Crowdsourcing Blog. May 27th 2006. http://www.crowdsourcing.com/cs/2006/05/birth_of_a_meme.html. Accessed 17th January 2014.

Howe, J. (2006b). The Rise of Crowdsourcing. Wired Magazine, June 2006. <http://www.wired.com/wired/archive/14.06/crowds.html>. Accessed 17th January 2014.

Howe, J. (2006c). Crowdsourcing: a definition. Crowdsourcing Blog. June 2nd 2006. http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html. Accessed 17th January 2014.

Hubble, N. (2006). *Mass-Observation and Everyday Life*. Houndmills-Basingstoke: Palgrave Macmillan.

iDigBio (2013). CITscribe Hackathon. <https://www.idigbio.org/content/citscribe-hackathon>. Accessed 30th January 2014.

Mueller, M. (2014). "Shakespeare His Contemporaries: collaborative curation and exploration of Early Modern drama in a digital environment". *Digital Humanities Quarterly*, Volume 8, Number 3. <http://www.digitalhumanities.org/dhq/vol/8/3/000183/000183.html>

North American Bird Phenology Program, n. d. About BPP. <http://www.pwrc.usgs.gov/bpp/AboutBPP2.cfm>. Accessed 17th January 2014.

Old Weather (2013a). Old Weather: Our Weather's Past, the Climate's Future. <http://www.oldweather.org/>. Accessed 17th January 2014.

Old Weather (2013b). Old Weather, About. <http://www.oldweather.org/about>. Accessed 17th January 2013.

O'Reilly, T. (2005). What is Web 2.0? 30th September 2005. <http://www.oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>, Accessed 16th January 2014.

Our Marathon (2013). About The Our Marathon Archive. <http://marathon.neu.edu/about>. Accessed 28th January 2014.

- Owens, T. (2012a). Crowdsourcing Cultural Heritage: The Objectives Are Upside Down. <http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/>. Accessed 17th January 2014.
- Owens, T. (2012b). The Crowd and The Library. <http://www.trevorowens.org/2012/05/the-crowd-and-the-library/>. Accessed 16th January 2014.
- Ridge, M. (2012). Frequently Asked Questions about crowdsourcing in cultural heritage. Open Objects blog. <http://openobjects.blogspot.co.uk/2012/06/frequently-asked-questions-about.html>. Accessed 18th January 2014.
- Ridge, M. (2013). Digital participation, engagement, and crowdsourcing in museums. London Museums Group blog. <http://www.londonmusemsgroup.org/2013/08/15/digital-participation-engagement-and-crowdsourcing-in-museums/>. Accessed 18th January 2014.
- Schantz, H. F. (1982). The history of OCR, optical character recognition. Manchester Center, Vt., Recognition Technologies Users Association.
- September 11 Digital Archive (2011), About the September 11 Digital Archive, <http://911digitalarchive.org/about/index.php>. Accessed 29th January 2014.
- Shillingsburg, P. L. (2006). "From Gutenberg to Google: Electronic Representations of Literary Texts". Cambridge, Cambridge University Press.
- Soldier Studies (2014). Civil War Voices, Home Page. <http://www.soldierstudies.org/>, Accessed 29th January 2014.
- SeeClickFix (2013). Report non-emergency issues, receive alerts in your neighbourhood, <http://en.seeclickfix.com/>. Accessed 16th January 2014.
- Silvertown, J. 2009: A new dawn for citizen science. *Trends in Ecology & Evolution*, 24, 9, pp. 467-71.
- Text Encoding Initiative (2014). P5: Guidelines for Electronic Text Encoding and Interchange. <http://www.tei-c.org/release/doc/tei-p5-doc/en/html/>. Accessed 29th January 2014.
- Weingart, S. B. (2013). Submissions to Digital Humanities 2014. The Scottbot irregular, <http://www.scottbot.net/HIAL/?p=39588>. Accessed 28th January 2014.
- 4Humanities (2012). All Our Ideas: The Value of the Humanities. <http://4humanities.org/2012/10/all-our-ideas-the-value-of-the-humanities/>. Accessed 28th January 2014.