**POWELL — Crowdsourcing**

**Crowdsourcing the Text: Contemporary Approaches to Participatory Digital Resource Creation**

Powell, D. J., Van Hyning, V., Wolfe, H., Tonra, J. and Fraistat, N.

**Session Topic**

Crowdsourcing in the digital humanities has reached a tipping point. Digital projects are more likely than ever to integrate participatory components into funding proposals, development plans, and scholarly objectives. A recently funded National Endowment for the Humanities (USA) Digital Humanities Startup Grant is typical: ‘The development of software tools that would facilitate citation and annotation of music notation and capture information about multiple participants’ contributions to collaborative digital projects. As an initial case study, the project would focus on an existing effort to compile a critical edition of Nicolas Du Chemin’s ‘Chansons Nouvelles’.1 Similarly, the Arts & Humanities Research Council (UK) awarded over £4 million in 2014 to various projects under its ‘Digital Transformations in Community Research Co-production’ scheme; many of these are focused on participatory creation of textual materials.

This trend is paralleled by an increased focus on the part of digital humanists. In her DH 2010 keynote ‘Present, Not Voting’, Melissa Terras helped focus digital humanists’ attention on the topic when she noted, ‘Crowdsourcing—the harnessing of online activity to aid in large-scale projects that require human cognition—is becoming of interest to those in the library, museum and cultural heritage industry, as institutions seek ways to publicly engage their online communities, as well as aid in creating useful and usable digital resources’.2 Described in the context of the Transcribe Bentham project, Terras framed crowdsourcing Bentham, and participatory resource creation in general, as a lens through which to discuss a variety of issues in digital humanities writ large, including our dual dependencies on primary sources and technology, legacy data, sustainability, money, serendipity, impact factors, and many more.3 Her talk ends with a call for robust, ‘proactive’ self-reflection to better understand the place of digital humanities in wider academic practices.

Our panel responds to that call. From 2011 to 2014, there have been a number of papers, posters, and workshops on individual crowdsourcing projects at Digital Humanities conferences (including a keynote in Nebraska). Despite this, there has, to our knowledge, never been a singular session or panel devoted to crowdsourcing the creation of textual data. Given the importance of such data to digital humanities research in general, as well as its increasing prevalence in funded projects, a sustained, focus discussion is more necessary than ever. This panel brings together a number of high-profile projects and initiatives to discuss the difficulties and affordances of engaging the public in co-creation of text-based scholarly materials. We will touch on the technical (developing infrastructure systems to facilitate public and collaborative work), the social (how best to work with user communities through effective feedback and communication), and the theoretical (what are the implications for scholarly practice as we now understand it when ‘scholarly’ work is distributed amongst divergent groups?). Our objective is to present a cross-section of current crowdsourcing projects in the humanities that will, collectively, allow us to move forward in conscientiously developing best practices for participatory textual creation in digital humanities.

**Presentations**

**1. Notes Towards a Participatory Humanities**

Powell, D.

Increasingly, academic communities faced with intractable amounts of data are turning to collaborative and participatory methods like crowdsourcing to catalogue, digitise, and effectively leverage their holdings. This includes the vast amount of handwritten and print material held by libraries and archives that has so far been inaccessible to scholarly inquiry on account of overall scale or individual difficulty. Institutions like the New York Public Library, the US National Archives, the British Library, and the University of Oxford are actively seeking to leverage public involvement to catalogue, transcribe, and otherwise process their holdings into usable form. Whether focused on restaurant menus, daily reports from World War I soldiers, or the expansive manuscripts of Jeremy Bentham, harnessing the cognitive surplus of the interested crowd has become a paramount concern for digital humanists and cultural heritage professionals. This presentation will contextualise this panel as a whole by providing a high-level overview of current and past crowdsourcing projects in the humanities that deal mostly or exclusively with the creation of textual content. It will also outline the structure of the panel from a pragmatic perspective.

**1.1 Crowdsourcing the Text: Contemporary Approaches to Participatory Digital Resource Creation**

According to Wikipedia and the *New York Times Magazine*, technology critic and author Jeff Howe coined the term ‘crowdsourcing‘ as a portmanteau of ‘crowd’ and ‘outsourcing.’4 In harnessing the undefined public, individuals and organisations can accomplish tasks with less investment, time, and attention. Although initially very much applied to the private sector, it has increasingly found a place in academia, in cultural heritage institutions, and in a variety of attempts to create scholarly knowledge. Archivists, librarians, cataloguers, humanists, and digital humanists have been especially quick to argue that crowdsourcing methods can be harnessed to enable the digitisation of large amounts of print and manuscript material that present unique challenges to traditional humanist processes. These difficulties often stem from the sheer scale of holdings to be digitised or from the lack of computationally tractable versions of documents (e.g., handwriting that cannot be read via optical character recognition). Crowdsourcing is popularly seen as a way to overcome the first difficulty by harnessing large numbers to undertake small tasks that add up to a greater product. It can overcome the second by the complex decision-making and intuitive decision-making power of large numbers of human brains. In both cases, users become co-creators of academic content that is, after crowd processing, viable for use by academic researchers.5

A number of recent projects have attempted to crowdsource textual information. This overview will draw on specific examples from the following:

• Transcribe Bentham (University College London)

• What’s on the Menu? (New York Public Library)

• Papers of the War Department, 1784–1800 (Center for History and New Media)

• Citizen Archivist (US National Archives)

• Australian Newspapers Online / Trove (National Library of Australia)

• Smithsonian Digital Volunteers Transcription Center (USA)

• Family Search Indexing (FamilySearch.org)

• Wikisource (Wikimedia Foundation)

• Zooniverse (University of Oxford)

• Early Modern Manuscripts Online (Folger Shakespeare Library)

• Crowd Consortium (multinational)

Of particular interest here is how we as digital humanists, as editorial actors tasked with preserving and often remediating cultural materials, deal with issues of oversight, authority, quality, and credit. Projects like those listed above, as well as those like *A Social Edition of the Devonshire Manuscript*, prompt us to consider what happens when crowdsourcing is applied to primary source materials as a way to render them computationally acceptable. What is the role of the humanist in a collaborative world where text overflows the bounds of the academy? How can we ensure scholarly quality, effective public engagement, and cost efficiencies in the new landscape of textual work?

**1.2 Panel Organisation**

This panel has evolved out of a number of discussions under way in the digital humanities, scholarly editing communities, and the GLAM fields, all centred on better understanding how the unique content of the traditional humanities—primarily textual in nature—can best be integrated with the open-source approaches evident in contemporary digital practices. Part of this introductory presentation will provide a succinct overview of the pragmatics of the panel as outlined here.

Panel participants include personnel from Zooniverse (https://www.zooniverse.org/) at the University of Oxford; researchers from King’s College London studying the applicability of collaborative methods to scholarly editing (http://dixit.uni-koeln.de/esr.html#c18485); researchers involved with the development of the Transcribe Bentham (http://www.ucl.ac.uk/transcribe-bentham/) and Ossian Online (http://demo.ossianonline.org/) projects; developers and humanists from the Folger Shakespeare Library (http://www.folger.edu/) involved in building crowd transcription environments; and digital humanists involved with the Crowd Consortium Initiative (http://www.crowdconsortium.org/), funded by the Institute of Museum and Library Services and dedicated to exploring effective crowdsourcing techniques in the GLAM fields. All panelists are engaged in the study of development of various digital humanities tools, projects, and initiatives devoted to understanding crowdsourcing in relation to textual artefacts. The individual presentations fall into three major categories: introduction, case studies, and a roundup of collective insights. Our final presentation—Engaging the Public: Best Practices for Scholarly Crowdsourcing—will move us directly into wider audience discussion during the open-forum section of our session.

The panel will be organised as follows:

• The panel chair (Daniel Powell) will introduce the panel and provide a high-level overview of crowdsourcing cultural heritage materials in the short presentation ‘Notes Towards a Participatory Humanities’ (10 minutes).

• Panelists will each present their short paper (10 minutes each; 40 minutes total).

• The projects, theory, and pragmatics brought up during individual presentations will be discussed in an open forum between panelists and audience members (30 minutes).

• The panel chair will summarise major points from the panel, reflect on ways forward, provide contact information, and close the panel (10 minutes).

The names and affiliations of panel participants are as follows:

• Neil Fraistat, Maryland Institute of Technology in the Humanities, University of Maryland, Crowd Consortium (USA)

• Victoria Van Hyning, Zooniverse, University of Oxford (UK)

• (Chair) Daniel Powell, King’s College London, Electronic Textual Cultures Lab, University of Victoria (UK, Canada)

• Justin Tonra, National University of Ireland Galway, Ossian Online (Ireland)

• Heather Wolfe, Early Modern Manuscripts Online, Folger Shakespeare Library (USA)

We are amenable to a panel chair who is not also a presenter but were unable to find anyone who could commit to attending DH 2015 in Sydney. It was suggested by some of those we approached that an Australian involved with Trove or the Newspaper Digitisation Project would be an ideal moderator.

**1.3 Conclusion**

This individual presentation will serve as an introduction to the overall panel, pressing questions in the field, and an argument for the relevance of examining crowdsourcing text in digital humanities contexts. It will also cover the pragmatics of organisation and the structure of the panel. As such, it is not designed to share original scholarly work, but rather to contextualise the other participants’ presentations so that we might all, together, work towards such insights.

**2. Unlocking Textual Corpora with Zooniverse**

Van Hyning, V.

This presentation will argue that crowdsourced transcription has the power to unlock and render accessible textual corpora that would otherwise remain inaccessible, unedited, and unknown. A number of efforts in this area are taking place under the aegis of Zooniverse (Zooniverse.org), the world-leading academic crowdsourcing organisation based at the University of Oxford and the Adler Planetarium in Chicago. Zooniverse began with a single crowdsourcing project called Galaxy Zoo (galaxyzoo.org/), which was launched in July 2007 by Dr Chris Lintott, now professor of astrophysics at the University of Oxford, and Dr Kevin Schawinski, who was an astronomy graduate student at the time. The goal was to process some one million images of galaxies from the Sloan Digital Sky Survey into two types: spiral and elliptical, a task that would have taken one person a minimum of three years of round-the-clock effort to complete. The several-thousand-strong crowd of volunteers who participated not only completed the task in a matter of months, but each image was classified an average of 38 times, as opposed to once, and thus rendered excellent quality data for each image.

The success of Galaxy Zoo led to the foundation of the Zooniverse in Oxford and the development of over 30 new projects in astrophysics, biology, climate science, and the humanities, including in the fields of music, papyrology, and the history of World War I. Zooniverse now has over 1.18 million registered users and partners with over 400 academic researchers, librarians, and archivists around the world. Following the successful generation of high-quality data by the crowd, Galaxy Zoo cultivated a science team that have published more than 40 papers. The complete list of Zooniverse papers from across all projects is available at zooniverse.org/publications.

Over the past seven years, the developers and researchers involved in Zooniverse have accrued extensive experience in creating user interfaces that facilitate lay participation in the sciences and humanities. Projects have evolved from the basic classification tasks required in Galaxy Zoo to more complex and demanding tasks such as text transcription.

**2.1 Building a Zooniverse for the Humanities**

Zooniverse now supports a number of crowdsourcing projects related to textual transcription, driven by a volunteer base of more than 40,000 registered users. These range from character-by-character transcription of ancient Greek papyri in the Ancient Lives project (ancientlives.org/) to word-by-word transcription in the Shakespeare’s World project, part of the Early Modern Manuscripts Online initiative that will be discussed in greater detail by my fellow panelist Dr Heather Wolfe (Folger Shakespeare Library). The line-by-line transcription approach applied to the sketchbooks and personal papers of artists from Tate Britain (Secret Lives of Artists) will be examined, as well as projects that require tagging and partial text transcription, as in Old Weather (oldweather.org/), concerning ships’ logs; Notes from Nature (notesfromnature.org/), an entomology project; and Operation War Diary (operationwardiary.org/), about WWI British field diaries from the Western Front.

This paper will outline the benefits and challenges of each transcription method deployed at Zooniverse, and reveal what our team has learned from designing, implementing, and sustaining each project. It will suggest how these findings might inform future crowdsourcing initiatives both at Zooniverse and elsewhere.

**2.2 Case Studies**

Zooniverse transcription projects vary widely in scope, content-area focus, and implementation; this talk will consider the following ‘citizen science’ and ‘citizen humanities’ projects and their use of crowdsourced transcription:

**• Ancient Lives**: Character-by-character transcription of approximately 500,000 Ancient Greek papyri fragments; a collaboration between the Imaging Papyri Project, the Oxyrhynchus Papyri Project, the Egypt Exploration Society, Zooniverse, and the University of Oxford.

**• Old Weather**: Transcribing millions of weather entries from ship’s logs; a collaboration between the National Archives and Records Administration (USA), the National Archives (UK), the Met Office (National Meteorological Services, UK), the National Oceanic and Atmospheric Administration (NOAA, USA), the National Maritime Museum (UK), the University of Oxford, and a number of others.

**• Notes from Nature**: Transcription of etymological specimen labels from over 200 institutions, including the University of California at Berkeley, the South Eastern Regional Network of Expertise and Collections (SERNEC) project—a collaborative collection from 222 herbaria across the southeastern United States—the Natural History Museum of London (NHML), Zooniverse, and Vizzuality, to name only a few.

**• Operation War Diary**: A collaboration between the Imperial War Museum (London), the National Archives at Kew (London), and Zooniverse. A tagging and transcription project that is designed to capture data for the Kew catalogues and for a second crowdsourcing project called Lives of the First World War based at the Imperial War Museum.

**• Secret Lives of Artists** (in production): A line-by-line transcription project of artists’ letters, diaries, and sketchbooks—a partnership between Tate Britain and Zooniverse designed to capture full text transcriptions that will eventually be incorporated into the museum’s online archives catalogue. Multiple users will perform each transcription, and these transcriptions will be automatically compared to one another and to gold-standard data to determine consensus.

**• Shakespeare’s World** (in production): A word-by-word transcription project that will enable users to acquire palaeography skills and the ability to transcribe early modern manuscript material in the Folger Shakespeare Library collection. Multiple users will perform each transcription, and these transcriptions will be automatically compared to one another and to gold-standard data to determine consensus.

**2.3 Conclusion**

Each project offers compelling and potentially reusable models of crowdsourced transcription and volunteer engagement that might interest researchers from disciplines and specialties across the digital humanities and the GLAM fields (galleries, libraries, archives, museums). Each of these projects consist of collections of anywhere between 50,000 and 1.5 million images, numbers that make traditional academic consideration—whether in terms of editing, cataloguing, or researching—at best impractical and at worst impossible. Within the Zooniverse, volunteers tackle work that would not be undertaken otherwise, typically because it would take an individual researcher or a small group of researchers decades to accomplish.

Well-structured academic crowdsourcing projects enable researches to ask a consistent question or questions of a large body of data, and thus make both qualitative as well as quantitative analyses in their research. For example, in the first seven months of Operation War Diary, users tagged and transcribed nearly 60,000 pages of material. This is roughly the equivalent of one academic researcher’s full-time effort for five years. The data generated by the crowd has shed new light on troop movement and experiences of life on the front, including armament, recreation, and soldiers’ physical and socio-psychological experiences of the First World War. Studied through the lens of large datasets—especially textual ones that may exist only through collaborative production—even a well-studied topic such as the First World War can be seen in a new light.

**3. Strategies for Effectively Crowdsourcing Transcription: Findings from Early Modern Manuscripts Online**

Wolfe, H.

What happens when you attempt to crowdsource an activity that was previously only the bastion of specialised scholars? That’s exactly the problem faced by the Early Modern Manuscripts Online project (EMMO) at the Folger Shakespeare Library in Washington, DC. Funded by a generous grant from the Institute of Museum and Library Services for three years, EMMO’s goal is to transcribe and encode all of the early modern English manuscripts at the Folger and make them freely available as a searchable corpus of texts. Our hope is that the availability of such texts, previously only available to scholars trained in the study of English secretary hand, will shift primary research from a print-centric view of English history, literature, religion, and politics to a much more nuanced understanding of how texts circulated. We also envision that this corpus will be invaluable for data mining and a wide variety of other research agendas and questions.

**3.1 Strategies for Effectively Crowdsourcing Text**

Currently in its second year of development, EMMO is beginning to experiment with a variety of formats for rapidly training our potential crowds in English secretary hand and basic editorial conventions before letting them loose on digitised images of manuscripts. In this talk, I will report and reflect on our experiences with four interrelated strategies for effectively crowdsourcing textual transcription:

• Online transcription sprints.

• Pedagogical partnerships.

• Intensive palaeography workshops.

• Transcribathons at university libraries.

*3.1.1 Beginners*

EMMO relies on both expert and beginner crowds, and there are different demands for each. Our beginner crowds—initially unknown to us, and most likely unfamiliar with secretary hand and with the orthography and abbreviations of early modern England—are most likely unwilling to sit through an intensive multi-level online tutorial before being allowed to transcribe. Zooniverse has been instrumental in helping us develop an online learning environment that minimises initial investment while making resources such as abbreviation and letterform guides readily available and easily accessible. Because our beginner crowd has never worked with early modern handwriting before, we have a ranking system for our digital images of manuscripts that allows us to direct easier examples to true beginners, and a hidden ‘test’ that allows the most accurate transcribers to play a larger role in vetting transcriptions. Transcriptions of each page image are made four times, at which point the transcriptions are collated so that errors can be highlighted and corrected by sending them out for further transcriptions or for correction by our super-users. Because we imagine that some veteran crowdsourcers will have no interest in transcription at all but are nonetheless interested in the project’s overall goals, we have also created non-transcription tasks that include creating bounding boxes around individual words.

We encounter other beginner crowds at transcribathons. These four- to 12-hour events can take place anywhere, but we have so far held them in university libraries. We typically offer prizes and incentives for our volunteers, as well as mini-training sessions at particular times throughout the transcribathon. Individuals can work on their own laptops or on teams, and we occasionally project ‘stumpers’ on the screen so that everyone can participate simultaneously. Because we are working with people with a variety of skill levels and time commitments, we design the transcribathons to meet multiple needs and interests.

*3.1.2 Experts*

Our expert crowds, largely based at universities, tend to have an interest in the time period and, in some cases, familiarity with secretary hand. We have developed an in-house module called Dromio for transcriptions made by communities at the Folger, in partner classrooms, and amongst an international group of palaeography contacts. Dromio is a TEI transcription/collation tool that pulls images from our digital image database and allows users to transcribe and encode them. The transcription window is a small box that sits on top of a zoomable page; it also includes buttons for our TEI tagset as well as for common early modern abbreviations. The transcriptions and editorial notes can be viewed in HTML or XML, and the names of the transcriber and the encoder are included in the metadata.

So far we have used Dromio for an advanced palaeography seminar that brought together 16 experienced palaeographers for an intensive week of transcribing some of the Folger’s most difficult and perplexing manuscripts, and we will be using it again in summer 2015 for an intensive week of introductory palaeography. EMMO has hired a full-time palaeographer and has a full cadre of interns and volunteers who use Dromio for transcription. We have been approached by professors from history and English departments who want to introduce students to primary resources in the classroom, and who are currently designing classes that incorporate specific Folger manuscripts. Dromio will be released as open-source software at the end of our grant, and we hope that the release will encourage other special collections to contribute to EMMO. Because accuracy is quite important for semi-diplomatic transcriptions, we are particularly interested to compare efficiencies in our partner Zooniverse to our in-house platform Dromio as we embark on the next phase of EMMO.

**3.2 Conclusion**

As indicated above, the Folger is approximately a year into this project. By the time of DH 2015, we will have had the opportunity to hold a number of additional workshops and transcribathons, as well as time to think through an analysis of our different strategies. This will allow us to make recommendations during this panel, based on our successes and failures. In particular, we will be in a unique position to comment on both transcribing difficult-to-parse manuscript content and on an integrated approach to managing and training volunteer communities. It is quite difficult for the ‘man on the street’ to transcribe English secretary hand with no training; combining such training with community management oriented towards producing effective transcriptions of difficult content can be seen as a sort of ‘best case’ litmus test for the efficacy of these models of knowledge co-creation.

**4. Crowdsourcing Texts of Many Dimensions: Transcribe Bentham and Ossian Online**

Tonra, J.

This paper analyses the theoretical and practical implications of crowdsourcing two different kinds of text: transcriptions and annotations. Two projects that adopt the model for these respective purposes are Transcribe Bentham and Ossian Online. They exhibit differing motivations for choosing this model, and aim to crowdsource tasks whose requirements and biases place particular demands and restrictions on participants. As a consequence, the accuracy of the term ‘crowdsource’ must be questioned for more subjective tasks that require the generation of original intellectual content. Collaboration plays a central, though differently focused, role in both forms of crowdsourcing, and investment in infrastructure and in the engagement of a community of volunteers is a constant imperative in this type of endeavour.

**4.1 The Complexity of Crowdsourcing Texts**

*4.1.1 Motivations for Crowdsourcing Text*

The motivations for these two projects’ adoption of the crowdsourcing model give an immediate sense of their typological distinction, and of the particular challenges and opportunities for each. Transcribe Bentham mounted a crowdsourced transcription initiative to produce manuscript transcriptions for the print edition of *The Collected Works of Jeremy Bentham* in a more economical manner than could be achieved by existing editorial staff (Causer et al., 2012, 121). The purpose of Ossian Online is to pilot an interdisciplinary model of scholarly collaboration by facilitating crowdsourced annotation and interpretation of a key 18th-century literary work. Specifically, Ossian Online aims to provide an experience of collaborative and interdisciplinary knowledge creation while testing the value of the crowdsourcing model to more subjective tasks in the workflow of the scholarly edition.

*4.1.2 Bias and the Crowdsourced Task*

The tasks that are required of the crowd in Transcribe Bentham and Ossian Online occupy different positions on the spectrum that runs from microtasking to macrotasking (Walsh et al., 2014, 382f.). Participants in Transcribe Bentham ‘transcribe Bentham’s manuscripts into a text box and, using a customized toolbar, encode the features of the manuscripts in Text Encoding Initiative (TEI)–compliant Extensible Mark-up Language (XML)’ (Causer et al., 2012, 120). The crowd, in this instance, engage in microtasking to solve problems of ‘the sort that would be difficult to solve computationally’ (Walsh et al., 2014, 382): transcription of manuscripts cannot (yet) be automated (though tranScriptorium is engaged in addressing this issue), nor can the encoding of structural features of Bentham’s manuscripts such as additions and deletions. Transcription is a largely objective process, but the addition of markup to the text moves into the realm of the subjective. Despite Transcribe Bentham’s explicit invitation to encode more objectively identifiable structural features of the manuscript, several volunteers considered this assignment more subjective, and ‘an aggravation to an already demanding task’ (Causer and Terras, 2014, 67). At its core, the fundamental and most successful work of Transcribe Bentham was objective: ‘a transcript of Bentham’s writing rather than original intellectual content’ (Causer and Wallace, 2012, 58).

In addition to presenting accurate texts and a critical edition of the Ossian poems published between 1760 and 1773, Ossian Online includes a tool for crowdsourcing annotation and interpretation of the texts. Ossian has previously been the subject of critical attention from a variety of traditional disciplinary and national perspectives, but despite the profound intercultural nature of the work, scholarship on the subject has been confined within disciplinary and national boundaries. In creating a virtual space for sharing intercultural and interdisciplinary perspectives on Ossian, the project aims to generate fresh debate and new knowledge through a synthesis of these disciplinary insights. The aspiration of generating new knowledge based on the specific skills of participants is a characteristic of macrotasking: a crowdsourcing endeavour involving more complex ‘open-ended, socially negotiated tasks [that] may go further than processing information’ (Walsh et al., 2014, 383).

*4.1.3 Demands and Restrictions on Participation*

The particular tasks required of the crowd in Transcribe Bentham and Ossian Online are related to the practices of scholarly editing but demand very different competencies. Transcription of manuscripts, in principle, demands little beyond literacy. Of course, some familiarity with palaeography and of the conventions of preparing transcripts for scholarly editing will hasten the process. But the basic act of recognising and reproducing text is not particularly specialised, demanding, or subjective. The demographics of Transcribe Bentham volunteers and the extent of their work thus far bear this out (Causer and Wallace, 2012, 39–52).

The kinds of contributions invited by Ossian Online—critical interpretation born of close reading informed by disciplinary knowledge—demand a more specialised contributor and inevitably call into question the *crowd* in crowdsourcing. Transcribe Bentham has already identified this fallacy, describing its experience as more accurately one of *crowd-sifting*: ‘beginning with the traditional open call associated with crowdsourcing, and then encouraging the emergence of a self-selecting, smaller number of individuals with the skills, desire and time to complete a complex task on a regular basis’ (Causer and Terras, 2014, 72–73).

Recent examples of crowdsourced interpretation offer different approaches to the task, while sharing a noticeable emphasis on *constraint*. Prism, a tool for collaborative interpretation of texts, allows users to highlight portions of text that correspond to sets of previously defined interpretive categories (*sound* or *sense* in Poe’s ‘The Raven’, for example). Meanwhile, *An Inquiry into the Modes of Existence* (AIME) is an augmented digital book containing sections of glossary, documentation, and interpretive contribution in parallel with the text of Bruno Latour’s book of the same title. Freely accessible on the Web, contributions to AIME are restricted to a hierarchy of contributors, collaborators, and mediators, all of whom have been approved by the project. Constraint in contributors is matched by constraint in contribution: the project has identified and defined 15 topics (or *values*) that will shape the contributions. Their content is also prescribed: a contribution ‘subjects the Inquiry to the trials for which is was designed’, and is explicitly ‘not a comment [. . .] nor is it like editing a wiki or reviewing’.

Transcription and annotation occupy different practical and ideological positions within the process of scholarly editing. Both are part of a tradition that the social edition, and its ‘embrace [of] social networking and commensurate tools’ (Siemens et al., 2012, n.p.), promises to inform and extend. Transcription, though mechanical and value-neutral, is an essential part in establishing and presenting an accurate text. Annotation can span a subjective range from demonstrable empirical observations about the texts (the information of critical apparatus), through contextual or documentary information, to literary-critical commentary. The more subjective and specialised the task, the more necessarily the crowd becomes circumscribed. The creators of Prism are correct in their view that ‘Reading [. . .] does not end with the identification of the words on a page’ (380) and that crowdsourcing can contribute to interpretive encounters with texts. However, the more that interpretation is prescribed, the greater the obligation to qualify the use of the term *crowdsourcing*, or to define precisely the model of collaborative interpretation.

**4.2 Conclusion: Collaboration and the Crowd?**

If the *crowd*, in its multitude and anonymity, is to be edged out of crowdsourcing in cases of macrotasking such as contributing interpretive annotations, it begs the question of why interpretation should be sought from a dispersed group. One of the fundamental reasons is the recognition that certain types of creativity and innovation are enabled by collaboration. The Internet has, since its origins in ARPANET, been a medium that facilitates dispersed academic collaboration and the development of research networks. More specifically, Ossian’s appeal to a broad range of academic disciplines is evident in the research produced by such fields as literature, history, Irish studies, Scottish studies, Celtic studies, romanticism, antiquarianism, textual studies, and book history. Ossian Online entices these existing disciplinary perspectives into dialogue with one another to enable the creation of new knowledge and better understanding and appreciation of an important cultural artefact. Such a practice inevitably unearths consonance and dissonance, and AIME’s treatment of the latter is instructive. The project creates a database of ‘crossings’ that occur ‘when there is a clash between two [interpretive] values’. On this principle, and a related understanding that heightened activity may occur at provocative or polysemic portions of the text, Ossian Online will visualise the scale of user engagement across the span of the text—thus mapping the extent of collaborative annotation and the particular interpretive concerns of the crowd.

**4.3 References**

**Barr, R. A. and Tonra, J.** (eds). Ossian Online. National University of Ireland, Galway.

**Causer, T. and Terras, M.** (2014). ‘Many Hands Make Light Work. Many Hands Together Make Merry Work’: *Transcribe Bentham* and Crowdsourcing Manuscript Collections. In *Crowdsourcing Our Cultural Heritage*. Farnham: Ashgate, pp. 57–88.

**Causer, T., Tonra, J. and Wallace, V.** (2012). Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *The Collected Works of Jeremy Bentham*. *Literary and Linguistic Computing,* **27**(2): 119–37.

**Causer, T., and Wallace, V.** (2012). Building a Volunteer Community: Results and Findings from *Transcribe Bentham*. *Digital Humanities Quarterly,* **6**(2).

**Latour, B.** *An Inquiry into the Modes of Existence*. Fondation Nationale des Sciences Politiques.

**Praxis Program Team.** (n.d.). Prism, Scholars’ Lab.

**Siemens, R., Timney, M., Leitch, C., Koolen, C. and Garnett, A.** (2012). Pertinent Discussions Toward Modeling the Social Edition: Annotated Bibliographies. *Digital Humanities Quarterly,* **6**(1).

**Transcribe Bentham.** (n.d.). The Bentham Project, University College London.

**tranScriptorium.** (n.d.). tranScriptorium Consortium.

**Walsh, B., Maiers, C., Nally, G., Boggs, J. and Praxis Program Team.** (2014). Crowdsourcing Individual Interpretations: Between Microtasking and Macrotasking. *Literary and Linguistic Computing,* **29**(3): 379–86.

**5. Engaging the Public: Best Practices for Scholarly Crowdsourcing**

Fraistat, N.

For the community engaged in leading crowdsourcing projects, as well as developing tools and applications to enable them, an important next step is to assemble our collective knowledge. What do we know at present about the parameters and potential scalability of crowdsourcing infrastructures, content, and tools? Who are the key players, and what are the major projects? What is the evidence of success in these projects? How can we best address the new challenges inherent in relying on ‘crowdsourced’ research resources? How can we establish standards for evaluating and incorporating user-generated contributions to crowdsourced projects? This talk will discuss the findings and results of a crowdsourcing workshop to be held at the University of Maryland in May 2015 in the context of key points made by other presenters in the roundtable.

**5.1 Reports and Findings from Engaging the Public**

Made possible by funding from NEH, IMLS, and the Sloan Foundation, Engaging the Public: Best Practices for Scholarly Crowdsourcing aims to culminate and then broaden the conversation about crowdsourcing begun in two face-to-face regional meetings (20 attendees each) and two webinars that are taking place through the auspices of Dartmouth’s 2014 IMLS-funded National Forum in Crowdsourcing for Libraries and Archives: Creating a Crowdsourcing Consortium (CCLA). Through this two-and-a-half-day capstone event, we will be bringing together 50 scholars from several disciplines and representatives from 10 funding agencies in order to consolidate the earlier work of CCLA and seek to advance a truly cross-disciplinary agenda for the future of scholarly crowdsourcing. We will also be seeking to support crowdsourcing efforts among digital humanities groups, museums, libraries, and archives by linking their work to computer science and social science communities and forging a collective consortium.

Throughout the workshop, our central concern will be on the question of how institutions might best adopt and employ crowdsourcing strategies for increasing public engagement, integrating data from contributors into existing collections, and increasing knowledge in the humanities and related domains. In obtaining support for the workshop from three different funders, each with their own distinct communities to bring into the conversation, we hope to ensure a rich cross-disciplinary dialogue and send a very public signal about the importance of these emerging practices, thereby increasing the overall impact of the workshop.

At its most fundamental level, the workshop has as its central question, ‘How might institutions best adopt and employ crowdsourcing strategies for increasing public engagement, integrating data into existing collections, and increasing knowledge in the humanities and other disciplines?’ Under this broad umbrella are a number of pressing questions that will guide the workshop, including,

• What is a crowdsourcing platform, and why is it important?

• What advantages does crowdsourcing offer, and under what conditions does it fulfil those promises?

• How might institutions best adopt and employ crowdsourcing strategies for use in collecting metadata, transcribing historic material, integrating data into existing collections, and increasing user engagement?

• Which types of institutions would benefit most from the use of crowdsourcing (e.g., libraries or archives with recently digitised collections, ones in danger of losing funding or closing due to lack of community involvement or patronage, etc.)?

• Can crowdsourcing strategies be used to increase public engagement in the physical spaces of cultural heritage institutions?

• How can crowdsourcing activities best be used in the classroom?

• How will material and crowdsourced data be collected, stored, and made accessible?

• How can interfaces such as games be effectively designed not only as data collection and processing tools but also as powerful means to motivate and mobilize the public?

• How can funders best support scholarly crowdsourcing?

In pursuing such questions, the workshop has the following aims:

1. Expand the conversation about best practices in engaging the public across the humanities and with the sciences.

2. Begin to establish standards for evaluating and incorporating user-generated contributions in research and in national collections. Determine what metadata and provenance are needed to preserve and contextualize public contributions.

3. Establish guiding principles that can inform the implementation of crowdsourcing efforts.

4. Collect ‘lessons learned’ from prior experiences with crowdsourcing and from the analyses of such work in the social sciences, including cutting-edge information and best practices in the field. Summarize lessons learned in an accessible way for the humanities, with accompanying principles, strategies, and best practices gathered into a useful visual brief.

5. Help establish a national consortium among these groups to examine collectively how tools and platforms to date can change the ways in which museums, libraries, archives, and research projects based elsewhere can use crowdsourcing technology to enhance collections and user experiences.

6. Provide a means for program officers from major funders to deepen their understanding of the current state of the art as well as the opportunities and challenges of crowdsourcing and to engage with the crowdsourcing community.

7. Produce a livestream feed and podcasts of keynote and panels, as well as a detailed written report documenting the proceedings and findings of the workshop.

8. Move findings and podcasts to CrowdConsortium.org, a group web presence, that will serve to house these best practices documents, reports, links to top research, and open-source, sharable resources for stakeholders interested in crowdsourcing.

**5.2 Conclusion**

This presentation will report on the key outcomes achieved during the workshop as a point of departure for larger group discussion of both the current state of the art (partly represented by the other presenters in the roundtable) and the future of scholarly crowdsourcing.

**Notes**

1. See NEH Grant HD51836, Enhancing Music Notation Addressability, awarded to the University of Maryland in March 2014, http://www.neh.gov/divisions/odh/grant-news/announcing-20-digital-humanities-start-grant-awards-march-2014.

2. See the abstract for her talk in the compiled abstracts from DH 2010, http://dh2010.cch.kcl.ac.uk/academic-programme/abstracts/papers/pdf/book-final.pdf.

3. See Melissa Terras’ blog for a full transcript, as prepared, of her keynote, http://melissaterras.blogspot.co.uk/2010/07/dh2010-plenary-present-not-voting.html.

4. See ‘Crowdsourcing’ on Wikipedia (http://en.wikipedia.org/wiki/Crowdsourcing) and an ‘On Language‘ column by William Safire from 2009 (http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html?\_r=3&ref=magazine&).

5. Mia Ridge, writing on her Open Objects blog, has a highly useful FAQ page that addresses more thoroughly the relationship of crowdsourcing to academic work; see http://openobjects.blogspot.co.uk/2012/06/frequently-asked-questions-about.html.