

Chapter 9

Crowdsourcing the Humanities: Social Research and Collaboration

Geoffrey Rockwell

Introduction

The phrase “Web 2.0” is supposed to have been coined at a conference brainstorming session about the new types of websites emerging after the dot.com collapse.¹ Web 2.0 sites, like the Wikipedia and Flickr, are often characterized by broad participation in content creation. They leverage the web to enable content to be created, not by those who manage sites, but by a community of participants. Web 2.0 participatory technologies are now being used by humanists to structure collaborations for research. This chapter will look at the opportunities for collaborative research in the humanities through two humanities computing projects that enabled collaboration through crowdsourcing. This will be framed by a discussion of collaboration in the humanities.

Humanists, in comparison to the sciences, are reputed to not collaborate on research. The image of the humanist as a “lone ranger” can be traced through images like Rembrandt’s “The Philosopher in Meditation”² or stories like Descartes’ *Discourse on Method*, where he tells his story of original research when he “spent the entire day closed up alone in a room heated by a stove, where I had all the leisure to talk to myself about my thoughts.”³ By contrast, the digital humanities are supposed to be characterized by collaboration, partly because of the need for different skills to complete digital projects and because we have technologies that facilitate collaboration.⁴ As Lynne Siemens summarized it:

Traditionally, research contributions in the humanities field have been felt to be, and documented to be, predominantly solo efforts by academics involving little direct collaboration with others ... However, Humanities Computing/Digital Humanities is an exception to this. Given that the nature of research work involves computers and a variety of skills and expertise, Digital Humanities researchers are working collaboratively within their institutions and with others nationally and internationally to undertake this research.⁵

In this chapter I will start by clearing the ground for crowdsourcing by discussing the limitations and promises for collaboration in the humanities. The point is not to praise collaboration, but to ask how it can be structured through social media for research. I will then discuss two social media projects I was involved in which used information technology differently to structure participatory or social research, and the lessons learned in these projects about collaboration. Both of these projects represent different ways of collaborating while thinking through technology. This will lead to general reflections on crowdsourcing in the humanities. The chapter will conclude by arguing that we need to rethink our models for how knowing is done in the humanities to account for the distributed knowing illustrated by social research projects.

Weed Control

In a section of the introduction to *Humanities Computing* called “Weed Control,” Willard McCarty takes exception to the way in which collaboration has been taken as a “transcendental virtue, to be applied regardless of context” (2005: 9). He agrees that even so-called solitary scholars are indebted to the work of others, but that that indebtedness does not mean that collaboration should become an end in itself, something it seems to have become. One could add that we already collaborate in all sorts of ways: in departments to deliver the curriculum, in program committees when we organize conferences,

when we correspond, when we edit journals, and when we review the work of others. These collaborations seem so obviously part of our academic life that we have ceased to notice them, calling instead for the new types of collaboration mediated by technology.

McCarty goes on, however, to make a stronger point about collaboration, that “[i]n the humanities, scholars have tended to be physically alone when at work because their primary epistemic activity *is* the writing, which by nature tends to be solitary activity.”⁶ McCarty seems content to leave the epistemological issue there, without questioning the primacy of writing in the generation of knowledge as he questions the perceived primacy of collaboration. I will return to the epistemological issue at the end of this chapter, but for the moment I note that McCarty concludes the control of the weed of collaboration with Ursula Franklin’s distinction between “holistic” and “prescriptive” technologies.⁷ McCarty, like Franklin, favors the holistic craft approach to technology where the (solitary) individual controls the process (and technologies) over the threat of prescriptive and industrial approaches where the individual is managed by technology. McCarty wants to defend solitary epistemic writing as research against enforced and managerial collaboration, and is right to do so. In this context, one can see why his model for humanities computing is that of a methodological commons where research methods are adapted for application to the problems of the traditional humanities without challenging the disciplines or their research practices.⁸

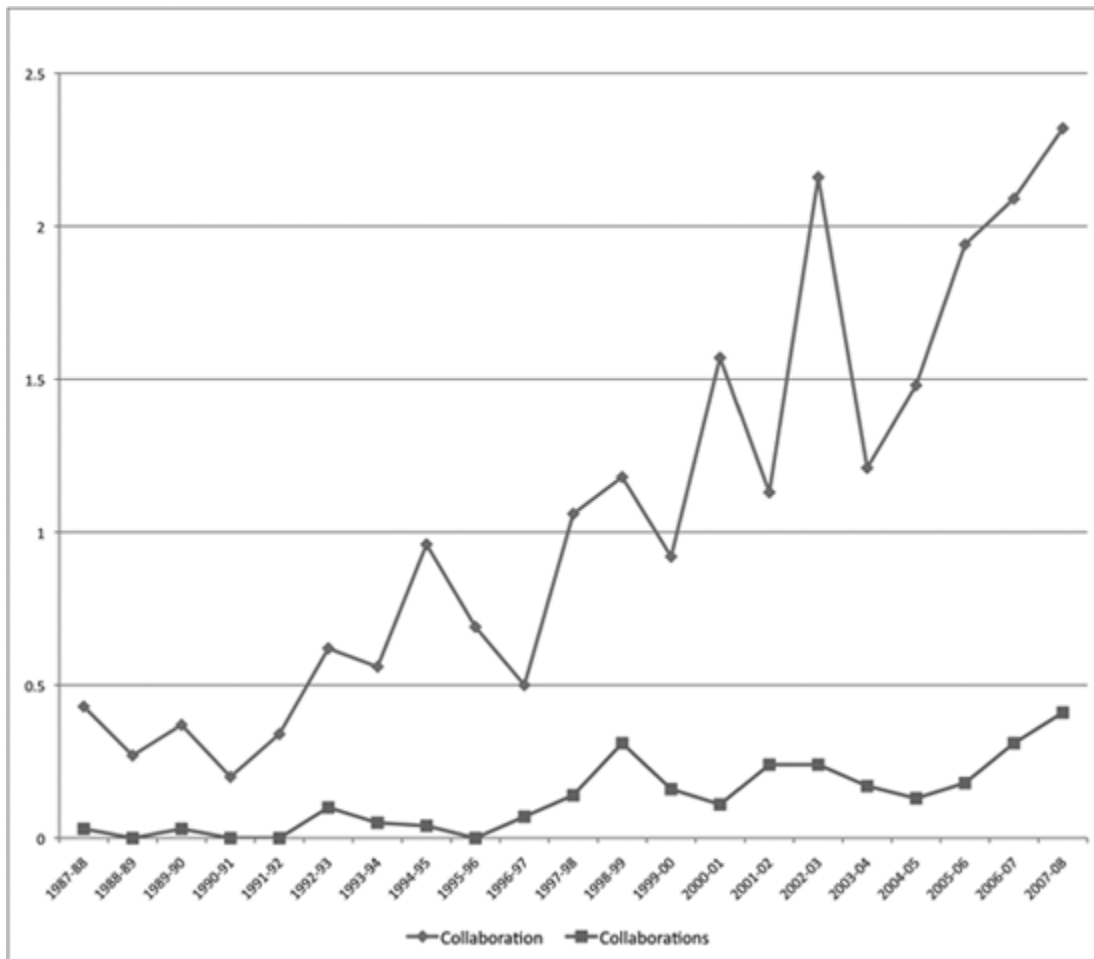


Figure 9.1 Growth of “collaboration” and “collaborations” across the *Humanist* corpus

Note: This graph was generated by Voyant Tools, S. Sinclair and G. Rockwell (2009) <<http://voyant-tools.org>>. The corpus was created using the text files from the *Humanist* discussion group archive at <<http://www.digitalhumanities.org/humanist/>>.

By contrast, digital humanists like Patrik Svensson, building on Cathy Davidson, believe the traditional ways of doing computing in the humanities for which McCarty is an advocate have been supplemented by a new wave of doing digital humanities or “Humanities 2.0,” where technology is used

Copyright © 2011. Routledge. All rights reserved. May not be reproduced in any form without permission from the publisher, except fair uses permitted under U.S. or applicable copyright law.

to undermine practices and epistemic assumptions.⁹ The very shift in names from “humanities computing” to “digital humanities” is indicative of the change from a model of applying computing in the humanities to a model where the digital is the catalyst triggering changes in the disciplinarity of the humanities. The new digital humanities for Davidson, like the Web 2.0 phenomenon, is characterized by the collaborative weeds of interactivity and participation.¹⁰ The digital decenters authority, challenging peer review, solitary writing and authority, and broadcast ways of teaching. One man’s weed is another woman’s spice. She is aware of some of the perils of social media like Facebook, but in the end she favors experimenting with forms of participation that the network enables as a way of renewing the humanities.

Crowdsourcing the Humanities

I tend to agree with both sides and believe that they can be reconciled. Collaboration is not a transcendent value, as McCarty points out, but it can, if properly structured, advance research and achieve valuable research ends—ends which the solitary scholar might not achieve writing alone. Both the 1.0 and 2.0 type projects are valuable—one builds on the other. Talk about collaboration should not squeeze out the traditional curiosity-driven practices, despite what funding agencies may want. The interesting question is how to structure collaboration properly so that it contributes to, rather than hinders, research. This is a variant on Steve Fuller’s fundamental problem of social epistemology:

How should the pursuit of knowledge be organized, given that under normal circumstances knowledge is pursued by many human beings, each working on a more or less well defined body of knowledge and each equipped with roughly the same imperfect cognitive capacities, albeit with varying degrees of access to one another’s activities?¹¹

One particularly promising way of organizing the pursuit of knowledge through technologically mediated Web 2.0 collaboration is what Jeff Howe calls “crowdsourcing.”¹² Crowdsourcing uses social media tools like wikis to enable a “crowd” or group of people to create something of value. The crowd of volunteers enabled becomes the human resource whose small contributions, if properly coordinated, can make a real difference. Perhaps the best-known crowdsourcing project is the Wikipedia, where volunteers have authored, edited and argued millions of encyclopedia articles in less than ten years without being paid to author and without editorial supervision.¹³ Before the Wikipedia, few people believed that any resource as important as an encyclopedia could be created through unsupervised volunteer work. As Shirky puts it: “Skepticism about Wikipedia’s basic viability made sense back in 2001; there was no way to predict, even with the first rush of articles, that the rate of creation and the average quality would both remain high ...”¹⁴ But the Wikipedia proved the skeptics wrong; now everyone is trying to figure out how to leverage crowdsourcing for commercial advantage, as it has been noticed that some of the most valuable internet assets are built on content that is not paid for, but voluntarily contributed. YouTube, Facebook, Flickr and other Web 2.0 phenomena are all built on our content. They provide us with a structured site to place content and connect it with others, but these companies also mine our data and return it to us with embedded ads.

It should be noted, however, that crowdsourcing projects are not free. While the labour of many generously contributes to them, they still need support for the center that designs the application, maintains the software and runs the servers. Even the Wikipedia has to raise funds like a public broadcaster. Even further, the design of a crowdsourcing application significantly constrains the contributions of the crowd. There is a way in which crowdsourcing applications combine the vision of a solo researcher or small team with the broad contributions of the many in a distribution of work that has a “long tail.” Crowdsourcing is not the opposite of solo work; it is an extension, where some have more influence than others.

Not all social media projects are commercial, nor are they all structured like the Wikipedia. A number of digital humanities projects have adapted crowdsourcing to research tasks in the humanities. One of the first was the *Suda On Line* (SOL) project that Anne Mahoney discusses in “Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia.”¹⁵ This project organized professional and amateur classicists to help translate passages from the Suda, a Byzantine Greek historical encyclopedia. This

project was not, like the Wikipedia, completely open. Instead it had a workflow where translations were checked. Other projects include the *Transcribe Bentham* project that is engaging the public in the transcription of the 60,000 papers by Bentham in the library of University College London,¹⁶ and the *Australian Historic Newspapers* project run by the National Library of Australia, that encourages the public to correct OCR'd text of old newspapers.¹⁷ Crowdsourcing, despite the skeptics, clearly works even in the small scale for serious research projects. What, then, can we learn from existing research projects?

Dictionary of Words in the Wild

At this point I will turn to two projects that I have worked on. They are by no means the only experiments in crowdsourcing in the humanities as noted above, but they both illustrate a way in which crowdsourcing can be used in research.

The first project I will outline is a social research site, the *Dictionary of Words in the Wild*¹⁸ where users can get accounts and upload pictures of public textuality which they tag according to the words that appear. This site resembles social photo sites like Flickr, but it has a research purpose which is to document the forms of text that appear outside of print, on signs, as graffiti, tattoos, and so on. The project does not start with a hypothesis of what public textuality (in the wild) is—it allows contributors to define it through participation.



Figure 9.2 Home page of the *Dictionary of Words in the Wild*

The *Dictionary* was developed with support from the TAPoR (Text Analysis Portal for Research) project at McMaster University. The first version was programmed by Andrew MacDonald in Ruby on Rails in 2006. Since then Lian Yan and Kamal Ranaweera have added to the code for the project and it has been transferred to the University of Alberta with the new domain name of <lexigraphi.ca>. The *Dictionary* has been reported on at *Digital Humanities 2008* at the University of Oulu and elsewhere.¹⁹

number of the users have uploaded most of the images, while most users have only uploaded one or two. According to the literature, this is common and in fact is one of the features of successful crowdsourcing projects—they harness not only the committed contributor but the long tail of less committed contributors.²⁰

How is the *Dictionary* a social research project? The project was started as an alternative way of asking about the text outside of print—the text in the wild. Compared to the river of research about published textuality, there is little written about the noisy textuality that surrounds us in our everyday contexts.²¹ Considering how much of it we read everyday, it is surprising that literary theory has not considered as discourse the text on T-shirts, signs, posters and painted on our walls. The *Dictionary* is a first step that tries to document textuality in the wild so that we can theorize it. Rather than simply post my photographs of what I considered public textuality, by creating a crowdsourcing project, the *Dictionary* invites others to contribute to the documentation without unduly constraining their imagination about what constitutes public textuality. The *Dictionary*, by enabling participation, has therefore gathered far more images of public textuality, and has a much greater diversity of images than if it had been a research project conducted by one person or a small team.

An important aspect of the project is that it is exploratory in the sense that it does not start from an assumption or theory. Instead, the project makes a simple point about the rich variety of textuality outside of the traditional covers within which we usually look. The sheer exuberance of the uploaded makes a visual argument about the importance of words in the wild that no narrative could. The project invites the reader to explore the variety of images or participate by adding new images, if you think we have missed forms of public textuality. As participants have noted, if you then start carrying your digital camera and looking for interesting examples of digital textuality, it changes your perception of the urban landscape.²² In short, the *Dictionary* can work as a different type of rhetoric that changes your perception through participation.

Day in the Life of the Digital Humanities

The second project I will discuss is the *Day in the Life of Digital Humanities (Day of DH)* project which brought together digital humanists around the world to blog one day of their work as a way of answering the question “Just what do computing humanists really do?” This project has been run twice, first in 2009, when we had 84 participants, and again in 2010, when we had just over 150. The project was organized by a team that included me, Stan Ruecker, Peter Organisciak, Megan Meredith-Lobay, Julianne Nyhan, and Kamal Ranaweera.

List of Day of DH 2009 Participants

Below, in alphabetical order by surname, is a list of colleagues who are expected to participate in A Day in the Life of Digital Humanities on 18 March 2009.

- William Allen - Arkansas State University
- Rafael Alvarado - Dickinson College
- Jon Bath - University of Saskatoon
- Craig Bellamy - King's College London
- Brett Bobley - National Endowment for the Humanities
- Matthew Bouchard - University of Alberta
- John Bradley - King's College London
- Susan Brown - University of Guelph
- Kai-Christian Bruhn (link to user page) - University of Applied Sciences Mainz
- Lou Burnard - Oxford University Computing Services



Figure 9.5 Alphabetical list of participants in the 2009 *Day of DH* project, with map

Participants who signed up agreed to blog their digital humanities work on the same day, March 18,

with text and images. Before the day, we used an email list to prepare participants and to discuss issues like securing permissions from people appearing in photographs. We used a wiki page as an anchor site for the project, with explanations and links to documentation. Here is the main page for the project in 2009.

Before the day itself, we created a blog for each participant. They were encouraged to enter an About entry before March 18 to introduce themselves and to test the system. This also gave participants a way to talk about projects that might not be dealt with on the particular day when we were all going to blog. The blog, of course, was primarily for participants to enter reflections as the day in question progressed. Many participants, including myself, ended up having to continue editing their entries and uploading photographs after March 18, but the idea was still to use the day as an organizing principle. Here is the start of my *Day of DH* in 2009 (see [Figure 9.7](#)).

Day in the Life of the Digital Humanities 2009

A Day in the Life of the Digital Humanities has wrapped!

You can browse the main feed of events [☞](#), find an individual participant [☞](#), read participants' biography pages [☞](#), and follow along with the comment feed [☞](#).


Also, the TEI data is now available for review [☞](#).

Thanks to all participants for making this project a success.

The Day in the Life of the Digital Humanities (Day of DH), a project looking a day in the work life of people involved in humanities computing.

Contents [hide]

- 1 What is it?
- 2 Process
- 3 Questions
- 4 Contact
- 5 Project Details
- 6 How Do I...
- 7 Other Links
- 8 Day of DH 2



What is it?

A Day in the Life of the Digital Humanities (Day of DH) is a community publication project that will bring together digital

Figure 9.6 Home wiki page for the 2009 *Day of DH* project



Figure 9.7 Day of Geoffrey Rockwell

On the day itself in 2009, we were surprised to find that the project got attention as it was happening, primarily through Twitter. A number of people hearing about it asked to join on the day of the project. In the second iteration, we tried to take advantage of this by suggesting a Twitter Hashtag.

After the Day in question was over, the organizing team slowly exported all the blog entries, proofed them, and tagged them, creating a dataset or collaborative publication available for others to read or analyze.²³ The hope is that we can sustain the project, repeating it each year with the help of others. This would add a diachronic dimension to the project, where people could compare reflections on the digital humanities over time.

The *Day of DH* project, unlike the *Dictionary*, was developed using off-the-shelf tools. The website that documents the project is a wiki created with MediaWiki. The individual blogs that participants used to document their day were set up using WordPress MU, which allowed us to rapidly create all the needed blogs. The only tools that we had to write ourselves were the scripts for processing the output from all the WordPress blogs.

The idea of the project was to document what we do in the digital humanities through a participatory project. The project can be seen as a form of autoethnography, where a community writes its own ethnographic narrative, or it can be seen as a type of one-day conference, since what effectively happened in both iterations is that the participants blogged about the digital humanities while reading each other's blogs.

This project, like the *Dictionary*, is also exploratory in the sense that it does not start with a hypothesis or theory about the digital humanities, but uses crowdsourcing to provide a communal answer. The *Day of DH* is, in effect, a different kind of answer to the question of what do digital humanists do, a collective answer rather than a reflective one. The project could have been conceived of as a survey, where people would be polled as to what they do and the answers summarized and reflected on, but that would not have had the same impact, and I doubt that we would have had as many volunteers if we had asked people to fill out a survey. Instead, through crowdsourcing, participants could negotiate the answer with each other and they could be assured that their voice would not be drowned in the summarization. For that matter, the full dataset of blog entries is available for analysis and summarization by everyone. We even provided a live RSS feed of the entries for those who wanted to run real-time analytical tools, as some like Stéfan Sinclair did. The *Dictionary* makes possible summative or reflective answers, but in and of itself it is a different type of answer, one that is communal and dialogical.

We chose the name *Day in the Life of the Digital Humanities* for the project, not only because it would indicate the expected commitment of participants, but because it might resonate with readers. The title, for example, was partly influenced by a lecture that a prominent digital humanist, Ed Ayers, gave on “What Does a Professor Do All Day, Anyway?”²⁴ It should also remind readers of Solzhenitsyn’s *One Day in the Life of Ivan Denisovich*, Joyce’s *Ulysses*, and Woolf’s *Mrs. Dalloway*, all of which are structured around one day. Finally, it might remind people of *The Day in the Life of ...* series of photography books, which bring together images of a country by many photographers, all taken on the same day. In the *Day of DH*, we also encouraged participants to upload pictures, though we had to worry about ethics requirements. The hope was that participants and readers would see the *Day of DH* as being in a tradition of organizing understanding around a day.

This project raises questions about authorship and academic recognition. We have submitted a paper about the project, “A Day in the Life of the Digital Humanities 2009,” to a peer-reviewed online journal in the field. This, of course, raised questions as to the authorship of the project. While this is still being negotiated, the solution that we hit upon was to have the organizing team as the co-authors of the paper that describes the project, and to also submit the dataset, along with a short “About the Dataset” document, as co-authored by all the participants and organizers. It was important as we were setting up the project that all participants felt that this was a collaborative online publication where all were co-authors. It is doubtful that any of us will get significant academic recognition for this project itself (as opposed to any papers in journals about the project). This, alas, is one of the many challenges that we face when collaborating in a humanities academic culture that still considers sole-authored written works as normative.²⁵ It is not, however, a problem particular to crowdsourcing projects—any co-authored project or digital project suffers the same problems when evaluated.

Making Crowdsourcing Work

Crowdsourcing projects, as attractive as they may be, run the risk of falling flat if no one volunteers to participate. The big question behind the design of social media projects is motivation—what motivates participation and how to design for motivation.²⁶ From my experience and from the literature, here are some general guidelines for developing crowdsourcing projects for humanities research.

Ask for Help. No one will help you in a research project unless you ask. Social research projects depend on volunteers to achieve research ends. To get volunteers, whether for a computer-supported project or not, you need to ask for help, and when asking, you have to be able to quickly explain what help is needed. Don’t be afraid to promote your project and outright ask for help.

High Concept, Good Title. For volunteers to decide to participate in a crowdsourcing project, they need to understand what they are committing to. It is important, therefore, to communicate efficiently the idea of the project. A title for a project like “Wikipedia” goes a long way towards describing the project—that is, it is an encyclo-pedia created with a wiki. I chose the name *Day in the Life of the*

Digital Humanities for a similar reason; it would communicate the commitment asked for (a day) and remind people of the “Day in the Life of ...” photography series, to give them a sense of what could be achieved.

Small Autonomous Tasks. Crowdsourcing is suited to tasks that can be broken into smaller, relatively autonomous tasks. Jeff Howe’s sixth rule is “Keep It Simple and Break It Down.”²⁷ Malcolm Gladwell, in a welcome critique of the revolutionary claims made for social media like Twitter, points out that social media are not suited for sustained and complex tasks, especially ones that carry significant risk (like real revolutions).²⁸ Instead, they are suited for lots of little tasks that carry little risk, are relatively quick and easy to complete, and do not call for lots of collaboration between participants. That is not to say that crowdsourcing cannot achieve complex things, but it often does so through leveraging many doing small tasks.

Support Imbalance. An imbalance between contributors is normal, so you should plan for it. A small number of participants will contribute a lot each, while a large number will contribute a little. Crowdsourcing projects should not neglect the long tail of participants who just want to give a little time. They are your word-of-mouth advertising and their contributions add up, so support them as well as the power contributors.

Interpretative and Creative. Crowdsourcing in the humanities does not have to be used only for interpretative tasks like correction, transcription, and translation. It is tempting to want to manage the chaos of a social project so that the crowd can only contribute safe tasks that can be moderated. But why not imagine projects that encourage original contributions. Crowdsourcing can also be used for creative tasks where users contribute new material to an open-ended project, like the *Dictionary of Words in the Wild*. The trick is to make small original contributions possible.

Support Diversity. Crowdsourcing should be structured to maximize diversity, not to flatten it. Collaboration isn’t necessarily a good thing if it leads to erasing difference. Eliot Smith (2007) discusses how groups can perform worse than individuals or fall prey to biases similar to those of individuals. For example, too much communication can lead to a convergence of opinion in a group, rather than maintaining healthy diversity.²⁹ Communication should be managed to avoid “bandwagon” convergence.

Vandalism. Vandalism of different sorts can be a real problem for crowdsourcing, especially automated or spam vandalism. For this reason, it is useful to anticipate and prepare for misuses of your site in ways that do not undermine the participatory character of the project. Holley (2009) gives an example of a thorough risk analysis.

Use Custom and Open Tools. You don’t have to create custom software for your crowdsourcing project. You can creatively use existing tools or customize tools. For the *Day in the Life of the Digital Humanities* project, we used off-the-shelf tools from a wiki to WordPress Mu. The *Transcribe Bentham* project has adapted MediaWiki rather than build a custom system. I expect that there will soon be good open-source crowdsourcing tools available that can be adapted easily to humanities projects.

Recognize Community. Crowdsourcing creates a community, even if the project is intended to be utilitarian. You need to support that community, without harassing it, through appropriate communication vehicles. How you do that will vary, depending on the project, but the sense of community is one of the motivations for volunteers to participate.

Organizational Team. You need a team to set up and run a crowdsourcing project—don’t try to do it alone. Some members of the team might emerge from the crowd of participants. Find ways to recognize the organizing team, whether they contributed programming, wrote documentation, helped others, or cheered you on.

Small Crowds Work Too. The success of a crowdsourcing project need not depend on large numbers of contributors. The numbers of collaborators in the projects discussed above are small, especially when compared to large-scale projects like the Wikipedia, but the numbers were sufficient to achieve progress on the research task.

Knowledge in Process. Like many digital projects, crowdsourcing projects should be thought of as processes not products. Articles in the Wikipedia are continually being edited—often, but not always, for the better. If you see an error, fix it. Crowdsourced knowledge is therefore more of a relationship than a finished object. It starts out tentative and then slowly gets refined. The history of that process is often available, which adds an open dimension to the knowledge in process. With Wikipedia articles, you can see their editing history and unwind it.

Distributed Knowing

The real limitation to the adoption of crowdsourcing is our imagination about what research is in the humanities. This brings us back to the question of epistemic activity raised by McCarty when he writes that, for the humanist, the “primary epistemic activity *is* the writing, which by nature tends to be solitary activity.”³⁰ While this may be true now, it has not always been that way and it need not continue to be this way. One could argue that dialogue was the primary epistemic activity from the Platonic Academy up until the emergence of the research university and its emphasis on publication. Or one could argue that the primary activity of the humanist is reading not writing, at least in terms of time spent by most people, and that writing is a specialization by professional humanists. Or one could argue that teaching is the primary epistemic activity by which ideas are transmitted. My point is that the usual move to argue for the primacy of writing is historically contingent and possibly not even true today, depending on what metric you use to establish primacy.

Nor is the primacy of writing a normative truth in the sense that writing should be our primary activity. Just as McCarty calls us to guard against treating collaboration as a transcendental value for the humanities, so we should also guard against writing becoming the fetish of a humanities driven by single-authored publication. A philosopher like Heidegger would call us to thinking as the primary activity. Heidegger and Wittgenstein would warn us against writing as a way of grasping the truth. For both there are dangers to writing, especially about truths that may be ineffable. Ironically, we know this through their writings, but that shows how important writing has been for the transmission and preservation of knowledge over time.

More recently, there has been a convergence of fields around the idea that thinking can be distributed, and much has been published on this.³¹ Fields like social epistemology and the sociology of knowledge look at how knowledge is formed by groups. Traditional epistemology is concerned with how individuals can know and specifically how they can be certain of the truth of their beliefs. This view has been challenged on two fronts: social epistemologists have argued that there is also knowledge in groups independent of the individual knowing, and some cognitive scientists have argued that distributed cognition is possible, where the act of thinking is distributed across different individuals, be they ants collaborating, nodes in a neural net, or people in some structured context like a law court. As Fuller argues, in the face of the common view of the primacy of solitary thinking:

I reject the Cartesian gesture of withdrawing from all social intercourse as a means of getting into the right frame of mind for posing foundational questions about the nature of knowledge. For even though the social world may appear to be a confusing place from which to deliver epistemic judgments—certainly more confusing than the privacy of one’s own study—it is nevertheless the *normal* (and really the *only*) place in which such judgments are delivered.³²

Psychologists and cognitive science researchers are also considering distributed cognition through technology, or how technologies can extend the cognition of individuals and co-ordinate the cognition of groups.³³ This convergence should interest digital humanists since it brings society, culture, technology, and history back to cognition. It turns out that one cannot understand cognition by studying

the brain alone and that one needs to understand how thinking is social, contextual, and technologically extended. It may not even be the case that writing is best described as a solitary Cartesian activity. It could be that the real activities of knowledge are best understood as distributed, in which case the question is again, how to best organize distributed cognition?

This brings us back to the problem of the imagination of knowing. To understand crowdsourcing in the humanities, we need to provide alternatives to the primacy of the solitary writer as the image of how knowledge is organized. One alternative image would be a trial in a law court. A trial is a classic example of the social generation of knowledge where a group of people with different roles negotiates actionable knowledge about a defendant. The judge manages the process, lawyers present arguments for and against the defense, and a jury deliberates. The end result is a verdict that goes on record as the best available knowledge judgment about the situation that can be acted upon. A trial is a way of organizing distributed cognition that has a long history. It is an alternative paradigm for understanding human knowledge that illustrates the importance of the organization of the group.

Now what we need are convincing paradigms of what crowdsourced research might look like that do not just treat it as a way of distributing labor in order to get traditional results. While we might use crowdsourcing this way, achieving traditional Humanities 1.0 aims, the promise of Humanities 2.0 is that it will offer new aims, new types of knowledge, and new outcomes. Not all these novelties will stand the test of time, but the possibilities are worth the exploration so that we can learn about research practices. The two projects described above were both imagined to be such exploratory projects. They were designed not to deliver traditional writings, but to develop collaborative knowings that work rhetorically as responses in the dialogue of the humanities. In trying to understand how they are a different type of research, we might better understand the epistemic activity of writing with which they are contrasted.

Conclusion

To conclude, the challenge of collaboration is not, I believe, to traditional individual research, but to the organization of professional scholarship in general. Clay Shirky (2008) describes how social media like blogs and Twitter call into question the social bargain around journalism as the fourth estate. Journalists were accorded special privileges, like the protection of sources, because we valued their investigations. Now that anyone with a blog is some sort of journalist and social media seem to be undermining the newspaper business, it is unclear that there is support for the social bargain that protected journalists. We do not know what will replace the newspaper, but the line between a professional journalist and an amateur is vanishing.

Likewise, I believe that the distinction between the professoriate and amateur researchers will also blur as more and more research is shared through social media.³⁴ While we never believed that the professoriate should have special legal privileges, we have acted as if we were a special caste of intellectual worker who should be supported by society and protected with administrative mechanisms like tenure. Unfortunately we live at a time when our social contract with society is being challenged and the humanities have not found a satisfactory way to answer utilitarian questions from the larger community.

The good news is that social research models offer one way to show relevance through involving a larger community. Crowdsourcing projects provide structured ways to involve the growing numbers of well-educated amateurs with time on their hands, so that they can contribute and be recognized for their contributions. As Howe points out: “An exponential rise in education has coincided with the emergence of the greatest mechanism for distributing knowledge the world has ever seen: the Internet.”³⁵

Such projects do not just provide educated amateurs with a meaningful way to use their leisure time, but they can create small communities of inquiry interested in advancing knowledge about a phenomenon—and that is what should matter. Such projects can create the network of friends that understand the value of the humanities because they are part of the community of inquiry. Such participation, where people understand the value of humanities inquiry, is a far better response to our

critics than any “gotcha” argument is likely to be. These amateurs vote with their feet, creating knowledge with us that is then available back to the larger community.

Further, crowdsourcing projects, especially in times of diminished funding, provide ways of doing “big humanities” without needing big grants as most digital humanities projects do. As Stanley Katz writes: “Let’s not whine about the humanities. Let’s watch, think and act.”³⁶ One way of acting when there is little funding is to imagine and develop projects that are inclusive and participatory. Both of the projects described above were not grant-funded. They were developed with the resources at hand.³⁷

Above all, crowdsourcing erases differences between professional scholars with degrees and those who are self-taught amateurs without formal training. This is the decentering of authority that Davidson writes about. That does not mean that the professoriate is likely to become irrelevant; after all, we are given time to do research and we have significant resources to help organize research projects, including the resources needed to set up crowdsourcing projects. What it means is that a different relationship between the professoriate and amateur researchers is emerging. The digital humanities is one of the sites where a collaborative relationship is being prototyped using networking. We have the skills to imagine different configurations that bring together professionals and amateurs in new forms of knowing together. Along with the opportunities, there is the danger that if the academy does not meaningfully reach out to amateur researchers, starting with our students, amateurs will organize social research sites on their own or crowd to commercially developed services. In either case, the academy could find itself left behind like print encyclopedias in the era of the Wikipedia. We could find ourselves irrelevant to and, what is worse, bypassed by the public that public universities depend on for support, when we could have evolved research gatherings that bridge the imbalances and support diversity of commitment. In short, I believe the best scholarship is done when professionals and amateurs are provided with mechanisms to collaborate for the advancement of knowledge. That is the real promise of collaboration, that it crosses sectors, and the digital humanities are in a unique position to help construct the needed bridging mechanisms.

Great ages of science are great ages of the humanities because an age isn’t a historical period but a construct, and constructs are the work of humanists.³⁸

References

- Ayers, E.J., ‘What Does a Professor Do All Day, Anyway?’ [Lecture] University of Virginia (1993), <<http://www.virginia.edu/insideuva/textonlyarchive/93-12-10/7.txt>>.
- Davidson, Cathy N., ‘Humanities 2.0: Promise, Perils, Predictions’, *Publications of the Modern Language Association of America* 123:3 (2008): 707–17.
- Descartes, René, *A Discourse on the Method of Correctly Conducting One’s Reason and Seeking Truth in the Sciences*, trans. Ian Maclean (Oxford: Oxford University Press, 2006).
- Franklin, Ursula, *The Real World of Technology*. CBC Massey lectures series. (Concord, ON: House of Anansi Press Ltd, 1992).
- Fuller, Steve, ‘On Regulating What Is Known: A Way to Social Epistemology’, *Synthese* 73 (1987): 145–83.
- Gladwell, Malcolm, ‘Small Change: Why the Revolution Will Not Be Tweeted’, *New Yorker*, October 11, 2010, <http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell?currentPage=1>.
- Gureckis, Todd M. and Robert L. Goldstone, ‘Thinking in groups’, *Pragmatics & Cognition* 14:2 (2006): 293–311.
- Hollan, James, Edwin Hutchins and David Kirsh, ‘Distributed Cognition: Toward a New Foundation for Human–Computer Interaction Research’, *ACM Transactions on Computer–Human Interaction* 7:2 (2000): 174–96.
- Holley, Rose, *Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers*, National Library of Australia, 2009, <<http://trove.nla.gov.au/work/22531631>>.
- Howe, Jeff, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business* (New

- York: Three Rivers Press, 2009).
- Hutchins, Edwin, 'How a Cockpit Remembers its Speed', *Cognitive Science* 19 (1995a): 265–88.
- Hutchins, Edwin, *Cognition in the Wild* (Cambridge, MA: MIT Press, 1995b).
- Katz, Stanley, 'Let's Not Cry for the Humanities—Yet', *The Chronicle Review's* blog Brainstorm, February 25, 2009, <<http://chronicle.com/blogPost/lets-notcry-for-the-humanities-yet/6687>>.
- Mahoney, Anne, 'Tachypaedia Byzantina: The Suda On Line as Collaborative Encyclopedia', *Digital Humanities Quarterly* 3:1 (2009).
- McCarty, Willard, *Humanities Computing* (New York: Palgrave Macmillan, 2005).
- McCarty, Willard, Geoffrey Rockwell, Andrew MacDonald and Eleni Pantou-Kikkou, 'A Carnival of Words: The Dictionary of Words in the Wild and Public Textuality', conference paper at *Digital Humanities 2008*, University of Oulu, Finland, 2008.
- O'Reilly, Tim, 'What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software', <<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>>.
- Organisciak, Peter, 'Why Bother? Examining the Motivations of Users in Large-scale Crowd-powered Online Initiatives', MA thesis in Humanities Computing at the University of Alberta, defended August, 2010, <<http://crowdstorming.wordpress.com/2010/08/17/on-the-motivations-of-crowds/>>.
- Rockwell, Geoffrey, 'Computing with the Infrastructure at Hand: Collaborative Research for the Arts and Humanities in Times of Scarcity', online at *Philosophi.ca*, <<http://philosophi.ca>>.
- Scollon, R. and S. Scollon, *Discourses in Places: Language in the Material World* (London: Routledge, 2003).
- Shirky, Clay, *Here Comes Everyone: The Power of Organizing with Organizations* (with an updated epilogue) (London: Penguin, 2008).
- Siemens, Lynne, "'It's a Team If You Use 'Reply All'": An Exploration of Research Teams in Digital Humanities Environments', *Literary and Linguistic Computing* 24:2 (2009): 225–33.
- Smith, Eliot R., 'Social Relationships and Groups: New Insights on Embodied and Distributed Cognition', *Cognitive Systems Research* 9 (2007): 24–32.
- Svensson, Patrik, 'The Landscape of Digital Humanities', *Digital Humanities Quarterly* 4:1 (2010).

1 O'Reilly (n.d.)

2 Rembrandt van Rijn, 1632, Musée du Louvre.

3 Descartes (2006), Part 2: 12.

4 See, for example, the HASTAC “About HASTAC” page where they define the network thus:

HASTAC (“haystack”) is a network of individuals and institutions inspired by the possibilities that new technologies offer us for shaping how we learn, teach, communicate, create, and organize our local and global communities. We are motivated by the conviction that the digital era provides rich opportunities for informal and formal learning and for collaborative, networked research that extends across traditional disciplines, across the boundaries of academe and community, across the “two cultures” of humanism and technology, across the divide of thinking versus making, and across social strata and national borders. (“About HASTAC” <<http://www.hastac.org/about-hastac>>). You can also see the importance of collaboration in Svensson (2010).

5 Siemens (2009): 225.

6 McCarty (2005): 12; emphasis in the original.

7 McCarty (2005). The distinction he draws on from Ursula Franklin is from *The Real World of Technology*.

8 See McCarty (2005): 119, Figure 3.1: “An intellectual and disciplinary map of humanities computing.”

9 Svensson (2010), paragraph 15.

10 Davidson (2008).

11 Fuller (1987): 145.

12 Howe (2009).

13 See the Wikipedia at <<http://www.wikipedia.org/>>.

14 Shirky (2008): 117.

15 The website for the crowdsourcing project is <<http://www.stoa.org/sol/>>. To learn more about the project, see Mahoney (2009).

16 For more, see <<http://www.ucl.ac.uk/transcribe-bentham/>>.

17 See Trove, <<http://trove.nla.gov.au/>>, or read Holley (2009).

18 Dictionary of Words in the Wild, <<http://lexigraphi.ca>>.

19 Willard McCarty presented a paper which we co-authored, “A Carnival of Words: The Dictionary of Words in the Wild and Public Textuality” (McCarty et al., 2008).

20 See the section in Shirky (2008): 122–30, titled “A Predictable Imbalance.”

21 An interesting exception is Scollon and Scollon (2003).

22 See Peter Organisciak’s blog post of April 20, 2009, “Words in the Wild” at <<http://www.porganized.com/blog/words-in-the-wild>>.

23 The 2009 dataset, as we call the complete collection of blog entries, is available at <<http://dl.dropbox.com/u/128635/Day%20of%20DH/DayOfDHwithoutImages.zip>>.

24 Ayers (1993).

25 Another collaborative project that I was involved in was the development of a wiki on “The Evaluation of Digital Work” for the Modern Languages Association, to deal with issues around evaluation and recognition of digital work. I started this on my personal wiki at: <<http://www.philosophi.ca/pmwiki.php/Main/MLADigitalWork>> when I was on the MLA Committee on Information Technology. The MLA then took it over and runs it now at: <http://wiki.mla.org/index.php/Evaluation_Wiki>, so that others can maintain it.

26 See Organisciak (2010) for an extended discussion of motivation.

27 Howe (2009): 285.

28 Gladwell (2010). Gladwell’s essay is an antidote to the generally enthusiastic press that social media have received. In particular, it deals with revolutionary claims made for social media. Gladwell does not think that social media is very effective at serious change:

It [social media] makes it easier for activists to express themselves, and harder for that expression to have any impact. The instruments of social media are well suited to making the existing social order more efficient. They are not a natural enemy of the status quo. If you are of the opinion that all the world needs is a little buffing around the edges, this should not trouble you. But if you think that there are still lunch counters out there that need integrating it ought to give you pause.

29 Smith (2007): 28. Smith is not writing about crowdsourcing projects, but about group work in general. He writes: “group members tend to have their opinions drawn in toward that average. Ultimately this process can lead all group members to converge, creating a destructive uniformity of opinions within the group, leading to the disregard of relevant evidence, and possibly even to what has been termed ‘groupthink’ ...” (2007: 28).

30 McCarty (2005): 12.

31 One place to start is Hutchins (1995a), or for more, his book *Cognition in the Wild* (1995b).

32 Fuller (1987): 146.

33 For an introduction to distributed cognition, see Gureckis and Goldstone (2006). For distributed cognition and computing, see Hollan, Hutchins and Kirsh (2000).

34 See the chapter in Howe (2009) on “The Rise of the Amateur.”

35 Howe (2009): 29.

36 Katz (2009).

37 I have posted an online essay, “Computing with the Infrastructure at Hand,” (Rockwell, n.d.) that goes into more detail about how we should consider what we have at hand. To be fair, both of these projects benefit from resources secured by other grants, especially infrastructure grants like those that funded TAPoR.

38 Davidson (2008): 707.