

# Crowdsourcing for Humanities Research

Zephyr Frank, Ryan Heuser, Sophia Paliza-Carre, Gabriel K.  
Wolfenstein, Kerri Young,

Center for Spatial and Textual Analysis  
Stanford University

[www.humanitiescrowdsourcing.stanford.edu](http://www.humanitiescrowdsourcing.stanford.edu)

Sponsored by the Andrew W. Mellon Foundation

## Crowdsourcing for Humanities Research

We begin with three intriguing and problematic terms. Crowdsourcing, in a digital context, is an emerging and contested practice. The meaning of the term ranges from the harvesting of cheap (free) labor to the intensive engagement with knowledge communities centered on complex materials. Crowdsourcing can refer to collective efforts at transcription; it can mean the application of human eyes and minds to non-textual problems that cannot be resolved by computers; the list of meanings goes on and on. At its root, the concept of the crowd suggests vast and undifferentiated numbers of participants. The crowd “sources” by generating resources, usually in a standardized format, through work in digital platforms. In turn, the digital platform offers a nexus where the raw goods to be “sourced” are served up and then returned after processing by the crowd. The term crowdsourcing can be dressed up in more elevated language, but the blunt description assayed above has the merit of calling things as they are. This definition begins to break down where the crowd becomes more active in its engagement. If the crowd provides the (re)sources and generates a community of knowledge and meaning-making, then crowdsourcing begins to shade over into something else. We will call this state of affairs community sourcing and stipulate that the role of the digital platform in this context is distinct from that associated with traditional forms of crowdsourcing.

Humanities. How best to define this term? Perhaps it is better to approach a definition indirectly. The term connotes a broad family of concepts and methods associated with the interpretive and critical study of human culture and society. Books and other texts are usually, but not always, involved. History, literature, and languages are the major fields of inquiry. The point is not to define this term, which is probably hopeless in any case, to but acknowledge how it sits rather uncomfortably alongside crowdsourcing. Everything about traditional crowdsourcing appears to be orthogonal to humanistic work in its traditional forms. Crowdsourcing is thin and undifferentiated whereas the humanities are deep and highly nuanced. The question: Is there a cat in this video? does not a humanities research question make.

And here, we arrive at research. What is humanities *research* and how might it relate to the emerging practice of digital crowdsourcing? Again, no explicit definition is needed. Rather, by pointing to the typical practice of humanities research, we can see what is at stake. Research calls up a complex field of activity that would include some if not all of the following elements: questions (hypotheses), theories, searching, sources, reading, debating, collating, revising, interpreting, selecting, narrating. Clearly, there is a lot going on under the heading of research. The crowd (or community) can weigh in on most, if not necessarily all, of these elements.

Crowdsourcing for humanities research, which serves as the title of our project and of this paper, is therefore an experiment in testing the boundaries of these three terms and reporting the results to the scholarly community. As an experiment, it has been designed to yield information of a comparative and, where possible, quantitative nature. This report proceeds in two parts. In the first part, we explain the rationale behind our original research proposal and describe in brief terms the critical amendments we made to the plan over the course of the study. A full account of the original projects is provided in the appendix

matter at the end of this paper. There and elsewhere in this paper, we provide these legacy materials in the spirit of putting all the cards face-up on the table. One of the chief limits to the accumulation of knowledge is the tendency to report only what worked in a given research project. Not all cards are trumps; knowing this is of paramount importance if we are to make progress. The second part of this report offers a detailed assessment of the three projects we conducted in order to experiment with different sources, crowds, and research questions. Additional data and analysis regarding the research can be found in the appendix.

## I.

### Original Study Design and Evolution

This project originated with a core idea and a common question. We believed that crowdsourcing had the potential to advance scholarship in the humanities. As new digital tools emerge, scholars engaged in digital humanities work have begun to make use of these tools to pursue research along new frontiers of knowledge, as well as to address traditional humanistic questions. As part of our initial planning process, we recognized the existence of a number of successful crowdsourcing projects with possible connections to humanities research – think Old Weather or the Transcribe Bentham projects. However, in spite of these instances of apparent success , there had been no systematic study of the process of developing and implementing crowdsourcing for the humanities at a detailed, project-based level. The state of the art had been well described in a meta-study by Hedges and Dunn, but the details regarding the methods and results of crowdsourcing for humanities remained largely hidden from our view.<sup>1</sup> Thus, the question at the center of our study is this: can we understand the process of crowdsourcing for humanities research and thereby assess its potential through careful experimental design aimed at providing a transparent and broadly comparative portrait of the affordances and limits of these methods? Put in blunt terms: does crowdsourcing truly offer an important tool for humanities research? Our answer, as will become clear in this report, is that experiments in crowdsourcing suggest that perhaps under certain conditions, and subject to several important qualifications, this is a tool worth deploying. Conversely, our study leads us to conclude that many research topics do not lend themselves to crowdsourcing, whether because they remain dependent on the kinds of close reading and individual analysis that characterize many humanities projects, or because it is prohibitive in terms of time or cost to recruit a crowd, engage with it, and harvest data of humanistic interest . Finally, our research suggests that most humanities crowdsourcing projects can enhance their chances of success by adopting a flexible stance toward the interests and needs of the crowd itself. When researchers view the crowd through the lens of community and seek scholarly engagement beyond the walls of the university, there is

---

<sup>1</sup> This isn't to say that work had not already been done exploring humanities crowdsourcing. There were clearly already projects out there – see, for instance, Transcribe Bentham. And work has begun to appear that examines the wider field, as the excellent work by Dunne and Hedges demonstrates (Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research, 2012). But our research through the literature and active projects suggested that no one had yet really tested these techniques with the question of whether they could actually be useful in mind.

potential for the creation of new forms of knowledge that derive from the interplay of researchers and collaborators in the public realm.

*The original grant*

Crowdsourcing offers an opportunity to advance humanistic work by applying distributed human ingenuity to solve problems that computers alone cannot and by fostering broad engagement with humanities materials beyond the walls of the university. In order to pursue these opportunities effectively, humanities researchers and their collaborators in interface design and dissemination need tested working models on which to build efficient projects. We need to know what works best with what kinds of material, and we can begin to answer these questions by conducting research on tagging novels, producing maps and metadata regarding environmental change, and collecting expert information and archival material from communities of interest.

We proposed three discrete research projects designed to highlight particular challenges and opportunities associated with crowdsourcing. The range of the projects were selected to be indicative of the methods we would employ to test engagement across a series of open toolsets for transcription, tagging, metadata improvement, and map warping in the interest of specific humanities research questions. The need for comprehensive and rigorous research along these lines is clear. Whereas a growing literature speaks to the use of crowdsourcing in academic contexts, no study directly and comprehensively addresses the range of *challenges* facing humanists wishing to enter into this space and harness the power of distributed effort via the internet.

Our projects were selected on two criteria. First, we considered the range of questions we wished to ask regarding crowdsourcing. In this regard, a selection was made to create a portfolio of research projects covering a range of topics, from tagging text in novels to identifying locations and providing metadata for historic photographs of railway lines. With this approach, we expected to be able to identify both the common tools and potentialities of crowdsourcing as well as the important differences owing to the kinds of questions asked and the sources consulted. Second, the projects were selected on the basis of their strong support in terms of faculty and institutional resources. Our ambitious goals and timeline required that the projects be sufficiently developed and staffed such that clear research questions could be posed and ample resources be put in place to make the most efficient use of the Mellon Foundation's support for our work. In addition to these criteria, we considered the degree to which each project could profit directly from collaboration with our Historypin partners and their expertise in web platforms and community engagement. In the end, our three original projects, as proposed, were:

1. Tagging 500 Novels (later called The Emotions of London),
2. The Year of the Bay
3. The Western Railroads Project (later called Living with the Railroads)

These projects met our requirements and provided enough diversity in content and public interest to make them excellent candidates for a comparative study.<sup>2</sup> From the outset we determined that there were potential tools we could use with the crowd. But there was a missing component: the crowd itself. Although humanities researchers have great expertise in their specific fields, and staff professionals working in the digital humanities have expertise in a variety of digital tool sets, most humanities researchers and technical staff have little expertise in interacting with the wider community as research partners. The project therefore sought to test the value of partnering with groups outside the university with expertise in online community engagement. The final form of the project came together with the establishment of a partnership between established humanities researchers (members of Stanford's Center for Spatial and Textual Analysis) and experts in the realm of interface design and web-based community engagement (Historypin). If humanities scholars in traditional disciplines, such as history and English, often lack the requisite skills and industry connections to develop and test crowdsourcing materials, it is equally true that without a connection to research projects led by established scholars, crowdsourcing in general, and interface design and community engagement strategies in particular, will remain on the periphery of humanistic research. This project brought together the vital inputs in the form of humanities questions and the necessary technical competence and global visibility required to evaluate the utility of crowdsourcing across a range of materials including text, tables, maps, and photos. We recognized from the start that even if we were to use these various tools, it would require a great deal of work to integrate them into our projects. That is, each project would have to modify and integrate existing tool sets to make them functional. This would not be a plug-and-play situation.

### *Tagging 500 Novels*

[

This sub-project, led by a research team from Stanford's Literary Lab, originally sought to extract and quantify character networks in 19th century fiction in order to investigate traditional literary hypotheses regarding the extent to which social settings (primarily urban or rural) correlate with social networks and their size, type, and density. More specifically, these hypotheses suggested that changes in novelistic form, plot, and character might be correlated to changes in real-world societal conditions, especially those involving revolution and industrialization. In rural settings, for example, characters tend to organize and interact in what might be seen as "familial" structures. The networks made possible by urban settings are different: In *Atlas of the European Novel*, Franco Moretti (1998) argues that urban settings tend to create character networks of a less dense and more superficial nature. Moretti writes that "the narrative system becomes complicated, unstable: the city turns into a gigantic roulette table, where helpers and antagonists mix in unpredictable combinations" (68). With these ideas in mind, we conducted a series of tests

---

<sup>2</sup> The name changes are important with regard to projects 1 and 3 above, and we will discuss the implications of these modifications in the detailed case-studies provided in the next part of the paper. Here, we wish to emphasize that these alterations in titles are an immediate indication that our projects changed significantly as we began to interact with the crowd. This points to a central finding, which we will emphasize throughout: if humanities researchers want to make use of crowdsourcing possibilities, they will often need to be flexible, in terms of both material *and* research questions.

with mode interfaces. These tests suggested that the problem as stated was too complex for crowdsourcing for two reasons: 1) complex narrative systems are too subjective and detailed – even trained academics could not come to satisfactory agreement; and 2) answers to questions about selected text snippets often required knowledge about the context of the passage that could not easily be provided to or assumed to reside in the crowd. Thus, we shifted our approach to the examination of emotions associated with particular locations mentioned in text snippets. Owing to their subjective and contextual nature, emotions associated with places provided a research question that could make use of the crowd.

Most humanities questions worth asking are at least moderately complex and require some degree of thoughtful commitment. They thus would seem more complex and to require more expertise than classic crowdsourced tasks, where task complexity, attention, time, and cognitive load are minimized. However, our work with Amazon's Mechanical Turk suggests that sophisticated research questions can still make use of material generated by these classic crowdsourced tasks. Although we had to change the kind of question we were asking, and the kind of data we collected, crowdsourcing still demonstrated itself to be useful and productive.

*Year of the Bay*

[

The Year of the Bay — 2013 in the San Francisco Bay Area — gave us an opportunity to test how to engage a broad public audience in crowdsourcing by tying one of our projects to a prominent regional event with compelling public and media opportunities. The Year of the Bay included the America's Cup, one of the world's top sporting events and the opening of a new Bay Bridge span which was an event that received a great deal of media publicity, but also events and exhibits at many of the region's historical and cultural institutions, including museums and libraries, which were interested in partnering with us in a regional effort to mobilize volunteers to rectify historical maps (mapcha), digitize shapes, and identify elements on the maps; place historical photographs and documents in the landscape, and identify entities in the photographs and documents, and thus improve the metadata and usefulness of these. From this, our research goal, driven by PI Jon Christensen's research on the environmental history of the San Francisco Bay, was to build from the ground up an environmental history of the San Francisco Bay Area. This project as initially envisioned, would rely on diverse datasets and will result in a new crowdsourced collection of digital, geospatially and temporally located objects that can be used by scholars to address crucial contemporary research questions in environmental history and humanities and historical ecology, including:

- 1) The original project sought to help uncover historical landscape conditions on a detailed scale, so that we can better understand the past, and the possibilities of these environments in a changing climate with looming water shortages. Ultimately these things happened, but not on the scale or quality that we had hoped.
- 2) 20th-Century Modernization and Environmental History. Our plan was to grow digital collections that would allow us to examine these changing relationships of resource use, recreation, and aesthetics and create a multi-layered history of the changing, but enduring relationships with the bay's environments, even as these

environments and communities experienced rapid modernization. Again, though some success was achieved here, it fell short of our initial plan.

3) The Possibilities of History, the Present, and the Future. Through this crowdsourcing experiment, we would be able to assess whether exploring and practicing the work of doing history — finding and interpreting historical sources and placing them in space and time — changes participants' understanding of the possibilities of present landscapes, and features that may be hidden from view, such as buried creeks, and the potential for future environments. This component never materialized.

We hoped the project would result in rich datasets allowing scholars to use these historical sources 1) to document complex changes in the landscape that have not been previously understood by the dominant narratives of environmental change, most of which have been written at a regional scale; 2) to understand human adaptation to environmental changes and how the landscape was used by different people over time (much of this complexity of quotidian human relationships to the changing land has not been captured by other sources); and 3) to interpret the ways in which changes in the landscape and changes in human relationships to the landscape were represented differently over time in the sources, particularly in the visual sources, photographs and maps. Though we did develop a rich archive of images, they do not seem to support these scholastic goals.

#### *The Western Railroads Project*

In 2009, Richard White commissioned a preliminary proof-of-concept study to explore whether it would be possible to use repeat photography to document the changing landscapes surrounding the railroad by replicating Alfred Hart's famous images of the birth of the railroad age in the West (<http://www.stanford.edu/group/spatialhistory/Visualizations/Hart/>). The project involved locating the modern site of the Hart photographs and retaking the photographs from as close to the same position as possible. The project was successful in that we were able to replicate virtually all the Hart photographs in California and Nevada, give them GIS identifiers, and plot them on Google Earth. But, at the same time, we recognized that repeat photography is both arduous and expensive. It provided matched pairs that provide evidence of landscape change (or the lack of it) and the social development, but it tells you little directly about the process of change in the intervening century and a half. The project inspired us to attempt a much larger, "open-source," project to collect, tag and interpret the thousands of photographs of railroads in the United States, which is why it was such a likely candidate for a crowdsourcing experiment.

Working with our partners in Historypin, the Hart collection photos were geotagged and inserted in the Historypin web interface which receives an average of 20,000 unique user visits per day.<sup>3</sup> Our ambition was to expand the Hart website in both depth and breadth.

---

<sup>3</sup> This for the site as a whole. As you will see, our specific Historypin site received far fewer viewers than 20,000 a day. We did, however, expect those kinds of numbers. Historypin has a number of project pages, as well as thousands of individual users. The point was that there

By expanding it in depth we intended to solicit photographs of the same sites that Hart photographed that fill in the century and a half between the original photograph and the repeat photograph. These would not be repeat photographs since they will not be taken from Hart's original vantage point, but since many of Hart's photographs are of cities or iconic landscape features we should be able to capture enough similar photographs to begin to record environmental and social change. By expanding the site in breadth, we intended to fill in the spaces that Hart did not photograph with other railroad photographs. However, finding such photographs was only part of the problem. We needed to locate them in space, and crowdsourcing could provide a solution to both problems.

Concomitantly, White's research has focused on the use of maps to reconstruct the spaces of settlement, environmental conditions, and conceptual elements of the space shaped by the railroads. Connecting photos to places provides the ability to connect two fundamental media (photos and maps) in a cohesive online interface. By doing so, this project planned to bring together two communities of interest with proven passion for their subject: railroad buffs and map lovers. We hoped that crowdsourcing would allow us to create a spatial database that we otherwise could not achieve on our own and open up the David Rumsey map collection to spatial, social, and environmental analysis that is now impossible except on a very small scale.

Realizing, even at the outset, that the railroad enthusiasts were a community that would need direct engagement, we reached out to the railroad enthusiasts by providing short stories and compelling calls for engagement for their blogs, newsletters, and email lists. We also planned to develop symbols of engagement such as badges that would symbolically reward and recognize their involvement in the project. This aspect of the research design, unfortunately, was never implemented. In part, our team came to recognize that the railroad enthusiast community was not likely to be motivated by such rewards coming from outside their own close-knit group. Eventually, we discovered that soliciting content and material from expert communities would require sustained outreach on our part, and so this project took on a longer duration than the other two. Though we did collect a great deal of material, to this point, we have yet to collect the metadata as envisioned in the original proposal.

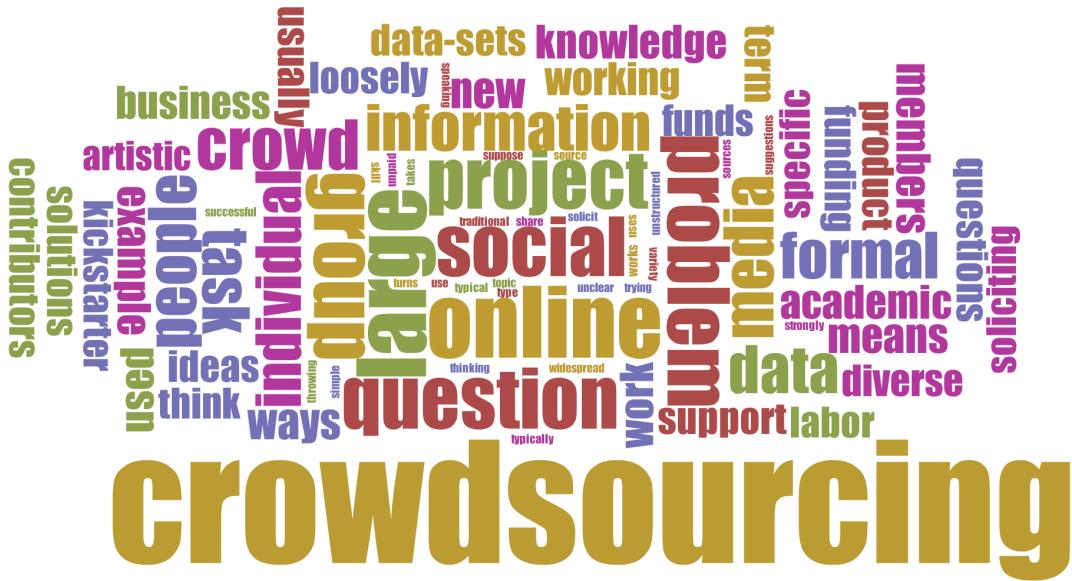
## II.

### **Concepts and Definitions: Knowledge Developed from Experimentation**

Before we delve into our own definitions of what the term means, it is worth considering how crowdsourcing is popularly understood in academic circles. In an admittedly ad hoc survey, we asked academics at a range of institutions and ranks, to weigh in on the question: what is crowdsourcing? The answers we received were instructive and even amusing in some cases. On the whole, the words our colleagues used to describe crowdsourcing suggest a complex set of ideas and a considerable degree of confusion, as the following word cloud demonstrates.

---

was a large population already using the site, and we hoped to parlay that into traffic for our site within the Historypin community.



The comingling of terms like “kickstarter” and “business” with those of “social” and “group,” indicates the way academics think of crowdsourcing both as something commercial and business-related and also something about groups and social production of knowledge. There are tasks, questions, and data-sets and these might relate to work and labor. The word community did not appear to any significant degree. There is a gap between the assumptions of the academic community, where crowdsourcing is more a mass/commercial endeavor, and our own experience in this project, where we have come to see crowdsourcing on a continuum with community outreach and engagement.

Between Crowdsourcing and Communitysourcing

Crowdsourcing and Communitysourcing may seem like different words for the same thing, but that is not the case. Though related, the former is generally a more extractive process, whose primary goal is to gather information or ideas from a group that may remain largely anonymous to the researching body. Communitysourcing, on the other hand, generates information and idea by engaging with already extant groups, or by growing new groups. These groups, often smaller and more local than the more amorphous crowd, require high touch and engagement. These later kinds of groups will often have a greater stake not just in the research topic, but in the process itself. Though we entered the project with some understanding of different approaches to crowdsourcing, what we were actually doing as part of each of our projects only began to emerge, or we only began to deeply understand the crowd/community dynamic, as we moved forward with the experiments.

How did our understanding of crowdsourcing change over the course of our project? As we calibrated our approach to each of the three projects, we were forced to confront critical definitional and conceptual questions. We learned by doing; and we learned by continuing to read the related literature on the subject. For instance, in his recent book, *Crowdsourcing*, Daren C. Brabham defines his subject as an “online, distributed problem-

solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals. Online communities, also called *crowds*, are given the opportunity to respond to crowdsourcing activities promoted by the organization, and they are motivated to respond for a variety of reasons.”<sup>4</sup> Here, the important thing to note is that crowd is neither left to its own devices, nor is it used as only a provider of information or mechanical work. Rather, Brabham makes a strong claim that for something to be called crowdsourcing, there must be an almost symbiotic relationship between the organization (or researcher) and the crowd. This distinguishes true crowdsourcing from things that have come to be called crowdsourcing: Wikipedia, open source software, and the like (which are driven bottom up, by the community), or a Mars poll to find a new color M&M, as control is held completely by the company. Though we agree with Brabham that precision of definition is important, and that the best of crowdsourcing projects are communal in the way he describes, we also feel that the things that have popularly been called crowdsourcing still count, even if they do not meet the definitional components he articulates. This is because we privilege the added value of community engagement and data collection as the most important elements of crowdsourcing, for humanities and social science researchers.

Mere definitions, as such, perhaps matter less than understanding the group you are engaging with and how you want to engage with them. Amy Sample Ward offers us a useful framework here.<sup>5</sup> She visualizes the difference between the crowd and the community in a set of concentric circles, centered on your organization. Furthest out is the crowd. You do not know them, and they “need consistent and compelling messages to take any action.” In the middle is your network. These are people you do not know directly, and cannot contact directly, but they are known to your community. The center circle is your community. These are the people directly connected to you, whether through attending your events, active on your social media like Facebook or Twitter, etc. This concept, however, might also be extended laterally, to include the older definition of community, as a group of people with shared interests and values that are consciously articulated.

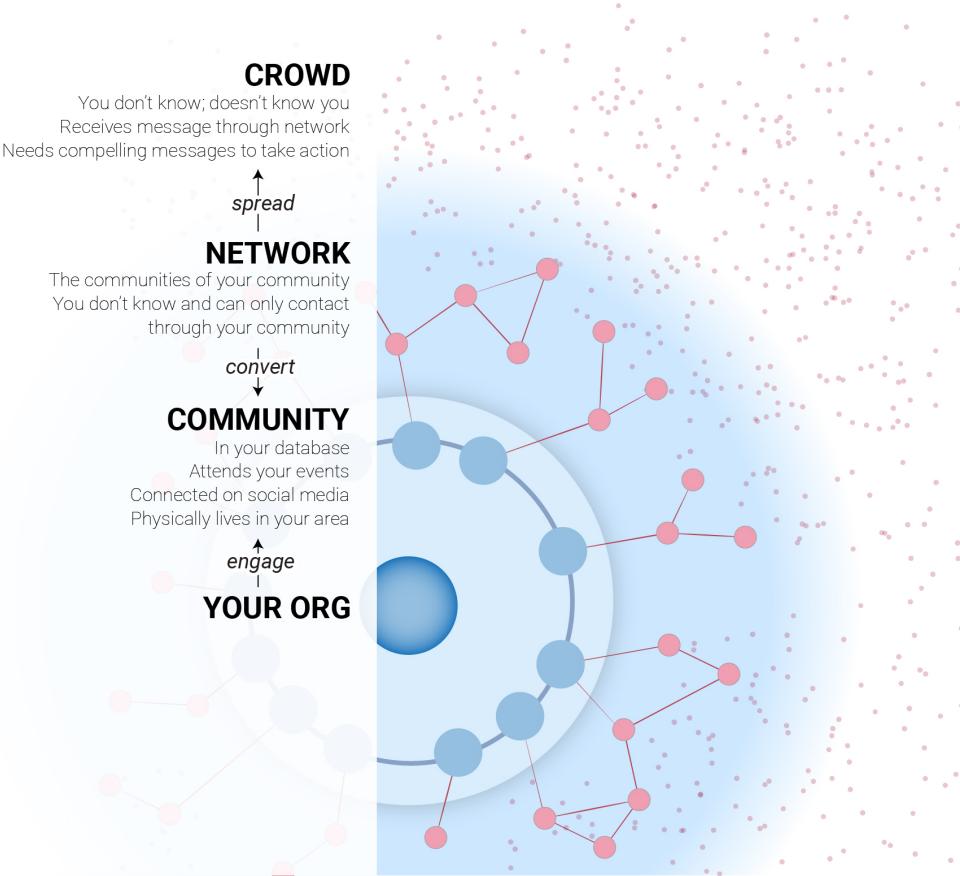
---

<sup>4</sup> Daren C. Brabham, *Crowdsourcing* (Cambridge: The MIT Press, 2013), p. xix.

<sup>5</sup> <http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>

**FIGURE**

**Relationships between the Community and Crowd**  
Adapted from Amy Sample Ward



Distinguishing the crowd from the community represents a valuable step and the process of distinguishing the one from the other matches our experience in developing the projects for the research reported herein. The distinction is valuable because it suggests two things, one having to do with engagement, the other having to do with the work itself. First, different strategies are required when dealing with the crowd as distinct from the community. Under our definition, a community is already in some way engaged with materials relating to your questions, and is, therefore, predisposed to engage with your project. The people transcribing Bentham probably already knew about Bentham and valued him. As such, this group corresponds more closely to the concept of community than to the crowd as an undifferentiated and anonymous mass. The community is focused on a specific person, issue, or place. In contrast, there are research questions that do not easily fall under the head of community interest beyond the walls of the university. If research involves questions and materials that do not have a natural community of interest extant out in the world, the work of engagement becomes one where the term crowd takes hold and alternative strategies are required. The crowd will need to be prodded to engage, whether through your message, the project itself, payment, or some other kind of incentive. What is more, given that the crowd will not have a deep well of knowledge and interest to draw upon

with regard to the research question at hand, there is also a necessary element of education or, at the minimum, scaffolding, required in order to garner useful input.

This may sound like it is not worth the trouble; why work hard to get people to do work when you may already have an interested group? Or, for that matter, why use a crowd (or community) at all? There are central questions a researcher needs to ask themselves, because sometimes you are better off doing it yourself. Though there are exceptions, if your task requires a specific skill set, particularly a skill set you have developed in your professional capacity, doing the work yourself, or using your resources to hire graduate or undergraduate research assistants will likely yield better results. If your tasks are smaller, are completed in short time segments, and do not require deep training, the crowd or community may be a route to explore. Additionally, engaging with a crowd in the way that Ward suggests offers a diversity of participation. It offers voices or data or analysis that you would not necessarily have access to. It encourages a breadth of participation that may let you ask questions that would not otherwise be possible. Remember, though, that even if your project is amenable to crowdsourcing, such a route will still require significant resources.

Second, you need to have different sets of expectations of the kind of work your groups will do—that is, what can your community accomplish and how is that different from what the crowd can do? The diagram, above, is useful because it suggests that the crowd and community are part of a larger ecosystem; they are not necessarily oppositional as much as they can be placed on a spectrum. Taking that spectrum into account is very important. If you misjudge where the group you are engaging with is found, you are not likely to generate useful material. This, the closer to the community you get, and therefore the closer to your organization, the more work people will do, and the more work they will be capable of doing. If your community, for instance, is a large group of academics, they will already have the skill set to do the analysis of the material you are looking for (assuming, of course, you have figured out how to motivate them to participate). They are predisposed to value both the topic and the mode of inquiry. If your crowd, however, is a group of people interested in, say, history more broadly, you will need to engage with them very differently. Tasks need to be simpler, or at least more straightforward. And they need to be the kinds of things that don't take your constant and active direction.

There is a real difference between crowdsourcing and what we can call communitysourcing. It is vital that you figure out which you are doing if you want to have a good chance of success. It is also likely, as indicated in Amy Sample Ward's schema, that projects will partake from a bit of both. These considerations have been offered here in the hope that they can help to frame the narratives associated with the implementation and assessment of our three experimental projects. We have already described their genesis. Our basic concepts regarding crowdsourcing for humanities research have been expressed. Now, we come down to cases. What did we do? What worked and what did not, and how do we know the difference between the two?

### III.

#### Project Reports

##### Year of the Bay

The Year of the Bay experiment was designed to test the experience of conducting crowdsourcing for the humanities in conjunction with major public events. As we have explained, our hypothesis was that such a coupling could garner widespread attention and enlarge the audience for a crowdsourcing project, leading hopefully to a new and useful research archive that would allow Jon Christensen – and others – to better understand the environmental history of the San Francisco Bay. The metaphor of a funnel is often used to describe participant engagement with crowdsourcing because many more people interact with most projects superficially (at the wide end of the funnel) than the number of people who travel through the funnel and engage deeply with projects. In this case our aim has been to widen the funnel as broadly as possible in order to increase the number of participants who engage at the wide end, and interact with as diverse an audience as possible and increase engagement proportionally. What we found was that the funnel metaphor was too simple – there is not just one funnel. Each of the groups we engaged with needs to be seen as its own funnel. In this way, the Year of the Bay would be better seen as a *set* of community sourcing projects that ran in parallel.

Our hook for the “Year of the Bay” was, literally, the year 2013, during which the America’s Cup races were to be held in the San Francisco Bay and the new span of the historic San Francisco-Oakland Bay Bridge was to be completed. Other activities taking place in 2013 and relating to our project included the 150<sup>th</sup> anniversary of the Port of San Francisco, the opening of a new home for the Exploratorium on the bay in San Francisco, and exhibitions at the California Historical Society and the Oakland Museum of California focused on the bay. Our strategy was based on tapping into enthusiasm for the bay and for local history and archival materials evidenced by campaigns to save the bay, and by several earlier and ongoing history projects and blogs in the area. Our goal was to collaborate with media organizations and museums, libraries, and archives to bring people to Historypin — through a custom URL at [yearofthebay.org](http://yearofthebay.org) — to engage with archival materials by pinning them to locations and providing other metadata, as well as contributing new materials from organizational, individual, and family archives. Our study had two goals: 1) that participants would help generate useful, accurate, and meaningful metadata for archival sources lacking metadata, and 2) that new archival materials would diversify and enrich our understanding of the environmental history of the San Francisco Bay and different cultural understandings and practices related to the bay. The history of the bay has long been dominated by a standard environmental narrative that we hoped new sources would enrich and complicate, adding to new understandings of environmental history.

##### *Summary of Activities*

The Year of the Bay project was launched in November 2012 on Historypin in conjunction with an event that brought media and institutional collaborators on a historical ship from the National Park Service to Hunters Point, an underprivileged, neglected, former

industrial and working class waterfront neighborhood where the ship was built nearly a century ago. This event garnered media coverage in the San Francisco Chronicle, KQED (the largest public radio station in the country), and on other outlets. It also established the project's identity with partners. At the time, the ability of participants to engage online at Historypin was limited to contributing new materials and pinning them on a map, a limit that continued until September 2013, when new tools for adding metadata to sources were added to the site. At the same time, a lively Twitter account was established, with a Facebook page added shortly thereafter.

In April 2013, Curating the Bay: Crowdsourcing a New Environmental History, an exhibition at the California Historical Society, opened and provided a significant new public launch for the project with more media attention from the San Francisco Chronicle, the Associated Press, KQED, Stanford News Service, and others. Curated by Jon Christensen from the historical society's archives, this exhibition was organized around the themes of the project, highlighting the gaps in archives, the questions they raise, and the potential for citizen participation in crowdsourcing history. A Historypin pinning station, with live displays of the Year of the Bay project, was a prominent focal point of the exhibition, which attracted 4,607 visitors in total through its 5 month run. Of these 4,607, 852 were brought through specific events associated with the exhibition, many of which focused on crowdsourcing. A Historypin project officer also staffed the pinning station once a week for several hours. In mid-summer, a San Francisco Chronicle columnist began regularly featuring the project on Sundays in the newspaper and on his blog online.

Across the bay in late August, the Oakland Museum of California opened a new exhibition, Above and Below: Stories from Our Changing Bay. That exhibition has also touted the Year of the Bay project and Historypin worked with the museum to conduct engagement activities during the exhibition, which ran until February 2014.

To this point, 7970 pieces of content — mostly photographs, but also some video and audio clips, and some miscellaneous scanned materials — have been added to the Year of the Bay Historypin site. Google Analytics for the site began in late April 2013. Between April 2013 and June 1, 2015, more than 18,700 unique visitors have come to the site, spending an average of 4 minutes and 41 seconds, while viewing around 19 pages, for a total of over 29,000 visits and almost 566,000 pageviews and 377,800 unique pageviews. Visitors were divided between new visitors and returning visitors (64% new visitors, 36% returning), and both visited about the same number of pages, between 18 and 20. But while new visitors spent 2:25 minutes on the site on average when they first visited the site (and 21 seconds on average on each page), returning visitors spent nearly 9 minutes on the site during each visit. Visitor engagement exhibited a nearly prototypical funnel pattern in visits, with ever fewer people spending more time on the site, except for a bump up again in the middle at the 3 minute range (about the amount of time it likely takes to familiarize oneself with the site). The pageview statistics reveal that after the 3-minute mark engagement increases dramatically, and continues to increase even as the number of visitors who stay on the site falls off steadily. This is the behavior we expected to see from the literature on crowdsourcing.

## Mellon Crowdsourcing: White Paper

### FIGURE

#### Session and Pageview Statistics

Year of the Bay Historypin Site, April 2013 - June 2015

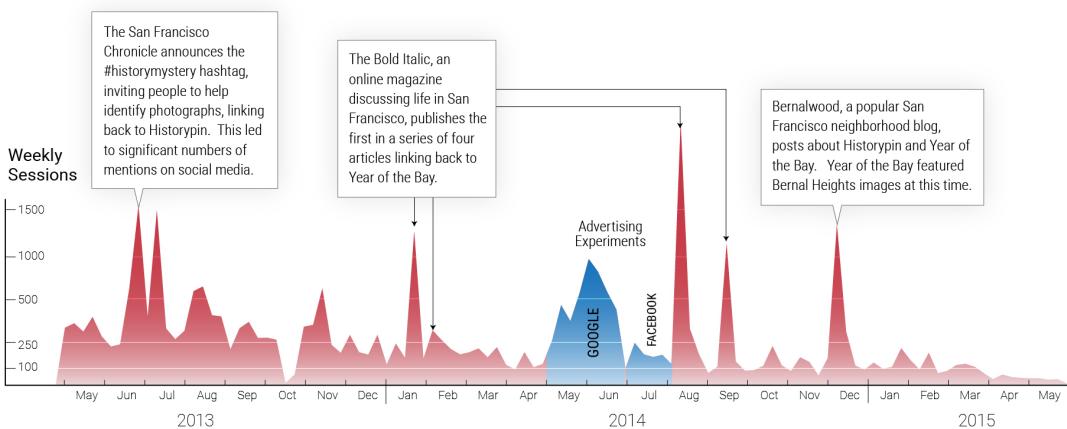
Session Duration	Sessions	Pageviews
0-10 secs	13,457	57,157
11-30 secs	3,546	29,197
31-60 secs	2,427	32,497
1-3 mins	3,818	94,221
3-10 mins	3,064	134,123
10-30 mins	2,016	124,864
>30 mins	970	93,906

The pageview timeline reveals that media coverage was a significant driver of traffic, with the largest peaks in traffic associated with appearances of Year of the Bay in the San Francisco Chronicle, in print and online. Other peaks were driven by events, such as a Historypin “hackathon” at the California Historical Society, Twitter (such as when the Lost San Francisco account shared a photo with its 20,000 plus followers), advertising through Google and Facebook, and by the introduction of “history mysteries,” in which coordinated efforts were made to share interesting items lacking metadata via social media, and people are encouraged to help the project solve particular mysteries. These efforts tended to focus on specific, highly engaged audiences and individuals and their social networks.

### FIGURE

#### Weekly Visitor Statistics

Year of the Bay Historypin Site, April 2013 - June 2015



This data – or at least the way we have reported it thus far – masks a vital part of the process of collecting it, or at least glosses over it. We only started to achieve significant successes in data collection when we began devoting significant resources towards

community outreach, both at Stanford and at Historypin. On the Historypin side, the addition of Year of the Bay Project officer Kerri Young significantly increased visibility and use of the Historypin site by the public. Through her active engagement with social media, and direct outreach to various communities, such as historical societies and blogs, both the visibility of the project and engagement with it grew. This addition went hand-in-hand with the addition at Stanford of a project manager for all the crowdsourcing projects, and the deployment of research assistants to engage in parallel outreach efforts with Young. This underpins one of our key messages from our two projects which involve active engagement of the public: without spending significant resources engaging with the crowd or community, your project will have a much lower chance of success.<sup>6</sup>

Examining, here, how the addition of a Project Officer impacted the Year of the Bay, is instructive on this point. While we did end up focusing on niche Bay Area groups and influential history bloggers from the outset, before engaging with these groups we spent a substantial amount of time experimenting with different outreach efforts. These ranged from personal meetings, phone calls, emails, use of social media, and the like. But our early efforts lacked coordination and focus. Even though outreach efforts such as press releases and magazine articles, aimed at more general audiences, were equally as important in boosting the project's visibility and providing credibility to our outreach groups, we realized such efforts required constant social media, a role that could not have existed independently of the Project Officer. Though we did not have substantial technical resources and manpower to administer regular in-person events such as pinning sessions, the Project Officer planned and administered these whenever possible. All of the above tasks, providing valuable user to researcher interaction, unfortunately did not always align with Year of the Bay's second-year goal of getting as much content as possible; we learned over time that building new relationships, and developing existing ones, is a slow process requiring the building of trust combined with the right timing.

The primary tasks of the Project Officer were:

1. Experimenting with and implementing different outreach techniques with colleagues at CESTA (drawing people to the project, getting them started).
2. Administering small live events whenever possible (mystery-solving, drop-in pinning, etc.).
3. Acting as one of the main resources for troubleshooting on behalf of Historypin.
4. Administering social media (including online campaigns, sharing new content being added to the project, etc.).
5. Providing qualifiable and quantifiable data for all above activities, again in conjunction with CESTA.

What this list should indicate is the amount of time such a role takes. Although we will describe this in further papers in more detail, we want to emphasize that this aspect of crowd/community-sourcing is time and labor intensive. For those imaging that such

---

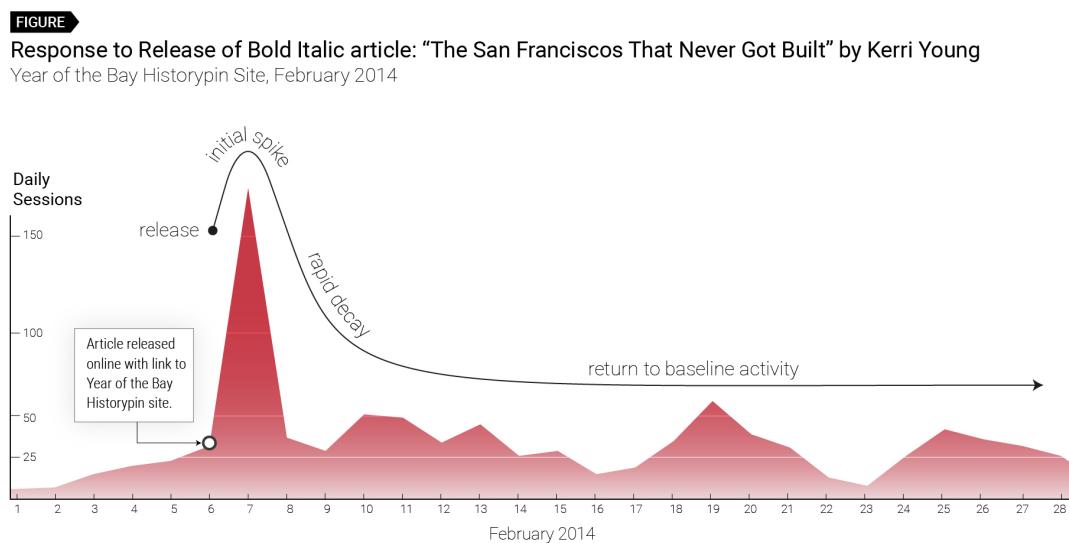
<sup>6</sup> Success can be measured in a variety of different ways, ranging from gathering usable data to engaging the wider community. But we found that no matter which metric you choose, you will have to cultivate your crowd. Just building it is not likely to get people to come and engage with your project.

projects will save labor or go quickly, we seek to dispel that belief. Though ultimately rewarding in terms of the collection of material – and ongoing engagement with the project – this took the time of a dedicated member of the project on a daily basis to maintain the visibility of the Year of the Bay, and to draw in new participant groups and individuals, all of which was central to the growth of data collection.

Drumming up positive media coverage, as well as social media outreach, was very important in raising awareness of our project to more diverse audiences. For example, after the release of the [Stanford News Service](#) story, Associated Press [story](#) on our crowdsourcing collaboration with the California Historical Society, and SF Chronicle writer Carl Nolte's [weekly calls](#) for a few months to visit Year of the Bay in his popular blog, the average number of homepage views rose 500-1000 more than the daily average. Also, in conjunction with Nolte's posts for Year of the Bay, doing bulk uploads of SF Chronicle photos to Year of the Bay also drove more traffic across the site.

In another attempt to reach more general audiences, the Project Officer collaborated with the popular local magazine, *The Bold Italic*, early in 2014 to create posts featuring the history of individual neighborhoods in San Francisco (also part of our campaign to solve local mysteries). These posts used photos pinned to our project by the San Francisco Public Library, which graciously shared higher resolution photos for this purpose. Young also wrote an article about a plan for San Francisco that [was never carried out](#), again using the opportunity to share SFPL photos pinned to the project. After the release of these posts, as with the Stanford News Service and Associated Press stories, there were large spikes in traffic to the Year of the Bay project pages. While there was positive social media engagement in the form of hundreds of likes, comments, and shares about these articles (over Year of the Bay and *The Bold Italic*'s social media feeds), this did not translate into significant participation within our project. We will come back to this later.

*Google Analytics for the Year of the Bay project, showing a large spike in visits on February 6, 2014 after The Bold Italic published Kerri Young's article "The San Franciscos That Never Got Built."*



Another valuable method of outreach was curating guest posts from active pinners, and sharing them over social media. For a project as broad as Year of the Bay, sharing examples of what others contributed within our project was valuable in demonstrating to potential users what they themselves could pin. While these blog posts included new archives sharing content in Year of the Bay, especially highlighting stories and content from individual users was important in demonstrating that anyone, not just local archives and museums, could contribute stories valuable to our project.

Also supporting all of the above activities, Young tried to maintain a strong daily presence over social media with examples of good pins and pinning practices. Highlighting pieces of content already pinned to the project was a good opportunity to target other groups or individuals over Facebook and Twitter, and an alternative to initially reaching out by email. Using local events and anniversaries as incentives for sharing memories (the opening of the new Bay Bridge, the last 49ers game at Candlestick Park, Outsidelands Music Festival, Loma Prieta, etc.) was a good way to ask for specific kinds of stories within a broad Year of the Bay project. It is true that posting on big moments in the history of the Bay or famous landmarks drove the most traffic to our site, but watching even a few users engage around more locally-based histories, or topics that trigger recent memories, have been just as valuable in making such topics more discoverable. Over the course of the project, daily social media outreach certainly increased engagement in the form of retweets, shares, and comments, but did not necessarily garner participation on the site. For our #FarewellCandlestick campaign for example, we received photo memories over social media because we ourselves offered to pin them to our archive, which significantly lowered the bar to participation by not making contributors to go through the pinning process on Historypin.

However, social media outreach is extremely time-consuming. Having only one part-time Project Officer meant that not enough effective time can always be dedicated to it; with this in mind, in future, we believe a better use of a part-time Officer's time is to be reaching out to and fostering relationships with and between known community leaders, rather than focusing too much time on social media experiments. This is not to say a similar project would not need to do outreach experiments in the lead up to the launch of their project: getting to know your audience is a vital component of crowdsourcing, but so is spending your resources wisely. If you don't have a full time officer, you will need to spend time cultivating the community. That said, we have four recommendations as to effective, and ineffective, ways to directly communicate with the crowd.

#### *Four Types of Direct Engagement*

##### Pinning Events – Public events with specific Community Groups

Over the course of the project, we held a few pinning events dedicated to particular community groups. Sessions with the Bernal Heights History Group demonstrate what worked and did not work for general pinning events, as well as mystery-solving events (later in this section). Bernal Heights History is a small group of mostly older local history enthusiasts who meet once a month to discuss the history of Bernal Heights, a quiet neighborhood on the southern edge of San Francisco.

With the Bernal Heights History Group, we had a community power contributor in the group's lead, Vicky Walker. The Project Officer was able to easily reach out to her and begin to plan an in-person event. She was able to provide a community space (the local public library), and send the word out to the rest of the group about the details. However, even if a community leader such as Ms. Walker is familiar with Historypin, we cannot assume that the rest of the group is. So the Officer will have to spend time explaining the project. This kind of introduction, combined with the pinning activity itself, requires very effective time-keeping.

Other time issues emerge in relation to the groups you are engaged with. Small community groups are often not technically savvy. Having the space and group organized and ready to meet is just one aspect of the event-planning process, one that the community leader can provide. But without technical resources, such as a projector, scanner, or even internet connection, it is usually not at all effective to conduct pinning activities. So you will need to provide the technology, if your project requires that component as ours did. With that in mind, having alternative methods for sharing project content are a must, even if it is simply having analog community forms for participants to fill out amongst themselves.

In addition, even with the invaluable help of a dedicated community leader helping to organize a small pinning event, technical as well as time constraints usually leave pinning itself trumped in favor of discussion. The latter is absolutely the kind of outcome we want for a community session, but the original intention of pinning photos during an event usually falls short as it disrupts the flow of a community gathering of friends who would rather talk amongst themselves. With community groups such as Bernal, these are people who regularly meet and are wholly comfortable with one another. This makes for a very informal atmosphere. For the Project Officer, this is very good in that group members will not have a problem having a meaningful discussion with one another. On the other hand, trying to maintain a set-schedule of event activities usually never goes as planned, as members would rather have a group discussions over single pieces of historical content rather than share in groups.

To help mitigate this issue, being able to gain access to the group beforehand on behalf of the community officer, in order to remind participants what to bring and make yourself available to questions, is good practice and helps avoid confusion on the day of the event. One of the biggest issues for the first pinning session with Bernal Heights History was that most of the group's members had not heard of Historypin before, and thus we spent much time explaining what our organization was rather than collecting photographs. Preparing group members beforehand with group emails, links to our site and project page, etc., again can greatly help to save precious time on the day of the pinning session.

#### Pinning Events – Drop-ins at a regular location

As part of our collaboration with the California Historical Society (CHS) on their exhibit [Curating the Bay: Crowdsourcing a New Environmental History](#), we experimented with having drop-in pinning sessions each week at a small scanning and computer station. We marketed it with CHS as an opportunity for exhibit visitors to bring in their Bay photo

memories to scan during the lunch hour every Wednesday, where the Historypin Project Officer would be stationed to help.

The weekly drop-in pinning sessions for Year of the Bay at the California Historical Society were only marginally successful in terms of gathering content on the day, but fruitful in terms of generating conversations about the project and the ways in which people can contribute at a later date. It was a difficult ask for walk-ins to bring in physical ephemera to scan; weekly calls to contribute over social media, in CHS's email newsletters, a special Curating the Bay tout on the project's main page, and even targeting individuals with personal calls produced only a couple of walk-ins with physical materials over the course of the exhibit. Also, technologically speaking, the pinning station was often unusable because of the problems with Historypin showing up in the pinning computer's various browsers (we often used laptops as alternatives). More training of exhibit staff on Historypin itself, moving drop-in days to the weekends, calls for themed-contributions, decorating the windows of the building with calls to contribute, and preparing visiting tour groups before arriving are all contenders for improving this process, should you undertake it.

However, there were both walk-ins who either knew or did not know a Historypin representative was going to be present, who simply wanted to learn more about the project, platform, and whether or not the material they possessed in their home collections were appropriate contributions. Due to the broad nature of the Year of the Bay project, the latter question came up frequently. Also, it was unfortunate that there were no references to yearofthebay.org in the main Curating the Bay brochure, so many people did not know that they could visit the page outside of the exhibit. That said, from the perspective of a Project Officer, it was wonderful to have a meeting space to talk to members of the public weekly about Year of the Bay and Historypin, as well as to see that many of those who came in just to talk later went on to contribute from home at a later date. That is, you may want to take time to have a regular schedule for interaction with the public.

The outcome of this experiment can partly be attributed to the fact that this project's success relies on building trust from both individuals and institutions before they can contribute, and this takes many months after the project launch to gain traction (and more than the couple of months the exhibit was open). Briefly mentioned above, if we were to try a similar experiment in the future it is absolutely worth trying to develop engagement strategies around guaranteed regular visitors such as local tour groups and school groups. Groups like these are regular visitors to museums and historical societies, and often book in advance for a guide. Knowing this, it is not unreasonable to spend more time asking these groups to prepare contributions BEFORE their visit, and to develop strategies around effectively collecting such content.

This aspect, then, must be seen as a failure in terms of crowdsourcing. It further demonstrates the importance of cultivating active and involved persons. Though some crowdsourcing projects can derive benefits from work being done in dribs and drabs, a project like this one cannot rely on, or even make significant use of, unprepared visitors.

#### Mystery Solving – Gamification of your materials

Gamification has become a popular term not just in crowdsourcing, but across the world of digital education. At base, it suggests that by adding a game component to your site as a mode of engagement – creating leaderboards that measure some sort of activity, offering rewards, generally using digital game elements in your educational space, can entice both greater levels of participation, as well as greater levels of learning. We sought to test its efficacy in crowdsourcing by creating a new component in our Historypin interface: the “Mysteries” tab. Even here, however, we did experiments to gauge interest. Before the launch of Year of the Bay’s mystery solving interface in November 2013, each week we picked photos pinned to the project that needed more metadata, and released them over social media as a #historymystery. This gave us the chance to reach out to specific local groups and individuals to collaborate on mystery-solving, which in turn helped spread the word about our project to communities that otherwise may not have come across it. An example of one of the most fruitful collaborations around mysteries has been the instances of the [Lost San Francisco](#) Facebook page sharing our mysteries with its 20,000+ followers. As Historypin is a global entity, being able to tap into a community specifically geared toward local history, and in Lost San Francisco’s case local nostalgia, proves more fruitful in terms of user interaction over social media. A spike in views, shares, and “likes” in the hundreds, as well as dozens of meaningful comments around our Year of the Bay photos characterize these instances.<sup>7</sup>

---

<sup>7</sup> We cannot underestimate how much having an institutional partner in CHS helped during this early process of mystery-solving; during the time of our collaboration for *Curating the Bay*, they worked with me in sharing every mystery that we shared over social media with their own networks to maximize the daily reach of the posts. This helped give our project more credibility to those who were not familiar with us, as well as helping us to gain new followers and interactions. With future collaborators, laying down the framework for a social media strategy with them is very important in doing reciprocal outreach that will benefit both parties.

 **Historypin**  
Today's Year of the Bay #historymystery is an incredible photo from of a house moving-quite literally. Here is what the SFMTA | San Francisco Municipal Transportation Agency Archives know about it: "...when Alma de Bretteville Spreckels built the Spreckels mansion at 2080 Washington St., she insisted that several houses be moved from the property to other locations instead of tearing them down. Th... See More  
[Like](#) · [Comment](#) · [Share](#) · [Edit](#) · August 5, 2013

 Carmen Franch, Ken Wagovich, Janet Harper and 66 others like this.

 141 shares

 **Historypin** The photo on Street View: <http://bit.ly/12V0xi7>



*Example of how “Likes” and “Shares” skyrocket on our Facebook mystery posts when [shared by Lost San Francisco](#).*

Here, we see the breadth of outreach that is possible when connecting with already extant social media groups. In this case, by getting “Lost San Francisco” to share out posts, the post got 69 likes, and 141 shares. This is an increase of orders of magnitude from shares and likes had we posted the image and post to our Facebook page alone. In addition, this outreach resulted in an increase not only of likes of the image, but a spike in people following our Facebook page. The lesson here is that if you can find allies with an already established presence, you stand a higher likelihood of getting your material to more eyes.

## Mellon Crowdsourcing: White Paper

**John Barnes** Amazing photograph! I worked for a house mover for a year and never saw something like this!  
August 5, 2013 at 6:37pm · Unlike · ↗ 1

**Bruce Costello** I remember years ago we used to see lots of houses being moved lots of places. That practice began to fade around the late 60's.  
August 5, 2013 at 6:38pm · Like · ↗ 1

**Bruce Costello** Think it mostly ended about the time someone bought a handful of houses and moved them to a vacant lot off of Edgar st, and never installed them. The ruins are still there to this day... used to be a very nice lot, too, for flying kites and just looking at the view.  
August 5, 2013 at 6:47pm · Like

**Katie Amos** John Cunningham  
August 5, 2013 at 9:08pm · Like

**Scarlet White** Amazing how they did this!!  
August 5, 2013 at 9:14pm · Like

**Steve Gates** They would have had to rotate the house because of all the telephone poles, etc. What a massive undertaking.  
August 5, 2013 at 9:14pm · Like

**Thomas Rubarth** Neat to see the trolley "hole".  
Here in Omahawww some of our houses did not move down the street ...they simply moved down. The steepest hills downtown were shaved-off to some degree. Several blocks of existing homes and churches were lowered.

*Example of the numerous meaningful comments for our pinned content when shared with popular nostalgia communities like Lost San Francisco.*

Again, we see the benefit reaching out to large, engaged communities. Not only did we get more comments than was usual, but they often contained more content. In this case, the comment by Brice Costello added data about the movement of these houses, and what happened to the land. This is the sort of information that would be useful in terms of the general research questions for the project.

But this also highlights one of the difficulties we encountered with the project. We were able to find communities that were engaged with our images, but those communities, and the cyber-spaces they inhabited, were not on Historypin. For this data to be useful, ultimately, the relevant comments would have to be manually copied and pasted to the image on Historypin. This remains a challenge. People are comfortable in their existing communities; getting them to add more communities, or, even harder, to double post to Facebook and Historypin, is a very high barrier.



*An ideal for mystery-solving over social media: Groups that we target to solve mysteries in turn spread the word through their channels.*

As an example of what is possible, our initial posting of the shipwreck image (above) was titled “Wreck of a lumber schooner. Remains of old wreck in foreground, San Francisco, Cal.” By posting this to our Facebook and Twitter feeds – with re-posts and re-tweets -- we were able to determine that it is likely the wreck of the William L. Beebe, below the Cliff House, in 1894. Three Twitter users were able to point us to *San Francisco Call* articles from December of 1894 with drawings and articles that match the image. So, on the one hand, this was a successful use of crowdsourcing. We should note, however, that although a number of mysteries were solved, only a few users contributed to this process. Most were solved by various project team members, or a few regular users. Moreover, you have already seen the flaw in the process; the mystery was solved not via our Historypin platform, but through the social media platforms. Moreover, though this process can be effective for individual images, it would be time consuming to seek to identify large numbers of images in this manner.

### Hackathons

In our time collaborating with CHS we also tried organizing a large hackathon style event in their museum-like space. Like our community group mystery-solving, this hackathon also benefitted from having high-resolution images shared via a projector. With the ability to use CHS’s space for the event, we set out to bring both hackers and non-

hackers alike to solve photo mysteries of the SF waterfront. Beforehand, we asked visitors to bring laptops, books, and other materials that they could use for research during the night. For event content, we curated a handful of waterfront photographs from our partners at CHS, SFPL, SMTA Archives, and SF Maritime.

After targeted invitations to known Bay Area mappers and hackers (e.g. Burrito Justice), a group of about 50 people attended, a mixture of personally invited vs. open invites to the CHS member pool. Going through each photo one at a time on a projector, this mixed group worked collaboratively to solve date, location, and business information mysteries within them. Assigning tasks, collaborating on finding citations, and using a mixture of digital and analog research devices to look up information were some of the things we were able to adjust and experiment with to best collect data.

The results of a [hackathon survey](#) that we sent out to participants after the fact gave us some insightful feedback.<sup>8</sup> Participants expressed a desire to be able to browse the photographs on their own devices at their own pace as opposed to on a projector, to break into groups to research certain parts of a photograph (not enough time for this), as well as having more available research tools at their disposal. They did however like how they could come and not only consume information, but contribute to the conversation, and a major incentive to attending was being able to see photographs rarely or never-before-seen. This suggests that engaging the public with well known images is less likely to generate the buzz you will want in a project like this one.

This event demonstrated how a specially invited audience can clash with open invitations. Those who had been to a hackathon before (most of our special invites) came prepared with laptops for research purposes, and many of those unfamiliar with this type of event did not. Those with laptops, the fastest methods of research, spearheaded the task of looking up information as questions came up, which alienated some of those without them who wanted to do the same. Because time during this event was at a premium, we did not have enough space in the program to break into groups to do research on each photo, an activity that could have solved this issue and made the event even more collaborative.

#### *Conclusions – Year of the Bay*

Through Year of the Bay, we have come to see crowdsourcing as consisting of individual activities, engaging with questions, challenges, and materials. Sometimes this interaction occurs alone, but more often it takes place in the context of interaction with others, on a website or through social media. It may also take place in person, face-to-face, and at formal events. Using crowdsourcing on complex historical source material beyond the archives, usually preserved in libraries and museums, will likely depend on very high-touch, directed, sustained outreach efforts to specific individuals, community groups, and organizations, requiring sustained support to engage in these efforts. It will likely not be substantially less work than the very high-touch, sustained efforts required to collect significant physical archives.

---

<sup>8</sup> The sample size on the survey is too small to make any strong claims, but, anecdotally, they fit with that we have learned.

We have seen that a crowdsourcing project conducted in conjunction with high-profile community events and media coverage can drive significant additional traffic and engagement. Traffic on the Year of the Bay Historypin site increased dramatically when the project was featured in local media. We know anecdotally that this brought some people to engage more deeply with the project and contribute materials. We also know that a constant social media presence, made possible by the work of a full-time Historypin project officer, and regular new “history mysteries” can drive significant traffic (although not as much as media) and even more significant engagement with the site. We suspect that this pattern may be able to continue even after the media coverage for Year of the Bay subsides. The initial coverage helped increase awareness of the project — the “brand” as it were — but the momentum depends more on finding ways to engage with particular communities of interest and deeply engaged individuals around specific calls to action.

In terms of partnerships, we found that there was value in working with the California Historical Society, for the project, as a venue for events, and a physical focal point for the concept, but also in helping the society begin to work toward new models of accessibility for its archives and public programs. But this did not come without difficulties. In addition to 4,607 visitors, only 162 of whom were current members, the Year of the Bay exhibition brought 112 new members to the California Historical Society, more than any other recent exhibition, including a blockbuster exhibition on the 100<sup>th</sup> anniversary of the Golden Gate Bridge, which brought more than twice as many visitors to the society’s galleries. This is, perhaps, one measure of the ability of crowdsourcing to engage new audiences. On the other hand, the fact that the society was only able to contribute 47 digitized images from its archives to Year of the Bay — while mounting an exhibition of more than 150 items, many never before publicly displayed — is an indicator of the difficulties of working with a traditional archive on new digital projects. For the society, this was a useful, though at times wrenching process, highlighting gaps in its capacity, which it is striving to fill.

There is also the question of where material comes from. Substantial archival materials have come from other partners. The San Francisco Chronicle contributed greatly to awareness of the project, and also contributed substantial archival material, although it required additional support from an intern to digitize materials, highlighting capacity gaps even in a large organization focused on digital material, at least in the present, if not the past. The San Francisco Public Library has made substantial contributions of materials to the project, almost all of which lack metadata. One of our biggest contributors, the San Francisco Metropolitan Transit Authority, was already a Historypin user. Other partners have not contributed as much material as we had hoped, and it has become clear that many organizations, including even some of those most eager to engage in digital projects, will require substantial, time-consuming support in order to contribute meaningful, engaging material to these projects. Indeed, Jon Christensen has raised questions about the efficacy of crowdsourcing to projects like this one, suggesting that the material we collected has proved to be of limited use for the researcher.<sup>9</sup> At a minimum, then, the Year of the Bay project suggests that such projects will be a significant amount of work and time at best, and bring no guarantee of success.

---

<sup>9</sup> Jon Christensen, “Reporting a Negative Result in Crowdsourcing” *The American Historian* (May 2015)

## Living with the Railroads

Following the work of Richard White, Living with the Railroads sought to use crowdsourcing to learn more about the social, cultural, and environmental impact of the development and expansion of the railroads in the American West, and across the United States (White, 2011). The crowd in trains is neither the more amorphous crowd of The Emotions of London, or the crowd of those generally interested in the San Francisco Bay Area of Year of the Bay; rather, the crowd we seek to engage in Living with the Railroads is that of railroad enthusiasts. This project, therefore, had to explore and lay the groundwork for interacting with this specialized crowd in a different manner than either of the other two projects. Instead of interacting with a crowd where expertise lies on the side of the researchers (CESTA) or the content aggregators and community builder (Historypin), in this instance it is the crowd who has the expertise. This draws a different set of challenges, including how to approach the crowd, what will engage them, and how to get them to stay connected (Dunn & Hedges, 2012).

The goal, in conjunction with Historypin, was to get railroad enthusiasts to upload their data, identify images and documents whose provenance is unknown, and help facilitate the growth of connections between enthusiast and enthusiast communities. That is, this project is not just seeking data from people, but hopes to get people to see the Historypin site, and partnership with Stanford, as something to their benefit, with the side-effect for CESTA and Historypin being data collection. Spring and summer of 2013 was spent building a working group of PIs, RAs, and Train Enthusiasts to figure out the best to solicit broader enthusiast participation, as well as working with Historypin to develop the website. Thanks to this work, and a significant amount of research on the part of TAs in collecting material for the site to add to the content that Historypin already had, we were able to launch the site with a significant amount of material, and we continue to add material to it, as well as to maintain our engagement with the enthusiast community.

### *Summary of Activities*

As a reminder, the core intent of this project had two components. The first was to expand the project based around the Hart collection photographs to better understand the processes of change – social, environmental, cultural – spurred by the rise and expansion of the railroads in the American West, and beyond. The main idea was to crowdsource both images as well as their metadata. Originally, we planned to parlay Historypin's expertise in soliciting material from cultural content partners to populate a Historypin site with an initial set of material, which would then be published to the expert community who would geolocate the images and add any known metadata.

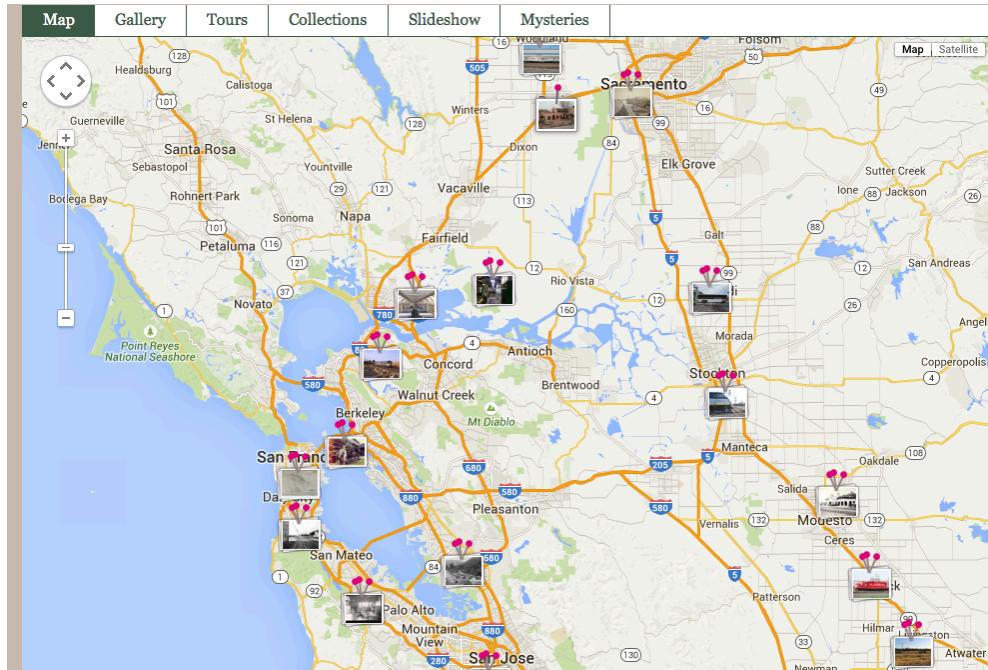
A second component centered around maps. Richard White's research has used maps to "reconstruct the spaces of settlement, environmental conditions, and conceptual elements of the space shaped by railroads." Following the success of his work, our plan was to offer the Rumsey map database, in addition to other maps collected by our researches, for crowdsourcing. We would geo-rectify the maps in the hopes that the expert public would locate details and images on the maps.

As indicated, in order to get this project going, groundwork had to be laid. Drawing on connections from both Historypin and CESTA, starting in April 2013 we began to have regular meetings with what we called our Crowdsourcing Trains Advisory Board. Though initial meetings with varying members of this group started as early as Fall of 2012, the group as a whole was not formalized until 2013. This group consisted of CESTA Staff, RAs, Historypin, and a 3-4 local railroad enthusiasts. This group helped the project drivers to understand the kinds of people they would be interacting with, and offered guidance and suggestions as to how best to interact with them, and as to what kinds of materials and what kinds of website functionality they would respond to positively. Through these meetings, two key things emerged. First, some enthusiasts would be interested in partnering with Stanford and Historypin if their data was secure and would be available over time. This led CESTA to make contact with the Stanford Data Repository, and we are currently in conversation with them to figure out how to house and curate data from this project. We hope that we will be able to make Stanford as one of the main repositories for digital materials on trains in the United States.<sup>10</sup>

But data security and longevity of curation is only one of the enticements that would encourage railroad enthusiasts and experts to participate in this project (it is important to re-emphasize that in this project the idea of participation, rather than CESTA collecting data, is central.) Crowdsourcing with expert communities requires establishing trust, as well as gaining entrée through personal connections. This makes the Railroads project doubly complex, and any project engaging with expert communities will need to keep this in mind. We also determined that in order to stimulate interest in the project we need to open it with data already present. That is, we cannot offer them a blank map and ask them to fill it in. In support of this, starting in April 2013, CESTA expanded its RA staffing. CESTA RAs did extensive work over the Spring and Summer collecting materials – maps, images, timetables, etc. – so we could prepopulate the Historypin site (they also worked on creating an extensive database of museums and railroad societies, and starting a test blog so that outreach can take place quickly once the site is fully launched). Rather than research broadly, we focused in on the area between San Jose and Sacramento. This led to launch of the site in what we hoped was an appealing form to the community.

---

<sup>10</sup> It became clear to us early on that, even were we to surmount the technical issues of integrating the maps interface with our Historypin interface, there was not a lot of interest on the part of the Railfan community in map analysis. Both our meetings with our advisory panel, as well as preliminary outreach efforts, suggested that we would stand the greatest chances of success if we concentrated on the image portion of our plan.



A current screenshot of the San Jose/Sacramento region. Note that multiple pinheads indicate that there are more than one image at those locations. In this case, there are hundreds of pins in this region.

But our advisory board also indicated to us that what we were offering, and what the community wanted, were not the only important things if we wanted to the project to be a success. These meetings indicated that in order to gain access to the community, we needed to do 3 things. 1. Present ourselves not as the experts, but as facilitators. If we approached these communities as closed off Ivory Tower researchers, seeking only material, we would find similarly closed doors. 2. We needed to get a sense of how the community would engage with the material we would offer. 3. We needed to offer something, whether material, platform, or both, that didn't already exist. The first point is clearly less one of material and more one of attitude. It governed our interactions, for instance, as we engaged with the rail enthusiast community on October 4, 2013 at the Southern Pacific Historical and Technical Society, which was the next step in our process.

#### Running a testblog – Laying the groundwork for success

One should not engage with expert communities blind, or without testing. So we need, here, to step back for a moment and look at the work we did to try to ensure our success at this meeting. Prior to our first major engagement, we made efforts to gauge the interest of the enthusiast community prior to the creation of the Historypin site. One major step was undertaken by one of the RAs, Sophia Paliza-Carre. She generated a test blog that she ran from June 24, 2013-July 24, 2013. The overall goal of this blog was to get a general sense of what kinds of material railroad enthusiasts were interested in, and what kinds of questions would prompt them to begin information-seeking discussions. Within this larger question she also sought more specifically to see if they would respond better to people-oriented photos or to landscape photos, whether more famous photographs were of distinctly less interest, and whether the time period of the photo mattered greatly. Finally, the hope for the blog was to see how difficult it would be to sustain interest and discussion about crowdsourced materials.

In order to test the interest of railroad enthusiasts in photos with only partial metadata, she initially put up 5 photographs posing a few introductory questions mostly asking about the details in the photos. Viewers were also asked to comment on anything in the photo. An example of questions is here:

*Please help us with our questions about this photograph, feel free to comment with any information:  
What is the purpose of this fence? What plants are pictured here? Are they native to California?  
What kind of change in the landscape can you see here as a result of railroads?*

The blog was created June 21st and began receiving views starting June 24th. Robert Bowdidge, one of our advisory panel members, kindly helped spark interest in the blog by sending it to and posting it to the following groups and crowd numbers. This, in fact, was an important data point. Robert Bowdidge is a member of the enthusiast community, as well as part of our advisory board. Having him, in effect, introduce us led to immediate acceptance. His introduction worked to vet us, sanctioning us as legitimate. Without this sort of exposure, the blog almost certainly would have received very little attention. And his help in putting the blog out there was also meant to simulate how the Historypin website would be marketed later. This is a key point when interacting with expert communities that we cannot emphasize enough: it can be difficult to gain entrée to these groups, and having an insider introduce you can be invaluable.

#### Blog Results:

According to Google Analytics, in a one month trial period the blog received 924 pageviews from 365 visits from 274 unique visitors. The visits were 95% from the United States, but also there were a couple visits from other places such as the UK, Japan, Australia, Germany, and Haiti. Within the United States, San Jose was the city from which the most pageviews came from. Other top cities include San Francisco, Sacramento, and San Mateo.<sup>11</sup>

Overall, there were a total of 25 comments, although these comments were submitted from about 7 different people (some were anonymous). This speaks perhaps to manner in which the blog was sent out-through the personal connections of one man. The comments were generally very informative and people engaged with each other in attempts to answer the questions or ask their own. They actually were able to correct the location of one picture that was listed as Fernberg, but was actually probably Ferndale. Only really one comment or two hinted that railroad enthusiasts are exasperated with the level of difficulty of the questions or their topics. In general, the environmental change seemed to be less of a concern, as one anonymous comment reads: “Hmmm---what are fences usually used for? Does it really matter if the plants are native--as long as they have a green card, what does it matter?” The most comments received on the blog were for the most recent picture, one from Special Collections that was from 1957. Viewers seemed most interested in photos they had not seen before and had good resolution/were conducive to examining small details. This matches with our experience in Year of the Bay, where images that are new to the community are the ones that gain the most traction.

---

<sup>11</sup> San Mateo is a city on the San Francisco Bay Area peninsula. This reflects the geographic location of the community interested in the material we blogged about, as well as the connections made by our local railroad enthusiast partners.

The addition of more photographs that related to a previous photo received little attention in terms of comments or pageviews. Paliza-Carre also added a general call for materials related to a map of Merced, that also received little attention. One issue here was that the blogger format does not alert users when blogs have been updated. Part of the other problem was that returning visitors only comprised 20% of pageviews, and most people that viewed the site (about 80%) were viewing it for the first time and did not necessarily come back (this is an important point for crowdsourcing that we will come back to below).

The work on the testblog, in conjunction with a significant amount of archival research and scanning, led to the Beta development of the Historypin site for Living with the Railroads. Our challenge, as we saw it then, was how engage the wider community. The blog suggested that people respond to people they already know. Our first deployment of the Historypin site at the major Southern Pacific Railroad convention reinforced this.

#### Post launch

As we continued to populate the site subsequent to its launch, we realized that the second major challenge is getting people to return to the site. Components of the site that encourage return visits, such as the ability to comment on photos and receive feedback, or the “Challenges” component of the site, which will offer images that need identification, are meant to overcome this issue. But, like with the Year of the Bay, this component achieved only limited success. Here, the problem is twofold. Although the site is engaging, and our ongoing Stanford railfan days<sup>12</sup> indicate interest in the platform, the railroad enthusiasts (as with all expert communities) already have their own groups, and prefer to engage with their known communities. Railroad enthusiasts were, and are, more than willing to answer specific questions, or look at particular images, but were not as willing, or, rather, not particularly interested in using the Histprypin site as a mode of engagement. Some were

---

<sup>12</sup> Railfan days were events at Stanford where we invited members of the railroad enthusiast community to Stanford to share their work, where we facilitated their access to the Stanford archives in Special Collections, and where we encouraged and helped them to use Historypin. These events are ongoing.

happy to upload material, but comments on uploaded images were sporadic at best.



Here is an example of an image uploaded by the railfan community. It is emblematic in 3 ways. First, it is from the middle of the 20 the century. Second, it is of a train or train car (railroad enthusiasts tend to not be interested in the environment as such). And, third, there are no comments to go with it from the community.

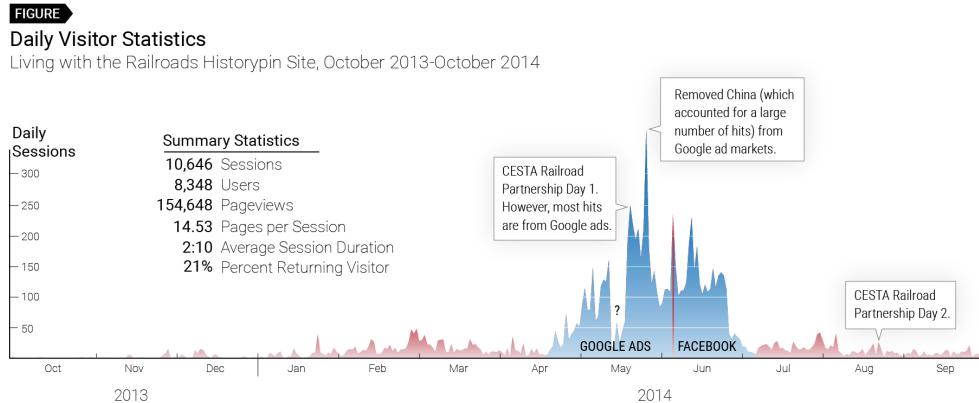
Finally, we found that if we wanted engagement, the most important thing is to present material of interest to railroad enthusiasts. Here, we achieved little success, if success is measured in terms of responses to images. Our researches have suggested that engagement will be more likely to occur if there is something rare or new to the community, or materials that are part of their lived experiences. This might make newer material more popular, though there are enough people with wider historical interests, or interests in family history, that we don't believe this will preclude our long term goal of learning about the impact of the expansion of the railroads in the late 19<sup>th</sup> and early 20<sup>th</sup> centuries.

Like Year of the Bay, Living with the Railroads has required significant, and ongoing, outreach. After the launch of the site, we continued to populate the site through scraping Flickr for train images. We also continued making contact with heritage sites, encouraging them to upload materials to the site. But we also continued to cultivate railroad enthusiasts directly. We began to hold semi-regular (Every 4-6 months) railfan events at Stanford. At these events, we would demonstrate new aspects of the site, help facilitate access to the Stanford Southern Pacific Railroad archives, and create space for railroad enthusiasts to present material or ask questions. This is a high touch, and labor intensive, component. It requires regular email and telephone communication, web interaction, event organization, and its results are not always evident. We believe that through continued outreach to the railroad enthusiast community we will be able to solidify Stanford's place as a center of railroad research, and thereby generate more material that will be potentially useful to the project questions. However, this emerged as a long term goal. If we measure the success of the project as generating data that is immediately useful to the PI, then we cannot see Living with the Railroads as a success. But if we measure in potentiality, there are still good possibilities for success.

### *Conclusions*

In some ways, all the groundwork we laid led to the site being successful. Since the launch of the Historypin website, the content on the site has grown from our initial upload of a few dozen images to more than 3,000 pieces of content, mostly photographs and drawings. Google Analytics for the site began in late April 2013. Between October 2013 when we launched the site and May 2015, more than 8,319 unique visitors have come to the site, spending an average of 2:10 minutes, while viewing around 14 pages per session, for a total of close to 10,603 visits and more than 153,724 pageviews and 101,776 unique pageviews. In addition, we have expanded our community engagement efforts from our small advisory group, to what is becoming a regular meeting of railroad enthusiasts at CESTA. This has resulted not only in more participation in the project, but also the growth of a new project which is geolocating Railroad mile posts. We learned from the community that most railroad materials are located by mileposts, and there was great interest on the part of the community in having precise latitudes and longitudes for stations and posts. This will make it much easier to upload mass upload materials to the Historypin site. In addition, the Railroads Research Assistants developed code and processes for downloading photographs from Flickr and uploading the material to Historypin. This allowed us to populate the site, making it more appealing to the enthusiast community.

As with Year of the Bay, retention of participants proved difficult. Of the total number of sessions (11,497), only 23.3% were from returning visitors. As expected, returning users had a much lower bounce rate (people leaving the site without doing anything) – 1.53% to 3.38%, and a higher average session – 5:12 minutes versus 1:27 minutes for the new users. In spite of the engagement work we did previous to the launching of the site, and the ongoing community building throughout the project, the numbers on our Historypin site were not significantly different from Year of the Bay. Moreover, most of the material on the site came from material tracked down by the CESTA team, or by a few major partners. In this sense, it is difficult to classify the railroad project and site as a success. Progress was made, materials were collected, but engagement with the community proved more difficult than expected (or perhaps just as difficult as expected). The problem, it should be emphasized, resides neither in the academic team nor in the community itself. It resides, rather, between these worlds, where the intermediating institutions and frameworks for collaboration have yet to be built. An academic-community partnership involving railroad enthusiasts will require more than a single crowdsourcing project.



This graph derived from our Analytics shows a snapshot from the period of October 2013 to October 2014, from the soft launch of the site through the periods of greatest activity. The various spikes between January and June correlate with our active use of social media (Facebook) as well as Stanford railfan events. The big spikes correlate to active advertising on Facebook and Google, especially Google. Though these efforts drew lots of traffic, they did not result in significant new material.

These caveats accounted for, we do not look at the project as finished. It represents a step toward the academic-community partnership we seek. This sense of continuation is important for most crowdsourcing endeavors. We feel the groundwork we have laid – through connections to the railroad enthusiast community and the collection of what we believe will be interesting and new materials presented in a digital format – will continue to broaden interest and involvement. In the coming year, we will parlay the work we have done so far into the ongoing data collection the project envisioned. This will include rail enthusiast metadata identification of documents already uploaded to the site, addition and identification of material by enthusiasts, dating of materials, and the addition and geolocation of maps.

There is another component to this crowdsourcing project; the obligation developed with the community of participants. A project like Living with the Railroads requires you to become, at least in some way, part of the community in order to access the knowledge it has. This means that you build up trust with the community, and if you do not want it to feel let down, you cannot just drop it when you have gotten your data. Put another way, the data gathering as part of crowdsourcing may make it seem like you are going to the archive and getting data, but because the archive is people, you owe it civility. One does not want to close the archive for future researchers. Crowdsourcing projects may require researchers to think about them as long term, or, at a minimum, researchers may need to commit themselves to maintaining the online space generated after finishing your project.

In our case we have not yet had to make this choice, as part of the unintended consequences of the project has been the emergence of the possibility of making Stanford a major repository for railroad knowledge about the American West and beyond. We have an incentive to maintain the community and continue to grow our involvement in it. But, one of the key takeaways from both Year of the Bay and Living with the Railroads is the need to be flexible. Crowdsourcing projects may not go the way intended for a variety of reasons:

researchers may not be able to engage with the community they want; the community may not want to give researchers the material they want; the community may not be able to engage with the site; or researchers may find that their initial plan does not gain traction. Researchers set on developing a crowdsourcing project need to be able to follow through with the crowd or community, and in some cases it will be necessary to refine and modify the research question based on what the crowd provides. For us, the jury is still out on Living with the Railroads. For Year of the Bay, it has not been as useful for the initial question the PI started with. Though it is a success in terms of an experiment, it has not been a success in terms of the research. This does not mean that it is not useful. Indeed, there is a large amount of material to be found, as there is with Living with the Railroads. But it is not useful in terms of the research questions we wanted to explore. So, there are no guarantees with crowdsourcing. However, you can improve chances of success – measured in terms of the crowd doing work – if the crowd receives some form of remuneration, as we see next.

## The Emotions of London

Payment to the crowd can significantly increase receiving actionable results, if researchers carefully ascertaining what the crowd is capable of doing. As stated above, the first goal of the 500 Novels project was to refine its research questions to ones more consonant with the expertise of Historypin and with what crowdsourcing research suggests was an ideal combination of characteristics: where prior expertise is unnecessary, crowd engagement is maintained, and the complexity of the task is minimized (Kittur, Chi and Suh 2008, Huberman, Romero and Wu 2009, Heer and Bostock 2010).

Our original research questions for this part of the project related to social networks and their development in nineteenth-century novels. However, in practice, this required the crowd to perform highly complex tasks such as anaphora resolution (for each pronoun, find its antecedent), and the annotation of each instance of dialogue for its speaker and recipient. Anaphora resolution proved prohibitively time-consuming, even for those familiar with Victorian literature, as it introduced complex hermeneutical problems that seemed to transcend the time-based limitations of crowdsourcing projects. For example, discerning the recipient of dialogue requires contextual knowledge sometimes only obtainable by having read the entire novel up to that point. In accordance with best practices in standard crowdsourcing projects and in the context of this particular project, we found that the best tasks were ones that required no prior expertise in the type of literature and no specific knowledge of the particular novel in question.

For these reasons, Principal Investigator Franco Moretti transformed the project into one more limited in scope and geographic in focus. This new approach, led by a research team from Stanford's Literary Lab with support from CESTA researchers, sought to analyze the ways in which physical space and mental constructions of location have worked to organize and represent changing experiences of the city throughout the nineteenth century, a period marked by both the democratization of the novel and rapid urbanization. Specifically, the project focuses on the emotions associated with particular urban locales in London during the nineteenth century, where “the narrative system becomes complicated, unstable:

the city turns into a gigantic roulette table, where helpers and antagonists mix in unpredictable combinations” (Franco Moretti, *Atlas of the European Novel*, 68). By tracking and mapping fluctuations in emotional spaces, Tagging 500 Novels hopes to search for trends among particular places and arrive at possible explanations for these correlations.

Among our research questions are:

- Which regions of the city are most represented in Victorian novels? Are some locales predominantly referred to or imagined, rather than rendered explicitly in the narrative?
- Does the spatial representation of certain kinds of emotional experience correlate to known class divisions in the city? Can we anticipate the rise of what Franco Moretti has called the “Third London,” a middle-class “wedge” in the city, through changing fictional representations of this region?
- Do characters of different genders or social classes have different spatial distributions, and if so, what would this mean? Does the mobility of women in fiction increase over time due to rising women’s activity in the public sphere?

We have found these research questions more tractable to crowdsourcing. In practice, the crowd-based annotations necessary to explore these questions can be derived from passages with no prior knowledge of the novel from which the passage comes. This has allowed us to reduce the amount of literature necessary to pass on to the crowd to only those passages mentioning particular London locales, a reduction of approximately 95% while still leaving more than 15,000 passages for analysis across a wide range of authors, genres, and periods. This more flexible, atomistic nature of the project has also allowed us to evolve the ways in which we ask the crowd for information with the speed and agility necessary for its success.

#### *Summary of activities*

Activity on the 500 Novels project before April 2013 was preliminary: timed tests were conducted of anaphora resolution and other activities demanded by the original vision of the project. Following this, it was determined the project needed to change in scope, so Franco Moretti and the Stanford research team reformulated the project’s goals and discussed with Historypin the ramifications of these decisions. More substantive work began in April 2013 with the hiring of six research assistants to provide additional labor for the new and still vastly complex project.

Our first task in realizing this vision was to gather a list of urban locales within London that play a prominent role in the nineteenth-century novelistic representation of that city. The CESTA research team compiled this list of locations from a variety of sources. First, researchers in the Literary Lab applied the industry-standard algorithm for detecting geographic proper nouns (place names, or toponyms): the Named Entity Recognition software developed by computational linguistics in the Natural Language Processing group at Stanford. Research assistants on the team then manually categorized the most frequently occurring toponyms for their geographic origin, culling from this list all locations within London. Second, this “unsupervised” approach was complemented by a more traditional, “supervised” approach to finding locations in London in nineteenth-century novels: traditional research into studies of urban spaces in Victorian novels. This dual-moded

process yielded 280 unique locations in London, including districts, streets, squares, parks, residences, prisons, bridges, churches, taverns, government buildings, markets, slums, hills, hospitals, inns, clubs, and other types of locales; and which range in origin from Roman to the late nineteenth century.

Beginning in June 2013, our next major task was to develop and refine an algorithm that would both recognize mentions of these locations in novels, and decide whether they are actually referred to as places or whether the mention occurred within an aristocratic title or for some other confounding factor. This required an iterative approach between programming algorithmic rules to decide whether each mention is a legitimate toponym reference; and the human assessment by research assistants of the passages still considered legitimate by the algorithm. The latter, manual side of this iterative approach involved annotating a random sample of passages for each location in order to discern whether each mention was a legitimate place-reference; and if not, why not. From this knowledge, we then developed our algorithmic rules identifying “legitimate” passages in our novels that mention these locations. Our manually created rules ensured, for instance, that mentions do not overlap each other (so that “London Bridge station” is not a mention of both “London Bridge” and “London Bridge station”) and that each mention is a proper noun toponym. For example, if the word “Lord” appears before a place such as “Hounslow”, this particular mention is considered an illegitimate reference to the place Hounslow, and eliminated from further consideration.

We then created a MongoDB database of the remaining legitimate passages. A passage we defined as, at minimum, a 200-word segment surrounding the mention of one of our London locales, but possibly longer so that sentence boundaries are respected and not split by the beginning or end of the passage. Through this data curation process, we were left with 15,256 passages for the 250 canonical 19C British novels we decided to focus on for our “first wave,” given their superior text quality and greater familiarity.

June-September 2013 was spent developing, testing, and refining three main crowdsourcing experiments as Human Intelligence Tasks for the Amazon Mechanical Turk interface: experiments into fictional setting, the status of characters, and the location of emotion. In general, our methodology was an iterative process, moving from:

- a. developing the “template” of an HIT, which requires executing both design principles in making the HIT intuitive to use, as well as research principles in best articulating the instructions of the HIT;
- b. testing the HIT by answering it ourselves as if we were the crowd, using Amazon’s “sandbox” version of its Mechanical Turk interface;
- c. possibly returning to step (a) in order to refine any idiosyncrasies or issues discovered in step (b);
- d. deploying the HIT to the actual crowd, and awaiting their responses;
- e. gathering data from the crowd once finished, and evaluating it for our two measures of data quality.

We will summarize below the results of this iterative process for each our experiments, because we feel it is important for prospective paid crowdsourcers to understand the amount of work that goes into a successful project.

Lastly, however, in parallel to these activities, the Stanford team worked to develop a web interface to display the emotional data received from the crowd via Mechanical Turk, using NodeJS, Leaflet, and MongoDB. This interface is structurally complete.

Even after all this work, the project was only beginning. That is, we had not yet deployed the majority of our passages to the Turkers. We had to actually perform experiments through the Turk interface before we could gather the data for the research. To put it another way, we had to do crowdsourcing experiments to do our crowdsourcing experiment. Because ultimately we had two sets of responses to our HITs -- the collective answers of our own research team, and those of the crowd -- we realized that it would be best to utilize two separate measures of data quality when analyzing the crowd's answers. The first measure is the "consensus" rate, or the percentage of Mechanical Turkers (MTurkers) who constituted either the majority answer or the largest plurality answer. For example, a binary question with 10 respondents would have a minimum consensus rate of 50%, as the most divided opinion would be an even 5-5 split. The second measure of data quality is the "correctness" rate between the MTurkers' answers and the correct answer, as determined collectively by the research team.

Using this methodology, we developed, deployed, tested, and refined three primary experiments, allowing us to gather three independent facets of data about a random sample of the place-based passages (curated in the manner described above). A summary of our initial findings, successes and difficulties with each of these three experiments follows. For each of the experiments, we do not include details about each one of its iterations, but only for the most refined versions of each. Our summary below will focus on setting, characters, and emotions.

### Setting

Our first set of experiments involved the relationship between the mention of a place and its surrounding passage. We wanted to know how the mentioned place manifests in the narrative of the passage. Does the passage represent this place fictionally, through a focused description or through a character's presence there? Or does the passage simply refer to or mention this place, without its being directly represented? Best capturing this distinction, we believed, was the fairly intuitive literary-critical notion of "setting." Separating place-as-setting from place-as-mention allows us to map the literary attention of Victorian novels to London with greater precision, enabling us to separate abstract and concrete kinds of literary attention to these urban locales. This distinction is important as we hope to discover which parts of the city were represented in Victorian novels; for instance, if lower class areas of the city are represented but predominantly in an abstract rather than a concrete manner, this would be a telling indication of the kinds of engagement these novels made into the evolving realities of the metropolis.

To gather this data, we designed an HIT with a simple binary question: is this place (**bolded** at the center of the passage) the setting of this passage? Possible answers included only yes or no. In the final form of the experiment, we had 30 unique MTurkers each answer this question for 10 unique passages. The consensus among the 30 MTurkers for these passages ranged from 53% (16/30) to 97% (29/30), averaging at about 77% with a standard

deviation of 16%. The overall correctness of this experiment was 90%: that is, for 9 out of the 10 passages, the majority opinion of the crowd correctly matched that of our own research team's majority opinion. Both consensus and correctness rates were statistically significantly better than would be expected by random chance. To test this, we designed a "Simulated Turk" algorithm, which simply chose "yes" or "no" randomly 30 times for each of the 10 passages. In the random simulation, the average consensus was 59% with a standard deviation of 4%, while the average correctness was 45% with a standard deviation of 9%. A t-test revealed that both consensus and correctness were significantly improved in the real Turkers versus their simulated equivalents, but in more human terms, consensus was 1.3 times better than chance and correctness twice better than chance. Therefore, this phase of experimentation was considered a success and is ready to deploy on a large-scale to thousands of passages.

### Characters

Our second set of experiments involved characters: who is in these passages surrounding London places? Are they male or female; closer to the lower or to the upper class; and are they actually "in scene," or are they simply mentioned by the narrator or other characters who are in fact in scene? This data would allow us to map the spatial distribution of male vs. female characters; upper vs. lower class characters; etc. The instructions for this HIT were to select all the characters in the scene, and then annotate their gender (Male/Female/Unclear), class (Lower/Upper/Unclear), and whether they were in scene (Yes/No/Unclear). The problems with this HIT began with the character selection. There was an intense lack of disagreement in character selection between MTurkers and even within our research team. The problem that arose with character selection concerned the total number of characters in the passage: there were issues with discerning whether certain pronouns had antecedents, which would take precedence due to the increased specificity. As each person highlighted a different number of characters for each passage, the character selection question did not yield useful data for determining the level of consensus.

Furthering this disappointment, each of our three sub-questions per character -- gender, class, and whether the character was in scene -- achieved consensus rates near the rate of chance. This was particularly surprising with gender. Although our research team achieved a 95% average consensus rate on characters we mutually selected, the crowd's average consensus was much lower, at 62%. The only almost entirely correct selections were characters with gendered titles (Lord, Dame, Sir, Mr, Mrs, etc) prefixed to their name. However, this reliance on shared cultural knowledge is not a secure enough method of assessing gender. The precariousness of this method was made clear when we encountered names whose gender we did not recognize such as Crinoline and Macassar. Whether the character was in scene or not fared even worse, with about an average consensus rate of 40% for both our own research team and the crowd. Worst of all was class, with a consensus rate of 50% for our own team and only 28% for the crowd. Because of the severity of these data, and the dwindling amount of time remaining in the summer term, we decided to prioritize the next phase of experiments while drawing some lessons from ways in which this experiment had been ill-suited to the crowdsourcing enterprise.

## Emotions

Our final phase of experiments was by far our most complicated, with the greatest number of iterations in the HIT that both we and the crowd completed. To describe the evolution of this HIT is not possible in the space allotted here; what follows is simply a summary of our most successful iteration, a simplified iteration of what had preceded it. This phase of experiments represents the most crucial aspect of our research project on the affective geography of nineteenth-century novels. In order to map emotions onto Victorian London, we needed to find a meaningful and crowdsourceable method of associating emotions to places. This we found in a phrasing of the HIT instructions that conflated categories we had previously separated: “Please select whether the emotion ascribed to, or experienced in, this location is positive, neutral, or negative.”

The consensus rates of both our research team (73%) and the crowd (69%) were similar for this HIT, which led us to question the validity of testing for correctness. Because correctness measures the correspondence between the majority opinion of the crowd and our own majority opinion, and because both majority opinions were equally embattled and uncertain, this test of correspondence seemed problematic: who was to say that one of these two equally uncertain groups was more likely to provide the correct answer? We decided to test this hypothesis that the consensus between so-called experts and the crowd was statistically indistinguishable, by passing along the HIT to a group of English Ph.D. students as a more patent group of experts. The result confirmed our hypothesis: their consensus rate was also 69%. Because of this, and because all of these rates were roughly 1.5 times better than chance, we consider this phase of experimentation as successful as can be ascertained statistically, and is ready to deploy on a larger scale.

### *The Larger Experiment*

These experiments allowed us to move on to the large deployment phase, based on the most successful model. For each specific place name where there were at least 10 passages, we deployed those passages to 20 Turkers per passage. Turkers were asked to annotate the place name as to whether the place was, indeed, the setting of the passage. Of the 20 Turkers, ten were asked whether the emotion of happiness was associated with the place, and the other ten whether fear was.

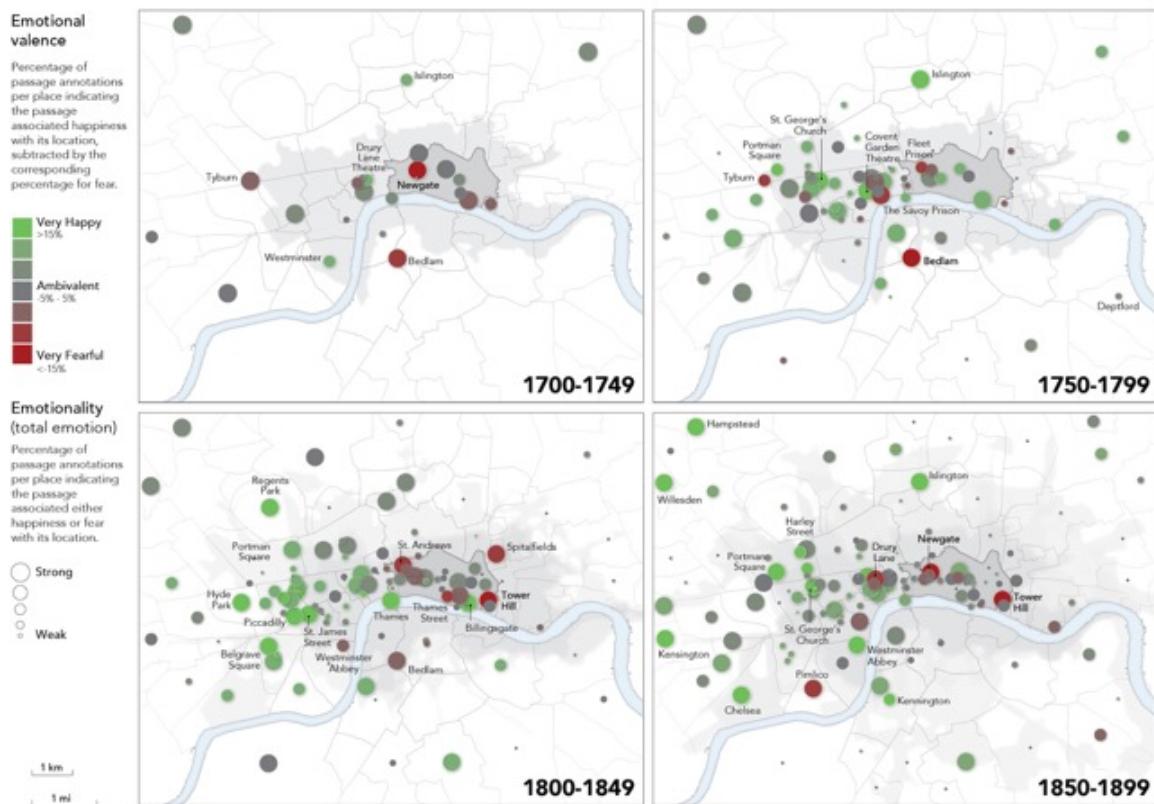
We then gathered the data and analyzed it first to make sure it was legitimately done (that is, that Turkers had not tried to game the system by answering all the questions in the same manner, or in obvious patterns). We did this by looking at agreement among the Turkers who had tagged the same passages. We used a variation of a weighted Cohen’s Kappa score to determine which Turkers agreed with their peers no more often than chance. For those Turkers, we would manually inspect their response, and exclude them, or re-deploy the questions, if necessary.

Using a variation of a weighted Cohen’s Kappa score, we compared the answers provided by each worker on each passage, to the consensus of the nine other workers on the same passage, weighting the importance of each consensus by how many other workers agreed. For example, a situation in which a worker answered that a passage was happy when all nine other readers answered that it was not, was weighted far more than when only five of

the nine other workers answered that it was not. Using this weighted Kappa score across each worker's entire output, we identified workers whose agreement with the consensus was no better than or worse than chance (resulting in a Kappa score approaching 0 or lower). Workers with a score of less than 0.01 across all of either their answers to emotion or setting were flagged for manual inspection. Those with especially high deviation from the consensus were excluded from our analysis.

From the resulting data, we created two maps; one for setting (Figure [2](#)) and one for emotion (Figure [3](#)). For both Figures 2 and 3, we represent places with polygons signifying districts and circles symbolizing buildings. To map setting, we quantified the likelihood that passages mentioning a place were actually set there in terms of the percentage of participants' positive responses to our setting question. The size and darkness of the circles in Figure 2 indicate this likelihood: larger, darker circles indicate that a passage mentioning that place is likely to be set there.. For the emotion data, we quantified the fearfulness and happiness of each place in each period by calculating the percentage of positive answers to both questions on the emotion of the place.

### Emotion in fictional London

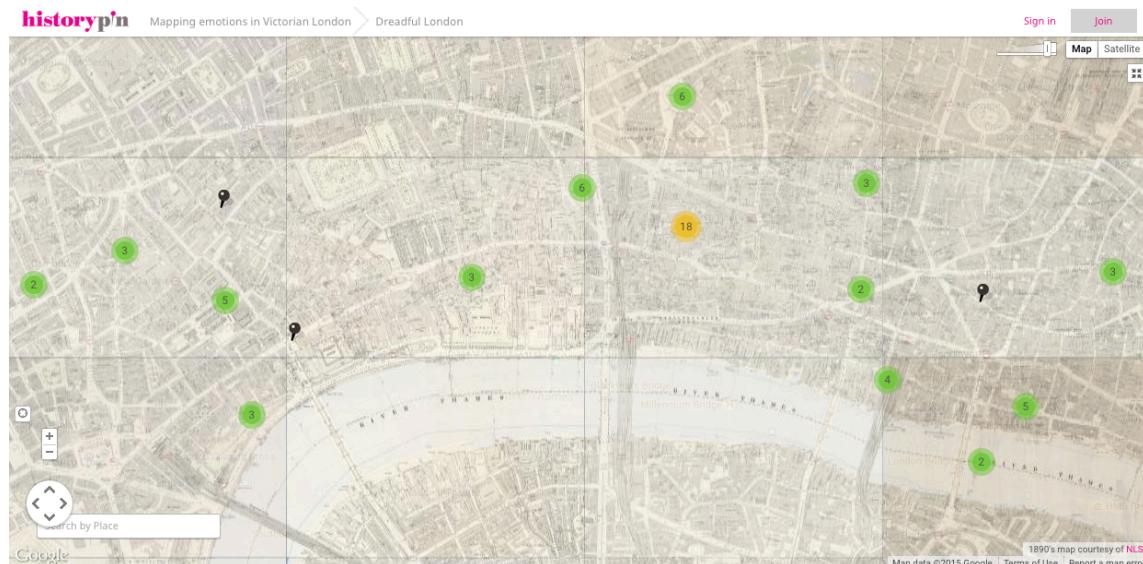


Fearfulness and happiness are therefore represented on a single spectrum, with the percentage of fearful responses subtracted from the percentage of happy ones. This spectrum, ranging from about -50% to +50%, is represented on the map as a red to green color gradient. The color of the circles indicates the emotional quality of the place along this fear-happiness spectrum, with darker reds representing a greater likelihood of fear, and darker greens representing a greater likelihood of happiness. The size of the circles indicates the strength of emotionality. Small circles in the middle of the color spectrum (yellow)

represent “neutral” places rarely found to be emotional; while large yellow circles indicate emotionally ambivalent places that attracted fear and happiness in equal measure.

### *Conclusions*

The varied results of our crowdsourcing experiments have generated several arguments about how crowdsourcing can function within the humanities. First, it seems that a strong central tension in humanities crowdsourcing is between the nuance humanists treasure as the hallmark of the discipline, and the simplicity of design necessary for crowdsourcing projects of this sort to yield data of sufficient quality. This tension played out most strongly in our emotional annotation experiments. Although we eventually settled on a tertiary emotional selection system, allowing the crowd to tag whether the emotion was positive, neutral, or negative, we began with a sophisticated “emotion wheel” developed by psychologist Robert Plutchik with 25 separate emotions. His theory of emotion captured nuances we considered invaluable at the time -- the difference between fear, disgust, and anger, for instance -- but which in practice only depreciated our data quality as input from the crowd became statistically indistinguishable from random selections. Humanists engaged in this kind of research need to learn to compromise on the precision of information they normally desire to find the optimal level of complexity for the tasks given to the crowd, such that they are nuanced enough to retain their humanistic meaning while simple enough to yield statistically meaningful results.



Our Historypin site displaying the data generated through the Mechanical Turk. Here, the points on the map indicate locations in London that were places of fear.

Another lesson for us has been the non-intuitive significance of correctness in humanities crowdsourcing projects. At first, we wanted each member of the crowd to return the “right answer,” and were disappointed when this did not occur. This disposition comes to us we believe from precedents in scientific crowdsourcing projects, where the crowd enters verifiable information such as the position of stars in the sky or the weather as recorded in a particular ship’s log. We discovered that for us, the notion of correctness can only emerge from the plurality of the crowd, so that each piece of information is important

only insofar as it contributes to a plurality, where the most selected answer becomes the opinion of the crowd as a whole. We found that in practice, the plurality opinion of the crowd was a far better indicator of what we considered the correct answer than would be the individual opinion of even a single member of our own research team. Also, we discovered that increasing the size of the crowd beyond a handful of members did not significantly increase the likelihood of consensus or correctness, so that this pluralistic method of gathering data from the crowd is not cost-prohibitive. Lastly, we believe that this method of acquiring humanistic information can possibly turn back around and challenge the ways in which information is gathered and presented in more traditional humanistic projects. We discovered that a same-sized group of Ph.D. students in English literature, and random members of the internet via Mechanical Turk, are just as likely to form a consensus on humanistic questions. From this we derive the lesson that individual contributions from either group are probably equally trustworthy: that is, they are trustworthy only insofar as they contribute to a wider pool of opinions, from which the plurality opinion is taken as representative.

## **Google and Facebook Advertising**

In our various experiments, one of the central questions we needed to answer is how to generate the crowd. For the Emotions of London answering that is simple: you pay people to complete tasks. You may need to do a lot of work to understand your crowd, but you don't need to create it. For the other two projects, however, we either had to create crowds whole cloth, or gain access to crowds who may not know about the projects. We therefore also experimented with digital advertising as a way to either grow or create new community members. In order to test the utility of paid advertising in building an audience for crowdsourcing projects, then, we ran tests using both Google Adwords and Facebook. Our reasoning behind these experiments was twofold: to measure the effect (and cost) of paid advertising and to provide a baseline against which to compare the effects of other forms of outreach, such as mentions in blogs and other social media venues. The results of our tests can be summed up in four findings.

First, advertising is a relatively easy and low cost way to drive the *quantity* of visitors to crowdsourcing websites [e.g., YOTB saw an increase of 36% in average unique pageviews per day in the period of advertising over an equal preceding interval as well as a roughly doubling in the number of users per day; for Living with the Railroads the effects were even larger--the average number of users per week increased by an order of magnitude from 71 to 744].

Second, the *quality* of the traffic thus obtained tends to be lower (lower ratio of unique pageviews [e.g., for YOTB, once advertising commenced, the ratio of unique pageviews to total pageviews dropped from 0.64 to 0.51], less time spent on the site per visit [in LWRR, the number of pages visited per session declined by 60% during the campaign]) than traffic obtained through other forms of outreach. Facebook, which is more targeted to people's interests rather than generic search activities, appears to provide better quality results for crowdsourcing than does Google.

Third, the magnitude of traffic derived from advertising is likely to be greater on average than that generated by strategic placements of projects in social media (blogs turn out to be key brokers of information and audience in this sphere), but less pronounced than specific high-profile placements. For example, during the period before the advertising campaign, a post in Lost San Francisco's webpage regarding Bernal Heights and YOTB caused a spike in users to a project high of 459 on Jan. 21, 2014. During the campaign, the highest daily user count for YOTB was 144. On the other hand, during the campaign, average user numbers per week were roughly twice (422) the average level (212) of project throughout its history prior to the launch of advertising for the YOTB. In sum, advertising works, but the quality of the resulting traffic requires careful cost-benefit analysis given the alternatives. In addition to these core findings, we can also report that paid advertising appears to be more effective in calls to action pitched to crowds who are focused on domains of knowledge rather than local affinities. The Living with Railroads advertisements in Facebook, for example, led to more instances of people taking action (N= 512, cost \$207.37) than did the similar advertisements regarding the YOTB project (N= 202, cost \$135.76).

Project	Number of People	Total Cost	Cost per Action
Living with the Railroads	512	\$207.37	\$.41
Year of the Bay	202	\$135.76	\$.67

The net result of these experiments in advertising shows the best traffic still comes from on-the-ground interaction. Traffic from Adwords did not lead to any appreciable *content generation* for either site. Facebook was more successful in leading to interest in the project – as measured by increase hits and commentary on the Facebook pages for the two projects. But it, too, generated fewer than 10 pins or comments. If you are looking to publicize your project or results, then paid advertising can yield significant dividends. If you are looking to increase participation, however, our results are less positive.

## Conclusions

The result of our experiment *in* crowdsourcing for humanities research can be summed up in three words: it is difficult. In spite of the hard work and considerable expertise committed to the task on the part of our Stanford team and our partners at Historypin, all three test case projects struggled to reach their academic research goals. Where we found success, this was often hard-won through a process of trial and error and a willingness to modify our experimental design. From the perspective of research *on* crowdsourcing for humanities, our study design yielded significant insights. Even so, our conclusions must be taken as provisional, given that our resources only allowed us to conduct three parallel experiments. In scientific jargon: an N of three provides no statistical power. It is also possible that the rate of success for crowdsourced academic research questions is less than one-in-three. Would a fourth (or fifth, or tenth) project have hit the

jackpot? We simply do not know. With these caveats duly noted, the following provisional conclusions can be assayed.

First, academic research in the humanities conventionally plays out over the course of years or decades. The median number of years between scholarly monographs published by members of the Stanford History Department is \_\_\_\_\_. The ethos of crowdsourcing is one of speed and scale. Success is often measured in terms of the quantity of material gathered divided by the time and resources expended. It should come as no surprise that there is a lack of fit between conventional humanities research and crowdsourcing. This is not a fatal problem; it is merely a question of forming realistic expectations and approaching crowdsourcing with the right perspective.

Second, we learned in all three projects that getting quality engagement from the crowd is a challenging problem. Paid advertising can generate large numbers of clicks on a website. It cannot, however, produce good metadata or newly uploaded material that is relevant to the scholarly question posed by academic researchers. Community outreach, blogs, and online mentions by “attention brokers,” far outstrip paid advertising on this metric. Yet these more effective measures are either costly or intermittent. In the end, even with higher quality engagement, there is no guarantee that the crowd will produce truly useful material.

Third, our study shows the strengths and weaknesses inherent in partnerships between academic researchers and third-party crowd/community sourcing providers. By teaming up with a proven and capable partner in Historypin, our project was able to reach far more people through a much smoother interface than would have otherwise been possible. On balance, this partnership was both necessary and worth the cost. Still, there are challenges in such a model. The academic side of the equation wishes to customize and experiment with the platform; the technical and design side wishes to streamline and maintain a consistent look and feel. Not every idea generated by our Stanford team could be implemented. Good communication and trust helped ensure that the basic foundation of the project was sound, but neither the academics nor the community engagement organization could avoid an inevitable degree of miscommunication, delay, and frustration.

Fourth, and finally, our experiments provide the most comprehensive and documented information available for humanities researchers interested in using the power of crowdsourcing to enhance their research. The takeaway message is simple: know your crowd and frame your question accordingly; partner with a strong organization with a proven platform for crowdsourcing; plan for setbacks and assume things will take longer to come together; have an exit plan—a means by which worthwhile results can be harvested in nearly any scenario. For an extended discussion of our recommendations for best practices, see Appendix section C.

## Bibliography

- Brabham, Daren C. *Crowdsourcing* (Cambridge: MIT Press, 2013).
- Christensen, Jon. “Reporting a Negative Result in Crowdsourcing”. *The American Historian*, May 2015
- Cialdin, Robert B. *Influence, the Psychology of Persuasion* (New York: William Morrow, 1993)
- Jon Christensen, “Reporting a Negative Result in Crowdsourcing” *The American Historian* (May 2015)
- Dunne, Stuart and Hedges, Mark. Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research (2012).
- Elson, David K., Dames, Nicholas, McKeown, Kathleen R. “Extracting Social Networks from Literary Fiction,” *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (ACL 2010). Uppsala, Sweden.  
<<http://www.cs.columbia.edu/~delson/pubs/ACL2010-ElsonDamesMcKeown.pdf>>.
- Heer, Jeffrey and Bostock, Michael. “Crowdsourcing graphical perception: using mechanical turk to assess visualization design,” *Proceedings of the 28<sup>th</sup> international conference on Human factors in computing systems* (CHI 2010). ACM, New York, NY, USA, 203-212.
- Heuser, Ryan and Algee-Hewitt, MARK. ‘Mapping the emotions of London in fiction, 1700-1900: a crowdsourcing experiment’, in P. Murrieta-Flores, D. Cooper and C. Donaldson (eds), *Literary Mapping in the Digital Age*. Farnham: Ashgate, 2015.
- Huberman, Bernardo, Romero, Daniel, and Fang, Wu. “Crowdsourcing, attention and productivity,” *Journal of Information Science*, 35 (6) 2009, 758–765.
- Kittur, Aniket, Chi, Ed H., and Suh, Bongwon. “Crowdsourcing user studies with Mechanical Turk,” *Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (CHI 2008). ACM, New York, NY, USA, 453-456.
- Mason, W. A. and Watts, D. J. “Financial incentives and the ‘performance of crowds’”. In KDD-HCOMP, 2009
- Moretti, Franco. *Atlas of the European Novel* (London: Verso, 1998).
- Ridge, Mia, ed. *Crowdsourcing our Cultural Heritage* (Surrey: Ashgate, 2014).

Ries, Al and Trout, Jack. *Positioning. how to be seen and heard in the overcrowded marketplace* (New York: McGraw-Hill, 2001).

Saldanha, Fabio Prado, Cohendet, Patrick, and Possebon, Marlei. “Challenging the Stage-Gate Model in Crowdsourcing: The Case of Fiat Mio in Brazil”. *Technology Innovation Management Review* (September 2014, 28-35).

Schultz, Don E., Tannenbaum, Stanley I., and Lauterborn, Robert F., eds. *Integrated Marketing Communications* (Lincolnwood: NTC Business Books, 1993).

Ward, Amy Sample. Crowdsourcing vs Community-sourcing: What's the difference and the opportunity? <http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>

White, Richard. *Railroaded: The Transcontinentals and the Making of Modern America* (New York: Norton, 2011).

Zarvit, Moshe. “How Crowdworkers Became the Ghosts in the Digital Machine”, *The Nation* (Feb 24, 2014)