Chapter 3

'Many Hands Make Light Work. Many Hands Together Make Merry Work'[1]: *Transcribe Bentham* and Crowdsourcing Manuscript Collections

Tim Causer and Melissa Terras

The philosopher and reformer, Jeremy Bentham (1748–1832), was a firm supporter of innovation and inquiry. Amongst other things, Bentham proposed a scheme for preventing the forgery of bank notes,[2] and in the designs of his proposed 'panopticon' prison, provided a detailed description of how the building would be heated, as well as a network of 'conversation tubes' which would allow the prison inspector to communicate instantly with individual prisoners in their cells.[3] Bentham's home, at Queen's Square Place in Westminster, was itself centrally heated.[4]

Bentham believed that modern, scientific enquiry was the most accurate means by which to investigate and solve social ills. His principles and methods were adopted by social reformers of the 1820s and 1830s, who achieved the amelioration of the criminal code, the ending of convict transportation to New South Wales, the widening of the electoral franchise, and the crowning glory of the 1830s: the abolition of slavery across the British Empire in 1833.[5] None of these reforms would have been possible, Bentham would have argued, without the widespread availability of knowledge and evidence.

Following in his example, the *Bentham Papers Transcription Initiative* (*Transcribe Bentham*) has utilised modern technology to digitise the vast

---

1   This quotation is from a Bentham manuscript dated 21 December 1793, discovered by volunteer transcriber Peter Hollis. The fuller quotation reads: 'Many hands make light work[.] Many hands together make *merry work*. Each to take the work of all the rest and critizise [*sic*] it' (emphasis in original). See http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/107/020/001, revision dated 15.37, March 18, 2013.

2   Pitkin, 'Slippery Bentham', 105.

3   Jeremy Bentham, 'Panopticon Postscripts: Section XXIII' and 'Section VIII', in *Bowring*, vol. 4, 110–18, 84–6.

4   O'Sullivan and Fuller, *The Correspondence of Jeremy Bentham: Volume 12* (*Collected Works of Jeremy Bentham*, hereafter '*CW*'), 280–2n.

5   For example, see Quinn, *Writings on the Poor Laws* (*CW*), vols 1 and 2.

collection of manuscripts written and composed by Bentham – held by University College London (UCL) Library's Special Collections[6] – and to make them available for scholars, students and the public at large to access and transcribe via a specially designed web platform. A collection of great historical and philosophical importance, previously only accessible on a research trip to London, is now progressively being made available to anyone in the world with an internet connection, and in a way which allows interested individuals to engage and contribute to our growing knowledge about this fascinating historical figure.

Transcribe Bentham: Why?

Bentham is perhaps best known for two things. First, for the aforementioned panopticon prison, based upon an idea conceived by Samuel Bentham, Jeremy's younger brother. In the panopticon, the prisoners' cells were to be arranged in a circle around a central inspection tower, exposing the inmates to what they had to assume was constant surveillance by an unseen inspector, and thereby causing them to modify their behaviour to avoid punishment. By this 'simple idea in Architecture', as Bentham put it, the deviancy of criminals could be cured, and the 'central inspection principle' would be equally applicable to poor houses, factories, insane asylums and schools.[7]

Second, Bentham willed that his remains – in the hope that others would be encouraged by his example to donate their own bodies to medical science – were to be publicly dissected, and then 'put together in such a manner as that the whole figure may be seated in a chair usually occupied by me when living in the attitude in which I am sitting when engaged in thought'.[8] Bentham left his corpse to his friend, Dr Thomas Southwood Smith, who dissected the body, and then reassembled and dressed the skeleton. For the next 18 years, Bentham's auto-icon ('self-image') sat in Smith's house, until in 1850 it was brought to UCL.[9]

However, the panopticon and the auto-icon tend to obscure Bentham's enduring importance in a wide range of fields. Bentham is one of the world's great thinkers, whose thoughts and ideas have had a profound historical impact and are still of contemporary significance. He was the founder of the modern

---

6   UCL Special Collections, http://www.ucl.ac.uk/library/special-coll/ (last accessed February 21, 2013).

7   Bentham, 'Panopticon, or, The Inspection-House', in *Bowring*, vol. 4, 39. See pp. 39–66 for the panopticon letters, and pp. 67–172 for the lengthy, detailed 'postscripts'. The key study of the panopticon is Semple, *Bentham's Prison*; the panopticon was never built, and its failure was the great regret of Bentham's life.

8   Bentham, *Auto-Icon and Last Will and Testament*. See also Marmoy, 'The "Auto-Icon" of Jeremy Bentham'.

9   See http://www.ucl.ac.uk/Bentham-Project/who/autoicon and http://www.ucl.ac.uk/Bentham-Project/who/autoicon/Virtual_Auto_Icon (both last accessed February 15, 2013).

doctrine of utilitarianism: that the right and proper end of all action and legislation is to promote the greatest happiness. Bentham laid out a systematic theory of punishment which emphasised deterrence, proportionality of punishment and reformation of prisoners, his *Nonsense upon Stilts* is an influential critique of the doctrine of natural rights (the forerunner of human rights theory)[10] and he was an important theorist of representative democracy. Bentham wrote on topics as varied as political economy, religion, jury reform and sexual morality (and this is only a summary).

Researchers and students wishing to access Bentham's thought, however, face a substantial obstacle: the edition of Bentham's works published between 1838 and 1843 by his literary executor, John Bowring is sorely inadequate for the needs of modern scholarship. The 'Bowring edition' is incomplete, as it omits several works published in Bentham's lifetime (particularly those concerning the sensitive topics of religion and sexual morality)[11] and substantial unpublished works which survive in manuscript. The edition also includes edited translations into English of 'simplified' French versions of some of Bentham's works, produced by another of his disciples, Etienne Dumont, so there is a question concerning the extent to which these texts are authentically Bentham's, as opposed to Dumont's and the translator's.[12] Finally, the edition's densely typeset text makes it a chore to use, and its biography of Bentham has been described as 'one of the worst biographies in the [English] language, out of materials which might have served for a masterpiece'.[13]

There was, then, until relatively recently, no adequate edition of Bentham's works which accurately represented his writings as he envisaged them. An attempt to rectify this deficiency began in 1959 with the foundation of the Bentham Project at UCL, which is engaged in producing the new, critical edition of the *Collected Works of Jeremy Bentham*, based on both Bentham's published works and his unpublished manuscripts, and returning to what Bentham himself actually wrote. It seems an almost Sisyphean task: UCL's Bentham collection runs to some 60,000 manuscript folios (estimated to contain c. 30,000,000 words), while the British Library holds a further 12,500 folios (c. 6,250,000 million words). Thirty of an estimated 70 volumes of the new edition have been published, and a total of around 28,000 folios have been transcribed. The majority of the Bentham Papers therefore remain untranscribed and their contents largely unknown, save for an outline index,[14] and the greater part of

---

10   Schofield et al., *Rights, Representation and Reform (CW)*.

11   Bentham's writings on sexual morality were published in Schofield et al., *Of Sexual Irregularities* (*CW*), in 2014. For a summary, see Schofield, *Jeremy Bentham: Prophet of Secularism*. The Bentham Project has published online a preliminary text of the third, unpublished volume of Bentham, *Not Paul, but Jesus*.

12   Schofield, *Bentham: A Guide for the Perplexed*, 19–43.

13   Stephen, *The English Utilitarians*, 225.

14   Milne, *Catalogue of the Manuscripts of Jeremy Bentham*.

the *Collected Works* has yet to be published. As a result, we only have a partial understanding of the true extent of Bentham's thought, as well as its historical and contemporary significance.

The purpose of *Transcribe Bentham* is threefold. First, it produces transcripts of Bentham manuscripts of sufficient quality for uploading to UCL's free-to-access digital repository for access, searching and to ensure the collection's long-term digital preservation and curation.[15] Second, it allows volunteers from around the world to contribute to humanities research: their transcripts will act as a starting point for editors of future volumes of the *Collected Works*, and volunteers will be fully credited in the volumes to which they contribute.[16] Furthermore, as many manuscripts have not been read since Bentham wrote them, there is also the potential for exciting new discoveries to be made which could change our perception of Bentham's thought. For example, the work of volunteers has shown that a substantial unpublished portion of 'Panopticon versus New South Wales',[17] Bentham's attack on convict transportation, exists in manuscript.[18]

Thirdly and finally, *Transcribe Bentham* was formulated as an experiment. The task required of volunteers is perhaps more complex and challenging than in many other crowdsourcing projects, and demands a high degree of concentration and engagement with a source material which is not, in many instances, the most immediately accessible or attractive. Would volunteers – who may not have had any palaeographical training, or have previously encountered historical manuscripts – manage to read and decipher Bentham's handwriting? Would they be able to identify the structural and compositional features of the manuscripts and mark these up in Text Encoding Initiative (TEI)-compliant Extensible Mark-up Language (XML), while also navigating Bentham's idiosyncratic style, along with his often challenging ideas? In addition, would the work of volunteers be of sufficient quality to act as a basis for editorial work, and for uploading to a digital repository for public access? And would *Transcribe Bentham* prove to be worthwhile both in terms of cost and time? After almost three years' experience, we are delighted to say that the answer to all of these questions is, to varying degrees, 'yes', as we will subsequently discuss.

---

15 http://www.ucl.ac.uk/library (last accessed February 5, 2013).

16 See Causer et al., 'Transcription Maximized'.

17 Bentham, 'Panopticon versus New South Wales', in *Bowring*, vol. 4, 173–248. Bentham wrote this work in 1802, and it was privately printed in 1803. It was not published for public consumption until 1812, and was reproduced in the 1838–43 edition of Bentham's works.

18 For volunteers' discoveries, see Causer and Terras, 'Crowdsourcing Bentham'.

Transcribe Bentham: What?

Transcribe Bentham is coordinated by UCL's Bentham Project,[19] in partnership with UCL Centre for Digital Humanities,[20] UCL Library Services,[21] UCL Creative Media Services[22] and the University of London Computer Centre (ULCC).[23] In October 2012, the British Library[24] joined the project consortium.

    Transcribe Bentham has, thus far, had three 'phases'.

Period 1: Design and Establishment (March 2010 to 8 March 2011)

Transcribe Bentham has its foundations in metadata compiled between 2003 and 2006 for the *Bentham Papers Database Catalogue*.[25] The *Catalogue* records 15 fields of information, including dates, headings and titles, for each of the 60,000 folios in the UCL Bentham Papers collection. It was initially conceived of as a resource for Bentham Project editorial staff and researchers consulting the UCL Bentham Papers, though it was hoped that one day it could be improved by adding transcripts and digital images.

    The *Transcribe Bentham* consortium was successful in securing a £262,673 grant from the Arts and Humanities Research Council's (AHRC) Digital Equipment and Database Enhancement for Impact (DEDEFI) scheme, a one-off call to fund projects for 12 months.[26] This was invested primarily in digitising around 12,000 folios, on the production of a collaborative transcription platform developed by the ULCC, and on the salaries of two full-time Research Associates to coordinate the programme.[27]

    At the heart of the project is the 'Transcription Desk', a customised installation of the MediaWiki software application, which incorporates the transcription platform and other elements important to the project (see Figure 3.1). The use of MediaWiki is a key factor in *Transcribe Bentham*'s success: it is perhaps the world's single most widely used collaborative open-source software for authoring online content, is stable, well-documented and has a global user base. Moreover, it is an interface which is instantly familiar to the millions of people who use

---

    19   http://www.ucl.ac.uk/bentham-project (last accessed February 5, 2013).

    20   http://www.ucl.ac.uk/dh (last accessed February 5, 2013).

    21   http://www.ucl.ac.uk/library (last accessed February 5, 2013).

    22   http://www.ucl.ac.uk/isd/common/creative_services (last accessed February 5, 2013).

    23   http://www.ulcc.ac.uk (last accessed February 5, 2013).

    24   http://www.bl.uk (last accessed February 18, 2013).

    25   http://www.benthampapers.ucl.ac.uk (last accessed February 18, 2013). The database catalogue was compiled by Dr Deborah Colville and funded by the AHRC.

    26   http://www.ahrc.ac.uk/FundingOpportunities/Pages/dedefi.aspx (last accessed February 5, 2013).

    27   See Causer et al., 'Transcription Maximized', 121–2 for allocation of the AHRC grant.
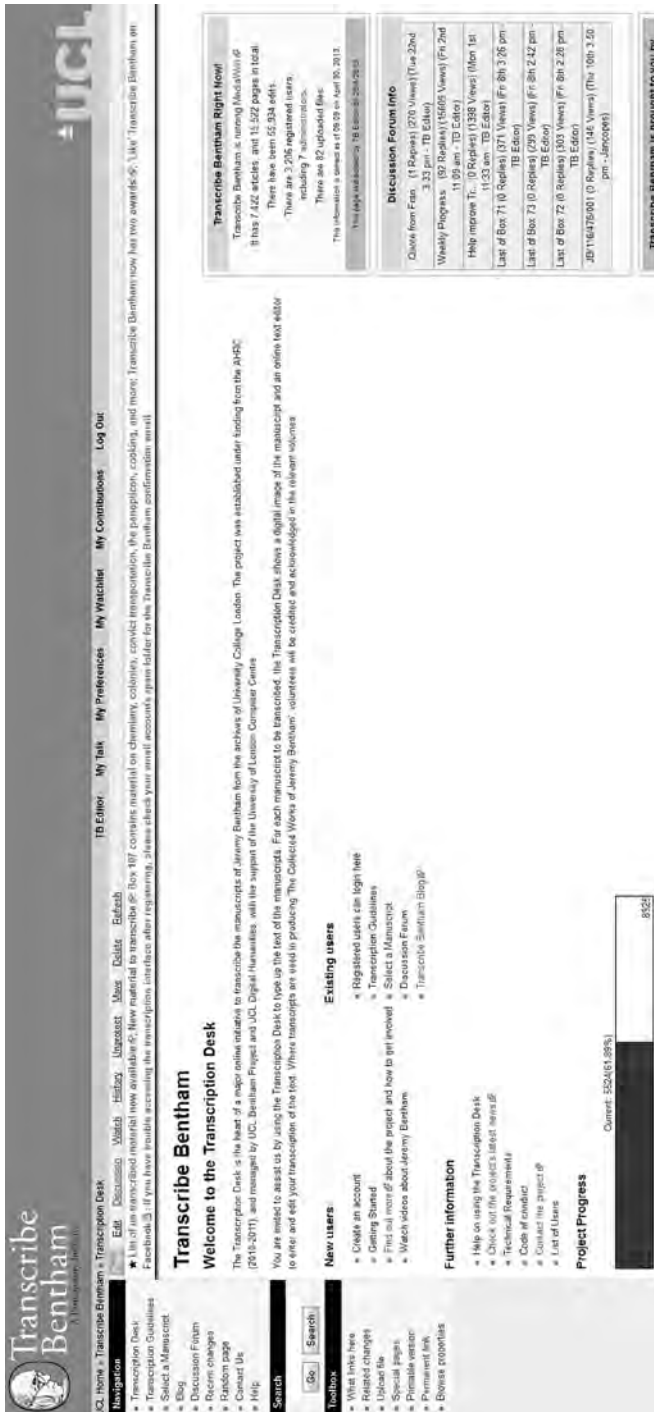
Figure 3.1   The *Transcribe Bentham* 'Transcription Desk' platform

Wikipedia each day, is customisable, easily maintained and offers a full revision history for every individual page in case of malicious or accidental edits, or spam.

As well as transcribing the text, volunteers also encode key features of Bentham's manuscripts in TEI XML. Use of TEI has become best practice for systematically encoding texts, whether prose, poetry, drama, primary source material and more besides. All the elements of the text can be encoded and identified with varying degrees of granularity, ranging from entire paragraphs down to lines, clauses and words, thereby allowing computers to read, understand and represent both the content and the appearance of the text, and to facilitate complex searching and querying of the transcribed corpus. *Transcribe Bentham* volunteers can, for example, indicate Bentham's deletions and interlineal additions through TEI mark-up, and these are represented in the rendered version of the transcript, providing an accurate digital representation of the original manuscript. TEI mark-up also has the added advantage of allowing the transcripts to be easily converted into any number of formats, and ensures their long-term preservation. Plain text transcripts of the material would be pointless: searching would be crude and imprecise, the transcripts would look nothing like the manuscript from which they were derived and conversion to other file formats would be labour- and time-intensive.

It was recognised that transcribers may not have any experience of mark-up, let alone TEI, and so a method by which volunteers could easily encode their transcripts was devised: the Transcription Toolbar.[28] Two MediaWiki extensions were developed by ULCC to facilitate the addition of mark-up by transcribers: *JBZV*, which adds an image frame next to an editing form, so that the manuscript could be transcribed into a text box and then saved; and *JBTEIToolbar*, allowing TEI mark-up to be automatically applied at the click of a button, and which renders the encoded transcript in the Wiki.[29] Using the toolbar, volunteers can straightforwardly indicate structural features of the manuscripts such as line-breaks, page-breaks, paragraphs and headings, and compositional features such as underlinings, additions, deletions and marginal notes, without necessarily having to learn the minutiae of mark-up (see Figure 3.2).[30]

In practice, a volunteer is presented with a zoomable image of a manuscript, a plain-text data entry box into which they enter their transcript and the transcription toolbar. When satisfied with their transcript, the volunteer submits it for assessment by a *Transcribe Bentham* project editor, who checks for textual accuracy and consistency of encoding. Changes are made to the text and mark-up, if necessary, the key question being whether appreciable improvements are likely to be made through further crowdsourcing, and if the transcript is of the requisite quality for public viewing and searching, and as a basis for editorial

---

28   For detailed discussion of the use of TEI in *Transcribe Bentham*, see Causer et al., 'Transcription Maximized', 121–5.

29   See http://www.mediawiki.org/wiki/Extension:JBZV and http://www.mediawiki.org/wiki/Extension:JBTEIToolbar.

30   Causer et al., 'Transcription Maximized', 122–3.

Figure 3.2 The *Transcribe Bentham* transcription interface, and transcription toolbar

work. If approved – if there are few or no unclear words or gaps in the text – the transcript is locked. If there are a number of gaps in the text, or the text is only partially transcribed, then the transcript remains available for editing. In either circumstance, an acknowledgement message is left on the submitter's user page. Though an unavoidably impressionistic and subjective judgement, the quality-control process does ensure that locked transcripts are a reliable guide to the contents of the manuscripts, and encourages volunteers by providing feedback and an acknowledgement of their work.

Transcribe Bentham was launched to the public on 8 September 2010 for a six month testing period, which ran until 8 March 2011.[31] During the first three months, the rate of transcription was steady but unspectacular: by 23 December 2010, 350 users were registered with the project, and 439 manuscripts had been transcribed or partially transcribed (see Figure 3.3). Only one volunteer regularly participated, while others took part on a sporadic basis. At this stage, the project did not seem all that successful, but a December 2010 *New York Times* article about *Transcribe Bentham* and crowdsourcing in the humanities had a transformational effect.[32] By way of illustration, from 8 September to 23 December 2010, an average of 25 manuscripts (c. 12,500 words) were transcribed or partially transcribed each week, whereas from 24 December 2010 to 8 March 2011, this increased to an average of 57 manuscripts (c. 28,500 words) per week. In short, the *New York Times* article and associated media coverage gave *Transcribe Bentham* momentum which has remained with the project ever since.

By the end of the testing period, *Transcribe Bentham* was in good shape. In total 1,222 volunteers had registered an account, and 1,009 manuscripts (c. 504,500 words) had been transcribed or partially transcribed, of which 559 (55 per cent) were complete.

Period 2: Consolidation (9 March 2011 to 30 September 2012)

Though the AHRC grant continued until 30 April 2011 to allow for reporting, full-time staffing of the Transcription Desk ceased on 8 March 2011, and this was communicated to volunteers.[33] We anticipated that this would result in a much reduced rate of transcription, and our fears appeared to be confirmed when all but three of the then seven regular transcribers ceased participating.

However, these concerns were ultimately misplaced, as the 18 months after 9 March 2011 proved to be a highly successful period for *Transcribe Bentham*, despite the project running only on small-scale funding provided by UCL, covering web storage costs and two days per week of staff time (there was no money for further

---

31 For *Transcribe Bentham*'s testing period, see Causer and Wallace, 'Building a Volunteer Community'.

32 Patricia Cohen, 'Scholars Recruit Public for Project'.

33 http://blogs.ucl.ac.uk/transcribe-bentham/2011/03/08/six-months-later/ (last accessed February 20, 2013).
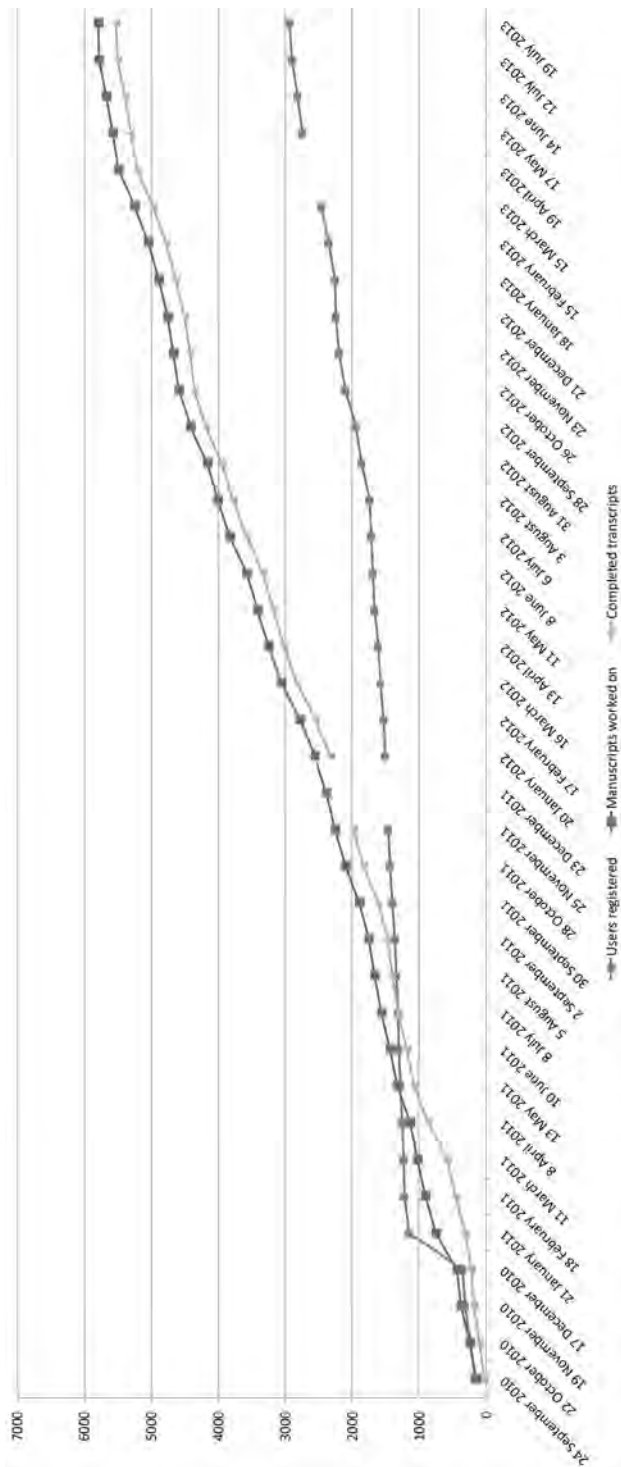
Figure 3.3 *Transcribe Bentham results, 8 September 2010 to 19 July 2013*

Note: Gaps in the data are owing to staff being away during UCL vacations.

digitisation or for modifications to the transcription interface). By 30 September 2012, 1,939 users had registered an account with *Transcribe Bentham*, and 4,412 manuscripts (c. 2,200,000 words) had been transcribed or partially transcribed, of which 4,185 (94 per cent) were complete.[34] An average of 42 manuscripts were worked on each week during this period, and the transcription rate was particularly high from mid-September 2011 to mid-March 2012, most likely owing to a *Sunday Times* article on scholarly crowdsourcing of 11 September which mentioned *Transcribe Bentham*.[35] Also that month, *Transcribe Bentham* received a major international prize: an Award of Distinction in the Digital Communities category of the 2011 Prix Ars Electronica.[36]

Despite *Transcribe Bentham*'s successes and the continuing engagement of a core group of volunteers, without further investment there was a danger that the project might stagnate if issues raised by volunteers in a survey of early 2011 were not addressed. Some of the survey's most important findings were in understanding volunteer motivations, and what dissuaded participants from transcribing more (or at all). Survey respondents reported that they took part mainly owing to interests in: Bentham's life and thought; history and philosophy; crowdsourcing and the technology behind the project; and a sense of altruism, taking part in something which will ultimately benefit the wider community. On the other hand, respondents told us that the main factors which limited their participation were: a lack of time in which to learn how to transcribe Bentham's handwriting; various issues with the Transcription Desk; the difficulty of deciphering Bentham's hand; and the TEI mark-up was considered by several volunteers as an aggravation to an already demanding task. A failure to address these very real concerns ran the risk of alienating regular *Transcribe Bentham* participants, and of limiting the recruitment of a wider pool of volunteers.[37]

Period 3: Expansion (1 October 2012 – onwards)

Fortunately, the Bentham Project and the *Transcribe Bentham* team were successful in securing further funding. For two years from 1 October 2012, the initiative is supported by a grant of £336,157 ($538,000) from the Andrew W. Mellon Foundation's 'Scholarly Communications' programme, for a wider scheme entitled the *Consolidated Bentham Papers Repository*, with the British Library joining the project consortium. This funding will, we believe, allow *Transcribe*

---

34   The large increase in completed transcripts was owing to project staff working through all partially transcribed manuscripts, and increased proficiency of *Transcribe Bentham*'s regular participants.

35   Kinchen, 'One Stir, Then I'll Discover a Galaxy'.

36   See http://archive.aec.at/#42434 for *Transcribe Bentham*'s citation, and http://www.aec.at/prix/en/ for more about the Prix Ars Electronica (both last accessed February 20, 2013). In 2009, this award was given to *Wikileaks*.

37   Causer and Wallace, 'Building a Volunteer Community'.

Bentham to achieve its full potential. Most of the remaining UCL Bentham Papers will be digitised, along with all of the 12,500 folios of Bentham manuscripts held by the British Library. Metadata will be compiled for the latter collection, and the 20,000 or so transcripts produced by the Bentham Project in Microsoft Word during the past 25 years will be converted to TEI XML. All of the manuscripts will be made available for crowdsourced transcription, and the images and transcripts will ultimately be stored in UCL's digital repository, thus reuniting Bentham's papers for the first time since his death.

Taking into account feedback from volunteers, and following work from ULCC, an upgraded Transcription Desk was launched on 15 July 2013, offering significant improvements to the user. Administrative processes, including the uploading of images, mapping them to the relevant metadata and the tiling of images for incorporation into the image viewers have also been automated. The code for this interface is available on an open-source basis, as a documented package for others to use and customise.[38]

The upgraded Transcription Desk aims to make transcription more straightforward for volunteers. Changes include an image viewer which allows volunteers to rotate the manuscript image, given Bentham's not uncommon habit of writing into or up the margin of a page, and at unusual angles (see Figures 3.4 and 3.5).[39] In order to take advantage of as much screen space as possible, 'maximise' and 'minimise' buttons have been added to the transcription interface; clicking the former clears from the screen all extraneous matter to show as much of the image as possible.[40] Other features are forthcoming, including making it more straightforward to select material to transcribe, and automated reporting of project statistics, which are currently manually compiled.

Perhaps the key change is the introduction of a tabbed transcription interface (see Figures 3.4 and 3.5): the 'Wikitext' tab displays the transcription area; 'Preview' generates a live preview of how the encoded transcript will look when saved; and the 'Changes' tab displays a highlighted list of changes the volunteer has made to the transcript. We hope that this will allow volunteers to understand better and more easily how the TEI mark-up works by allowing them to switch, at the click of a button, between their encoded transcript and a rendered preview. In the previous version of the Transcription Desk, participants had to save their work and leave the transcription interface to see their transcript rendered, causing a potential loss of concentration and making it much more difficult to compare the

---

38   https://github.com/onothimagen/cbp-transcription-desk (last accessed July 15, 2013).

39   For an extreme example, see http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/079/047/001 (last accessed July 30, 2013).

40   Bentham coined the terms 'maximise' and 'minimise'. Other Benthamic neologisms in general usage include 'international' and 'codification'; less widely adopted were, for example, 'circumgyration' and 'jentacularisation' (both for jogging, of which Bentham was a proponent).
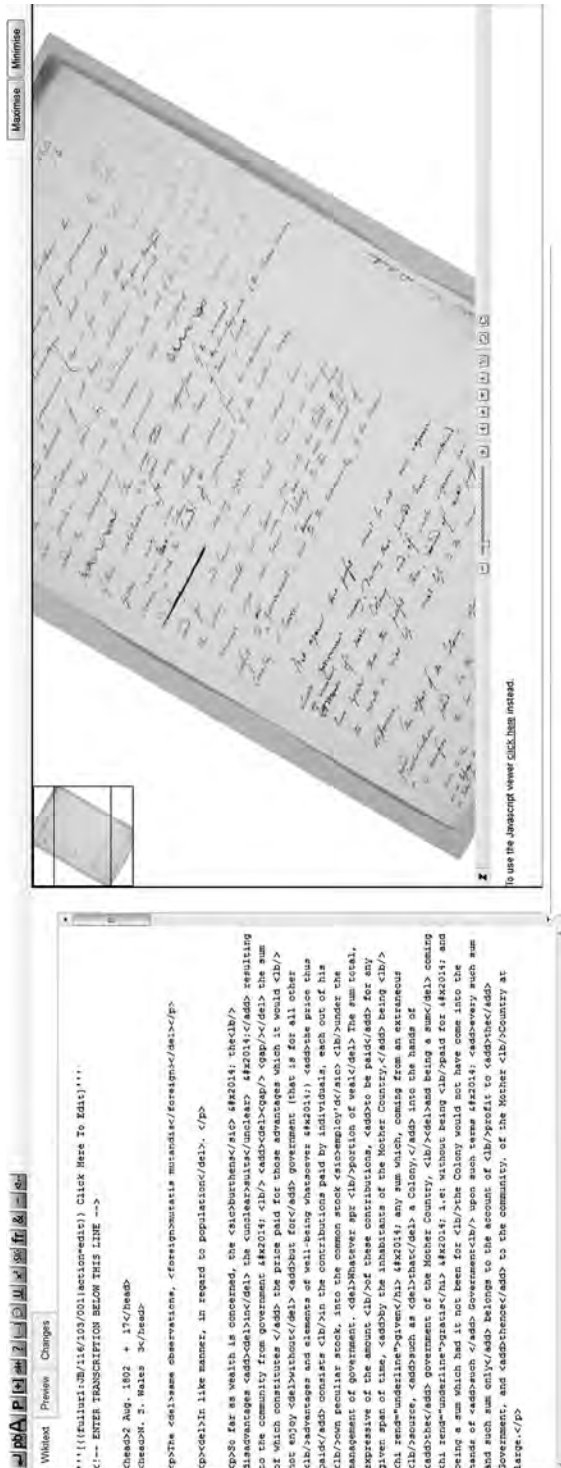
Figure 3.4 Upgraded Transcription Desk in 'maximised' mode, showing rotated image, transcription toolbar and tabbed transcription interface

Figure 3.5   Upgraded Transcription Desk in 'maximised' mode, showing rotated image and preview of encoded transcript

mark-up with the representation of the text. Likewise, comparing changes made to an earlier version of a transcript involved leaving the interface, and then entering the given page's revision history.

By making it more straightforward to see the functioning of the mark-up, it is hoped that the tabbed transcription interface will reduce the number of encoding errors, and further increase the efficiency of the quality-control workflow, particularly when dealing with lengthy manuscripts with a complex structure. Early indications suggest that the quality-control process is indeed quicker with the tabbed interface, though data need to be gathered over a significant period before firm conclusions can be drawn. We anticipate that the upgraded interface will help to increase user recruitment and retention, making it easier for new volunteers to participate, while also supporting the work of experienced transcribers. Feedback from volunteers suggests that they regard the upgraded website as cleaner, faster and more inviting. According to several transcribers, the text in the transcription box is better spaced and easier to read, the tabbed interface allows much greater flexibility and the rotatable image viewer is a boon. In particular, the 'maximise' button and the expanded transcription area it provides are much appreciated, especially by those using laptop computers.

While the improvement work was carried out, volunteers continued to transcribe at a healthy pace. As of 19 July 2013, 2,934 accounts were registered with *Transcribe Bentham*.[41] A total of 5,799 manuscripts (c. 2,800,000 words) had been transcribed or partially transcribed, of which 5,528 (95 per cent) were complete. Although, the transcription rate slowed a little during this period to an average of 34 manuscripts (c. 17,000 words) each week, *Transcribe Bentham* is currently in a healthier state than ever before.[42] In November 2012, the initiative received another award, coming second in the 'Platforms for Networked Innovation' competition, run by *KNetworks* project.[43]

The methodology and expertise developed in *Transcribe Bentham* will also be utilised and tested further in a connected project, entitled *tranScriptorium*.[44] This scheme is funded by the European Commission's Seventh Framework Programme (1 January 2013 to 31 December 2015), in the 'ICT for Learning and Access to Cultural Resources' challenge, and aims to develop innovative, efficient and cost-effective solutions for the indexing, searching and full automated transcription of manuscript images, using Handwritten Text Recognition (HTR) technology. *tranScriptorium* is led by the Universitat Politènica de València (Spain), with

---

41   This does not include project staff, robots and 647 blocked spam accounts. Spam on the Transcription Desk manifests as the creation of pages with links to commercial websites. All spam accounts and pages are blocked and deleted.

42   The latest progress statistics are updated on a weekly basis at http://blogs.ucl.ac.uk/transcribe-bentham.

43   http://blogs.ucl.ac.uk/transcribe-bentham/2012/11/22/transcribe-bentham-receives-award-in-knetworks-competition/.

44   http://transcriptorium.eu (last accessed February 20, 2013).

a consortium comprised of the University of Innsbruck (Austria), the National Center for Scientific Research 'Demokritos' (Greece), the Institute for Dutch Lexicology, UCL and ULCC. UCL will provide images and TEI transcripts of Bentham manuscripts and will, with ULCC and the other partners, develop and implement a crowdsourcing platform in which automated HTR transcripts of English, Dutch, German and Spanish manuscripts will be made available. Volunteers will be asked to correct these transcripts and help ensure that the software's future results are more accurate.

Incorporation of HTR technology into *Transcribe Bentham* affords exciting possibilities, though it is not without risk for a project with an established group of users. There is, for example, the danger that regular transcribers might feel their skills and role are devalued, and become alienated, if they believe they are being replaced by a machine. However, initial discussions with regular *Transcribe Bentham* participants suggests they would view HTR technology as being complementary to their work, and that it may even encourage an element of productive competition, as volunteers attempt to 'beat the computer' for accuracy.

Early results suggest that the HTR software can produce accurate transcripts of legible and standardised Bentham material, but it may cope less well with more complex manuscripts and their innumerable deletions and additions, and with material composed towards the end of Bentham's life when both his eyesight and handwriting deteriorated. In this scenario, we envisage that incorporating HTR into *Transcribe Bentham* will allow two tasks to be offered: full transcription of manuscripts by engaged volunteer transcribers, as happens now using the existing transcription interface; and correction of HTR-generated transcripts by text correctors who may not have as much time to devote to the project. There may be significant cross-over between the two groups: transcribers may wish to do the more straightforward task as light relief or when time is short, while the text correctors may wish to test their skills by moving on to full transcription.

More generally, *tranScriptorium*'s HTR software promises to be an extremely exciting development in making vast swathes of digitised manuscripts discoverable to the public, and we are delighted to be part of the project.

Participation

Crowdsourcing projects, from Wikipedia, to *Galaxy Zoo*, to the National Library of Australia's newspaper text-correction programme, have found that although they may have thousands of registered volunteers, most work is in fact done by a minority of users. *Transcribe Bentham* is no different: though 2,934 users had registered with the project by 19 July 2013, only 382 (13 per cent) had transcribed a manuscript, or a part thereof. Of those who did participate, almost two-thirds worked on only a single manuscript (see Table 3.1).

Every single contribution to *Transcribe Bentham* is greatly appreciated, whether it is the transcription of a sentence or an entire page. But the fact

Table 3.1   Number of manuscripts worked on by volunteers, 8 September 2010 to 19 July 2013

No. of manuscripts worked on	No. of volunteers (*percentage*)
0	2,552 (*86.9*)
1	238 (*8.1*)
2	68 (*2.3*)
3	25 (*0.9*)
4	6 (*0.2*)
5 to 20	26 (*0.9*)
21 to 50	5 (*0.2*)
51 to 100	6 (*0.2*)
101 to 200	2 (*<0.1*)
201 to 999	3 (*0.1*)
1,000+	3 (*0.1*)
Total	2,934 (*100*)

Note: 'Worked on' is defined as the volunteer having transcribed at least some part of a manuscript, and having clicked 'save' at least once to register their edit/edits.

remains that the great majority of the work has been carried out by a core of 17 'Super Transcribers' (see Table 3.2), of whom 10 currently participate. These expert volunteers sustain the project, and it is one of *Transcribe Bentham*'s great strengths that such dedicated, skilled participants submit high-quality work on a regular basis; in several cases, Super Transcribers have now transcribed more Bentham manuscripts than some Bentham Project staff. For instance, by 19 July 2013 volunteer Jfoxe had worked on 1,444 manuscripts (c. 722,000 words), Diane Folan 1,201 (c. 600,500 words) and Lea Stern on 1,044 transcripts (c. 522,000 words).

However, heavy reliance upon Super Transcribers does leave *Transcribe Bentham* in a precarious position: if one or more ceased participating, then the transcription rate would decrease precipitously. That the overall level of participation among all registered users is so low would indicate that a great many have found the task at hand to be too complex (a conclusion supported by our user survey), and that improvements had to be made to the user interface to attract more volunteers willing and able to participate regularly.[45] The reliance upon Super Transcribers also suggests that to say *Transcribe Bentham* is crowdsourcing is a misnomer, as the project does not have thousands of active users carrying out small tasks. *Transcribe Bentham* might be better described as 'crowd-sifting': that

45   Causer et al., 'Transcription Maximized', 127, and 130–33, and Causer and Wallace, 'Building a Volunteer Community'.

Crowdsourcing our Cultural Heritage

is, 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. We continue, as a result, to cast our net as wide as possible in the attempt to find more participants to join the cohort of Super Transcribers.

Table 3.2   Contributions of *Transcribe Bentham*'s Super Transcribers, 8 September 2010 to 19 July 2013

Username	Began participating	Currently participating?	Location	No. of manuscripts transcribed (*percentage of 5,799*)	Average no. of mss worked on per week (rounded)
Diane Folan	22 Sept 2010	Yes	UK	1,201 (*20.7*)	8
Carno	28 Dec 2010	No	US	89 (*1.5*)	7
Lidunn	28 Dec 2010	No	US	75 (*1.3*)	2
Mfoutz	29 Dec 2010	No	US	91 (*1.6*)	1
Clarabloomer	30 Dec 2010	No	US	71 (*1.2*)	8
RexL	31 Dec 2010	Yes	US	112 (*1.9*)	1
Lea Stern	4 Jan 2011	Yes	US	1,044 (*18*)	8
Jancopes	4 July 2011	Yes	US	201 (*3.5*)	2
Duyfken	1 Aug 2011	No	Australia	38 (*0.7*)	38[a]
Jillybean	15 Aug 2011	No	UK	106 (*1.8*)	4
Calico-pie	11 Sept 2011	No	France	37 (*0.6*)	3
Ohsoldgirl	11 Sept 2011	Yes	UK	320 (*5.5*)	4
Jfoxe	20 Sept 2011	Yes	UK	1,444 (*24.9*)	15
OlgaNM	9 Oct 2011	Yes	UK	69 (*1.2*)	1
Petergh	11 Nov 2011	Yes	UK	518 (*8.9*)	6
KeithThompson	21 Jan 2013	Yes	UK	90 (*1.6*)	4
Robmagin	29 May 2013	Yes	Canada	37 (*0.6*)	5

Note: [a]Duyfken transcribed 38 manuscripts over a two day period, accounting for the high weekly average.

Time and Money Well Spent?

A major concern about crowdsourced transcription is whether it is ultimately worth the effort. Would the time and money required to develop and deliver a platform, to recruit and manage volunteers and to check that crowdsourced submissions

are of a sufficient standard not be better off invested in simply employing experts to do the job? For example, staff at the University of Iowa's *Civil War Diaries* transcription project found that they spent a significant amount of time checking the work of volunteers, but did point out that this was not a prohibitive amount of time, and that the project as a whole would have been impossible without the cost-savings afforded by crowdsourcing. Sharon Leon (see also Chapter 4) likewise noted that while crowdsourcing 'makes new kinds of work for existing staff', engaging the public with otherwise underused resources is a worthwhile endeavour which could outweigh concerns over cost.[46] In this section, we will discuss the time and effort required to ensure that volunteer-produced transcripts meet the required quality standards.

In a paper published in 2012, we suggested that though *Transcribe Bentham*'s early results were encouraging and that the project facilitated engagement with an important resource and raised the profile of Bentham studies, checking submissions and managing the website was labour-intensive and time-consuming. During Period 1 (see earlier discussion), we found that the project's two full-time Research Associates each spent the equivalent of a month's full-time work checking submissions. Had they instead been employed to transcribe Bentham manuscripts on a full-time basis, they 'could have transcribed about 5,000 manuscripts between them over twelve months, or two and-a-half times as many as the volunteers would have produced had they continued transcribing at the same rate'.[47] (This calculation was made on the basis that volunteers then transcribed or partially transcribed an average of 35 manuscripts per week, whereas a full-time researcher could be expected to transcribe 40 to 50 manuscripts per week.) Based on these findings, David Weinberger was undoubtedly correct to suggest that: 'For now … the results of the Bentham project cannot be encouraging for those looking for a pragmatic way to generate high-quality transcriptions rapidly.'[48]

However, our finding is now out of date: *Transcribe Bentham*'s more recent results paint a much healthier picture. As mentioned above, at the end of Period 1 staff time spent on *Transcribe Bentham* was scaled back owing to the expiry of the AHRC grant. During Period 2, one Research Associate was responsible for *Transcribe Bentham* on a 0.4 Full-Time Equivalent (FTE) basis, that is, two days per week, one of which was spent on quality control. Despite the reduction in staffing, the transcription rate was higher than during Period 1: in Period 2, 3,403 manuscripts were transcribed or partially transcribed, at an average rate of 42 per week (see Table 3.3). At this rate, volunteers would produce 2,184 transcripts per year, whereas had the Research Associate devoted the two days per week he spent on *Transcribe Bentham* to transcription, he would produce only between 870 and 1,046 transcripts

---

46   Quoted in Zou, 'Civil War Project Shows Pros and Cons of Crowdsourcing'.
47   Causer et al., 'Transcription Maximized', 130–31.
48   David Weinberger, 'Crowdsourcing Transcription'.

Table 3.3　　Time spent on quality control process, 8 September 2010 to 19 July 2013

Period	Period length, days (weeks)	Manuscripts worked on (total)	Average no. of manuscripts worked on, daily (weekly)	No. of days spent checking transcripts[a]
Period 1: 8 Sept 2010–8 March 2011	182 (25)	1,009	6 (40)	50
Period 2: 9 March 2011–30 Sept 2012	572 (81)	3,403	6 (42)	81
Period 3: 1 Oct 2012–19 July 2013	291 (41)	1,305	4 (32)	17

Note: [a]Assuming one working day = 7.5 hours.

a year.[49] So, compared to Period 1, not only were manuscripts transcribed at a faster rate during Period 2, but the quality control process was more efficient.

The amount of staff time spent checking transcripts for Periods 1 and 2 are, however, only estimates. From 1 October 2012, a Research Associate returned to working on *Transcribe Bentham* on a full-time basis,[50] and since then the quality-control process has been measured in depth by recording several metrics. These are: the number of words transcribed, including and excluding TEI mark-up; the number of changes made to both the text and mark-up; the time spent checking each transcript before it was accepted; and the time spent correcting any errors in the mark-up after converting the transcript to an XML file using the *oXygen* XML Editor.

From 1 October 2012 to 19 July 2013, 1,394 manuscripts were worked on, and a total of 1,305 transcripts (94 per cent) were successfully submitted by volunteers for checking, at a rate of 31 submissions per week. The rate of transcription slowed a little during this period, compared to Period 2, owing to two main factors: the Christmas and New Year break, and three Super Transcribers – including the second-most prolific participant – took part less frequently than in Period 2, owing to personal and work commitments. These findings cover a nine-and-a-half month

---

49　Assuming: 52 weeks of five working days, less UCL's annual leave allowance of 27 days, and a further 15 days of public closures and bank holidays in 2013, at 0.4FTE = 87.2 days of transcription, at a rate of 10 to 12 transcripts per day.

50　Staffing was modified slightly from 1 March 2013: another member of staff began working on *Transcribe Bentham* on a 0.2 FTE basis, and the original Research Associate now works on a 0.8 FTE basis.

period, and though they might be regarded as provisional there are enough data to suggest reasonably that the trends identified below should continue.

Fifty-one individual volunteers submitted transcripts during Period 3, having collectively transcribed 419,464 words, or an average of 321 words per transcript not including mark-up; taking the mark-up into account, volunteers submitted a total of 588,203 words, or an average of 450 words per transcript. That the mark-up increases the number of words transcribed by almost 30 per cent is clear evidence that encoding is no small task.

Table 3.4   Summary of quality control process, 1 October 2012 to 19 July 2013

No. of individual submitters	No. of words, (incl. mark-up)	No. of alterations made to text	No. of alterations made to mark-up	Time spent checking transcripts (seconds)	Time spent correcting errors in XML file (seconds)
51	588,203	5,262	10,406	464,463	6,391

Transcribe Bentham's Research Associates spent the equivalent of 129 hours and 6 minutes' worth of labour checking submissions during Period 3, plus another 1 hour and 46 minutes correcting any errors which remained after converting the transcripts to XML files. On average, a given transcript took 357 seconds (5 minutes 57 seconds) to check, though there is, of course, great variation depending upon the length and complexity of the manuscript, the experience of the transcriber and the consistency of the mark-up. For example, the transcript of JB/051/271/003[51] comprised of 95 words (including mark-up) and took a mere 52 seconds to check: Bentham's handwriting is legible, the manuscript layout is unremarkable and the transcript was excellent. No changes were required to either text or mark-up.

At the other extreme, 1,931 seconds (32 minutes 11 seconds) were spent checking JB/050/135/001 (see Figure 3.6),[52] with 35 changes made to both text and mark-up. This was not because the transcript was of a poor quality; far from it. Rather, it was because the transcript was some 1,269 words long (including mark-up), and the manuscript is difficult to navigate owing to the numerous deletions, nested interlineal additions and multiple marginal annotations. After the transcript was converted to XML, it then took a further 14 minutes to locate a line-break

---

51 http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/051/271/003, revision dated 16.33, October 1, 2012.

52 http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/050/135/001, revision dated 15.37, October 15, 2012.
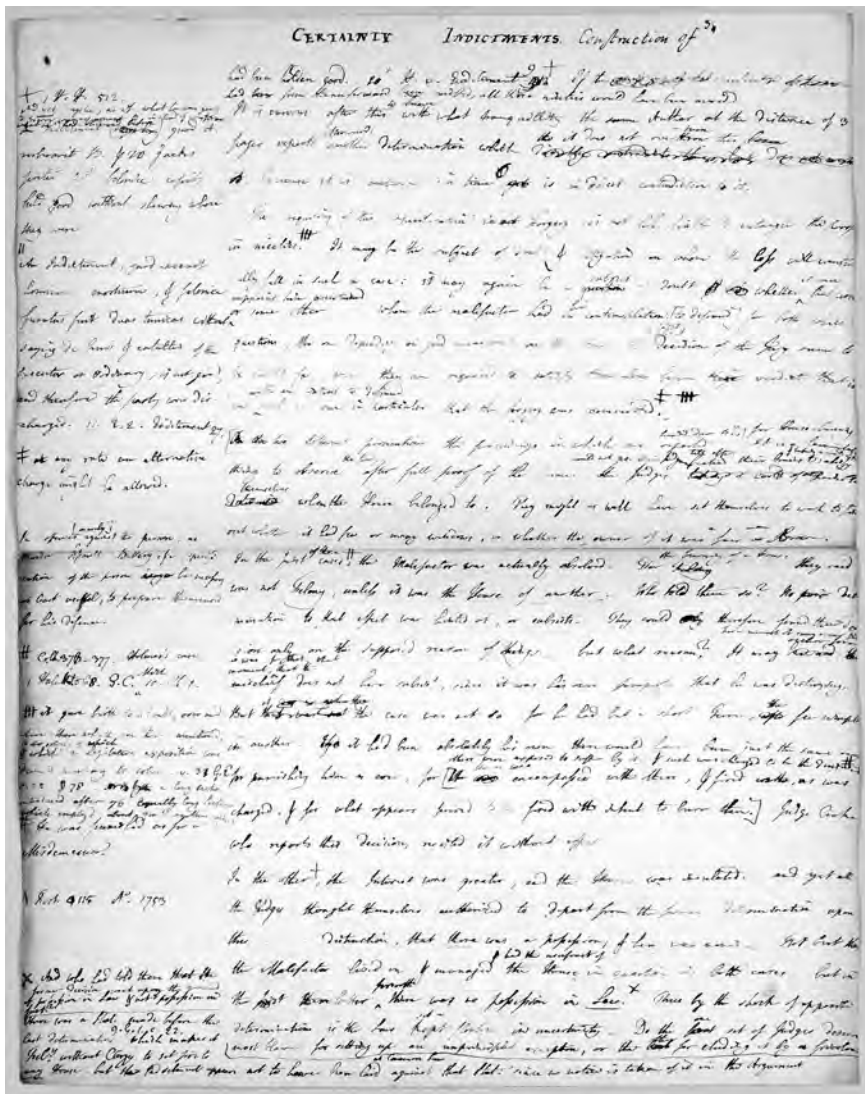
**Figure 3.6   Manuscript JB/050/135/001, courtesy UCL Library Special
Collections. Image taken by UCL Creative Media Services**

which had been mis-typed (not by the volunteer, we hasten to add), resulting in
oXygen XML Editor refusing to validate the file until the error had been located.
However, even in an extreme case like this, it is still far quicker for us to check a
transcript than transcribe it ourselves.

That a transcript can generally be checked in less than 6 minutes is extremely
encouraging (see Figure 3.7): it is a rare Bentham manuscript that we could
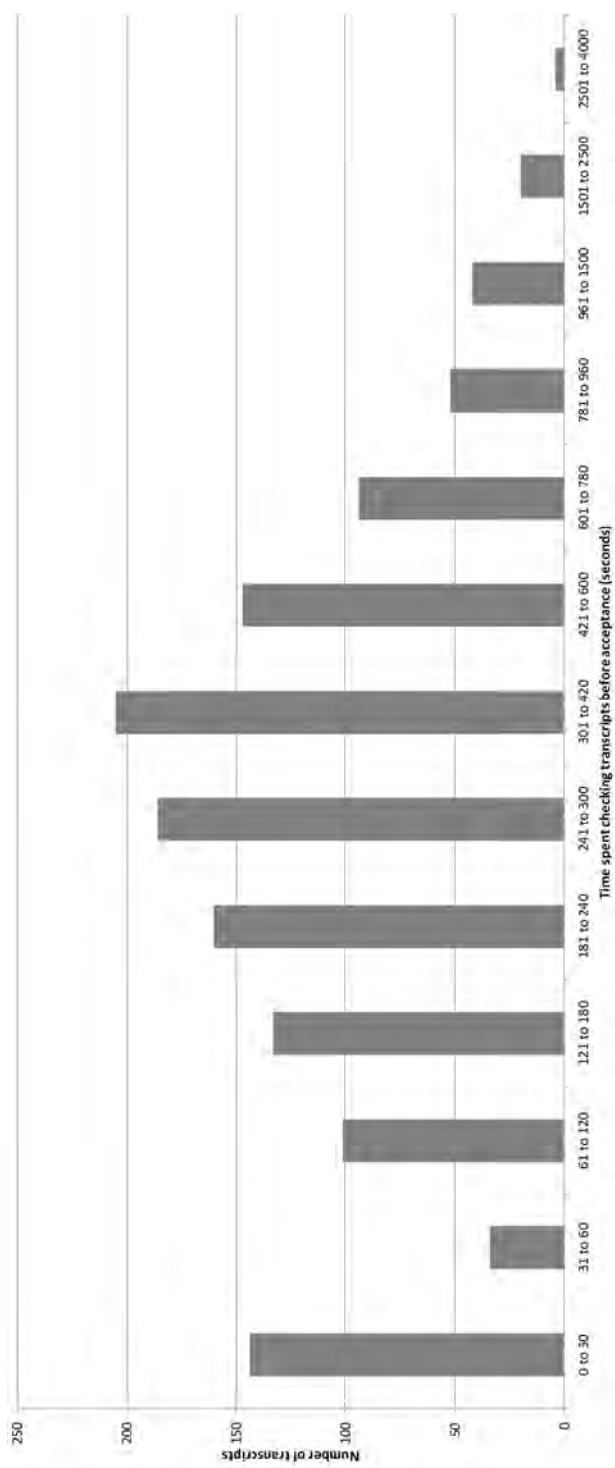
Figure 3.7   Time spent checking submitted transcripts, in seconds, 1 October 2012 to 19 July 2013

Note: Data were unavailable for six transcripts, owing to a software crash.

transcribe – let alone encode – in that space of time. However, about half the checking time is generally expended on ensuring that the XML mark-up is both well formed and valid. That such a relatively mundane and mechanical task can take such a disproportionate amount of time is less encouraging for *Transcribe Bentham*'s long-term cost-effectiveness, though the upgraded transcription interface (discussed above) should help to address this problem. In summary, the quality-control process is now much more efficient than before.

Quality: Just How Good Is the Work of Volunteer Transcribers?

Another regular concern about crowdsourced transcription is whether volunteers' submissions can be of a reliably high standard in comparison to the work of experts. In a project like *Transcribe Bentham*, where the material is frequently complex in both content and composition, and the crowdsourced results are used for public access and editorial purposes, some form of manual quality-control process is essential. Automated comparisons of transcripts would be difficult, and in the case of a lengthy, complex manuscript – such as JB/070/231/001[53] – it would be a waste of volunteers' time and effort to ask several of them to transcribe the same one, and of staff time to check and compare multiple variants.

We know that the vast majority of work done by *Transcribe Bentham* volunteers is of an extremely high quality, as 95 per cent of all transcripts have been approved by researchers experienced in reading Bentham's manuscripts. However, this figure does not illustrate what exactly happens to submissions during the quality-control process, and what alterations are required before a transcript is accepted.

Gaining an accurate representation of the manuscript text is the key task, and in this volunteer transcribers have proven more than able. Of the 1,305 transcripts submitted in Period 3, few required substantial changes to the text: 522 (40 per cent) were approved without alteration, and only 184 (14 per cent) needed eight or more changes before acceptance (see Table 3.5 and Figure 3.8).[54] Transcripts requiring extensive alterations to the text were few and far between: JB/002/567/001 required 69 changes because the marginal notes were not transcribed, and JB/100/001/001 required 213 alterations to the text, as the bottom-left panel of the manuscript was

---

53 http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/070/231/001, revision dated 14.39, April 23, 2013.

54 A change to the text is defined as: entering one untranscribed word; correcting one incorrectly transcribed word; moving an incorrectly placed word/portion of text. A change to the mark-up is defined as: addition/deletion of one piece of mark-up which only requires a single tag (e.g. <lb/>); addition/deletion of all or part of an opening and closing pair of tags (e.g. <p></p>).

untranscribed upon submission.[55] On average each submission received during Period 3 required only four alterations to the transcript's text before being approved.

The TEI mark-up, by comparison, causes more work for volunteers and staff: the 1,305 transcripts submitted during Period 3 required, on average, eight changes to the encoding. Only 313 (24 per cent) were approved without any modification of the mark-up, and 191 (15 per cent) needed only one change. A substantial number – 307 transcripts (24 per cent) required 10 or more alterations, and a disproportionate amount of time was spent on these lengthier submissions: 57 hours and 37 minutes – 45 per cent of all the time spent on moderation during Period 3 – was spent checking these 307 transcripts.

It would be fair to say that for volunteers, the XML mark-up complicates participation, and it has undoubtedly dissuaded many from participating more fully, or at all.[56] This conclusion should not detract from the efforts of Super Transcribers who have proven adept at text encoding, even though most had no experience of it prior to taking part in *Transcribe Bentham*. Rather, it is a warning

Table 3.5   Editorial intervention in manuscripts submitted between 1 October 2012 and 19 July 2013

No. of words	No. of transcripts	Average no. of changes to text	Average no. of changes to mark-up	Average time checking (seconds)
0	12	0	0	<1
1 to 25	24	<1	<1	16
26 to 50	6	0	<1	99
51 to 100	31	2	3	92
101 to 200	168	1	3	128
201 to 500	693	3	7	274
501 to 750	191	6	9	488
751 to 1000	92	7	14	645
1001 to 2000	79	12	25	1,053
2001 to 2999	8	16	19	1,807
3000+	1	66	3	729
Total	1,305	4	8	357

_____

55   http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/002/567/001, revision dated 14.52, October 22, 2012, and http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/100/001/001, revision dated 12.40, December 20, 2012.

56   See Causer and Wallace, 'Building a Volunteer Community' for volunteer opinions about the transcription process.
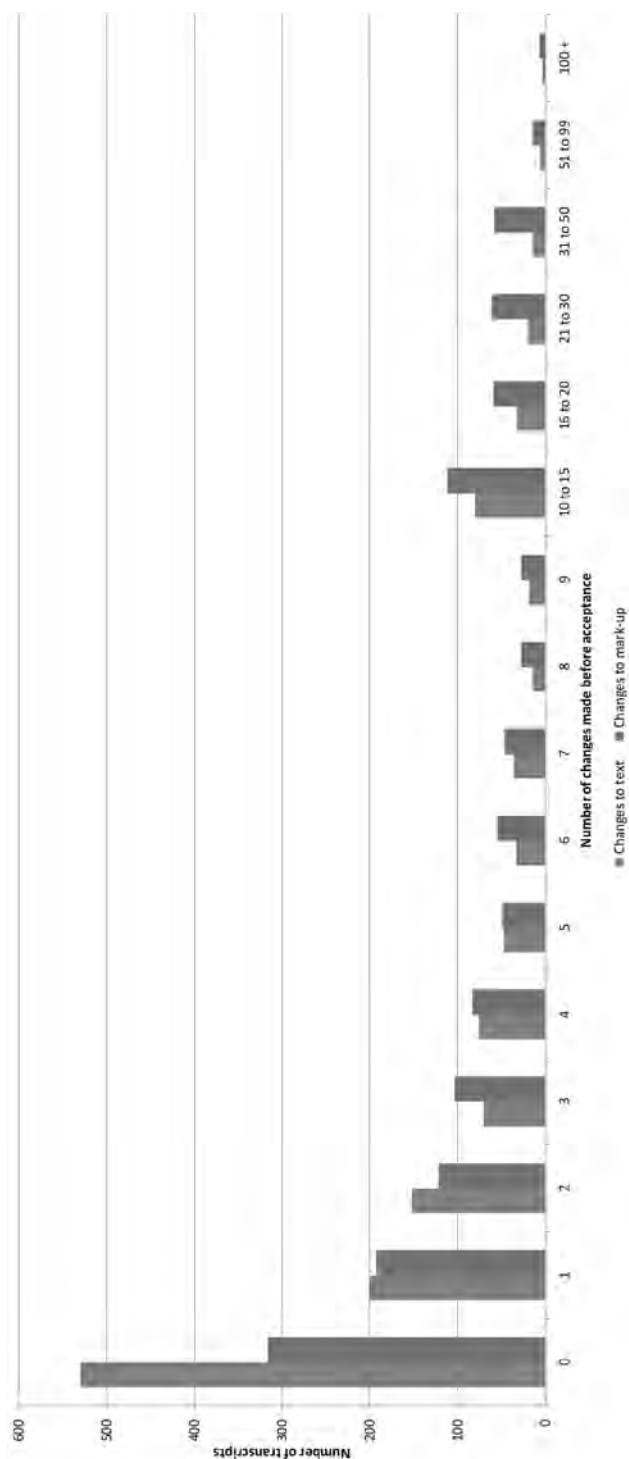
Figure 3.8   Changes made to text and mark-up of submitted transcripts, 1 October 2012 to 19 July 2013

that though crowdsourcing projects should not underestimate the capabilities of their audiences, nor should they test their volunteers' patience; the task at hand should be simplified as far as possible.

But why do some transcripts require more editorial intervention than others? First, the experience of the volunteer in question matters a great deal. The first few attempts at transcription are the hardest: acquiring an eye for Bentham's hand comes with time and practice (as Bentham Project researchers know only too well). During Period 3, 38 new volunteers transcribed a total of 188 manuscripts, which required on average 10 changes each to the text, and 20 alterations to the mark-up, before the transcripts were approved. By comparison, a Super Transcriber transcript requires, on average, four changes to the text, and seven to the mark-up.

The hand in which a manuscript was written makes a significant difference to the extent of editorial intervention. The majority of the Bentham Papers are written by Bentham, but substantial portions of the collection are neat, fair copies in the hands of secretaries and copyists, as well as portions of printed text (such as Acts of Parliament or newspaper clippings) which may or may not have been annotated by Bentham.[57] Manuscripts written by Bentham generally take more time to assess,

Table 3.6   **Quality control of submitted transcripts, 1 October 2012 to 19 July 2013[a]**

In whose hand?	Average no. of words (incl. mark-up)	No. of transcripts	Average no. of changes to text	Average no. of changes to mark-up	Average time checking (seconds)
Bentham	492	873	5	7	421
Bentham mostly, some by copyist	740	6	9	19	824
Copyist	335	343	3	9	207
Copyist mostly, some by Bentham	449	31	2	19	391
Printed text	581	29	3	10	222
Printed text, annotation by Bentham	635	12	<1	12	312

Notes: [a]Ten submitted transcripts were of blank sheets. [b]That manuscripts in the hands of copyists require an average of nine alterations to the mark-up before acceptance appears to be owing to the repeated misplacement of line-break tags.

57   For examples of manuscripts in the hand of a copyist, see http://www.transcribe-bentham.da.ulcc.ac.uk/td/Category:Box_107, folios 279–343, and of printed text, see http://www.transcribe-bentham.da.ulcc.ac.uk/td/Category:Box_116, folios 650 to 652 (both last accessed February 21, 2013).

usually contain more text and are often laid out in a less straightforward manner than the fair-copy manuscripts (see Table 3.5).

Finally, and unsurprisingly, the length of the document impacts greatly upon how long it takes to check. Transcripts submitted during Period 3 were an average of 450 words in length including mark-up, though 456 were of above average length. The above-average length transcripts, generally in Bentham's hand, required an average of seven changes to the text and 12 to the mark-up, and an average of 612 seconds (10 minutes 12 seconds) to check). By way of example, the transcript of JB/095/118/001 was 2,009 words long, difficult to follow, and required 32 minutes and 20 seconds to check.[58] On the other hand, transcripts of up to 750 words in length are generally checked in 4 minutes and 34 seconds or less (see Table 3.6).

Conclusion: The Worth of Crowdsourcing?

The amount of work put into the initiative by *Transcribe Bentham* volunteers should not be underestimated: they have braved a new experience in learning to read and transcribe Jeremy Bentham's handwriting, while at the same time adding TEI mark-up to their work. Transcribing Bentham is far from an easy task and we are extremely grateful to all of those who have committed to spending time with Bentham's manuscripts, in order to contribute to scholarship and widen access to the material. The success of *Transcribe Bentham* is owed fundamentally to their work, and is not taken for granted.

The first six months were undoubtedly the hardest for *Transcribe Bentham* (as they would be for any crowdsourcing project attempting to establish itself) and especially so as the material at hand is not necessarily well known or of immediate and popular appeal. Not only did volunteer transcribers experiment with a new system, but so, to an extent, did we: workflows for checking transcripts and maintaining the website had to be fully established, and as a result of this experimentation, the quality-control process proved more time-consuming than it otherwise might have been. Whilst a certain amount of time will always be spent checking submissions, this has been reduced dramatically as processes became easier and more familiar. There is also the possibility of recruiting experienced volunteers to moderate submissions as the project continues or if more volunteers take part regularly, but so far, the moderator/volunteer relationship has coped with the number and quality of submissions received.

With a supply of new material to explore, an upgraded transcription interface which addresses the concerns of volunteers, and a core of Super Transcribers, *Transcribe Bentham* is now at the stage where it can fulfil its potential, and the initiative's potential impact on the work of the Bentham Project (and, by extension,

---

58   http://www.transcribe-bentham.da.ulcc.ac.uk/td/JB/095/118/001, revision dated 11.38, December 18, 2012.

other institutions wishing to crowdsource transcription) can be starkly indicated. As mentioned earlier in this chapter, the UCL Bentham Papers consists of 60,000 manuscript folios, while the British Library holds another 12,500 folios. Without entering into a detailed discussion about how many of the manuscripts are folia and how many are bifolia, we estimate that the combined collection will require about 100,000 transcripts before it is fully transcribed.

Between 1984 and 2010, 20,000 folios were transcribed by Bentham Project staff – some 28,000 transcripts – at a rate of 1,076 per year, owing to the availability (or otherwise) of funding for editorial work. Assuming that sufficient funding was available for transcription to continue at this rate, the remainder of the UCL and British Library collections would not be fully transcribed for another 67 years.

Between 8 September 2010 and 19 July 2013, *Transcribe Bentham* volunteers worked at a rate of 2,024 transcripts per year. At this pace, the remainder of the collection would be completely transcribed by 2049 which, though some way off, is considerably faster than had *Transcribe Bentham* never existed. However, if the upgraded Transcription Desk and ongoing publicity campaign can recruit enough volunteers to produce between 75 and 100 transcripts per week (c. 3,900 to 5,200 per year), in this hypothetical scenario, the remainder of the collection could be transcribed in between 12 and 16 years. This still seems distant, but with continued support and volunteer effort, it is entirely possible that full digital access to all of Bentham's manuscripts and their transcripts could be provided within two decades, a prospect which was unthinkable just a few years ago.

Significant amounts of money – some £598,830 – have been invested in *Transcribe Bentham* by the AHRC and the Mellon Foundation, primarily on digitisation, software development and staff salaries, and it is only right to question whether this expenditure is worthwhile.[59] In the short term, *Transcribe Bentham* has produced nearly 6,000 transcripts in almost three years for this return, which does not sound all that impressive. Yet in the long run – and even taking into account a certain level of staff support to provide quality control, maintain the website and support volunteers – we estimate that should the remainder of the remaining 70,000 transcripts required to complete the collection be produced by volunteers, the Bentham Project could avoid staff costs of around £1,000,000, which will more than cover the investment in the initiative. Moreover, this does not take into account the incalculable public engagement value of *Transcribe Bentham*, and the creation of a hugely important searchable digital archive of Bentham's manuscripts.

Some institutions might consider outsourcing transcription, but for the Bentham Papers this is not a practical option, primarily because no organisation would ever give the Bentham Project sufficient funding to contract out the transcription of the tens of thousands of remaining untranscribed manuscripts: owing to their fragility and importance, the manuscripts themselves could not be sent offshore, and would still require digitisation (almost a third of the £598,830 invested in *Transcribe*

---

59   Robinson, 'Why Digital Humanists Should Get Out of Textual Scholarship'.

Bentham will be spent on digitisation). Given the difficulty in deciphering Bentham's manuscripts, the likelihood of the results being satisfactory without *Transcribe-Bentham*-style staff support and quality control, is rather doubtful.

It is difficult to overstate just how important and helpful it will be to the production of the *Collected Works* to have the content available in a digital, searchable format, providing an overview of the Bentham Papers which we have never had before. Nearly three million words have been transcribed by volunteers, providing us with a growing digital resource to build and experiment with, and it now seems time to consider using text visualisation techniques, intelligent search interfaces and text analysis to make full use of the richly encoded data created by volunteers, and discover new and insightful ways to explore the ever-growing corpus.

Transcribe Bentham would, we hope, have met with Bentham's approbation through the initiative's efforts to democratise the creation of, and access to, knowledge and humanities research, and its use of modern technology to enable the task to be completed in as efficient and timely a way as possible. 'Many hands make light work', wrote Bentham in 1783, but 'many hands together make merry work'. *Transcribe Bentham* continues to prove the truth of this particular maxim.

Acknowledgements

Funding

In regard to *tranScriptorium*, the research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007–2013) under grant agreement number 600707 – *tranScriptorium*.

References

Bentham, Jeremy. *Auto-Icon and Last Will and Testament*. Edited by Robert A. Fenn. Toronto: privately printed, 1992.

Bentham, Jeremy. *Not Paul, but Jesus*. Edited by Philip Schofield, Catherine Pease-Watkin and Michael Quinn, 2013. http://www.ucl.ac.uk/Bentham-Project/publications/npbj/npbj.html (last accessed May 1, 2013).

Bentham, Jeremy, *The Works of Jeremy Bentham*, published under the superintendence of his executor, John Bowring, vol. 4. Edinburgh: William Tait, 1838–43. Referred to as *Bowring* in the footnotes.

Causer, Tim and Valerie Wallace. 'Building a Volunteer Community: Results and Findings from Transcribe Bentham'. *Digital Humanities Quarterly* 6, no. 2 (2012). http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html (last accessed February 6, 2013).

Causer, Tim and Melissa Terras. 'Crowdsourcing Bentham: Beyond the Traditional Boundaries of Academic History'. *International Journal of Humanities and Arts Computing* 8, no.1 (2014): 46–64. Pre-publication version available from http://discovery.ucl.ac.uk/1354678/).

Causer, Tim, Justin Tonra and Valerie Wallace. 'Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *The Collected Works of Jeremy Bentham*'. *Literary and Linguistic Computing* 27, no. 2 (2012): 119–37.

Cohen, Patricia. 'Scholars Recruit Public for Project'. *New York Times*, December 27, 2010. http://www.nytimes.com/2010/12/28/books/28transcribe.html?pagewanted=all&_r=0 (last accessed July 26, 2013).

Kinchen, Rosie. 'One Stir, Then I'll Discover a Galaxy'. *Sunday Times*, September 11, 2011. http://www.thesundaytimes.co.uk/sto/newsreview/features/article772703.ece (last accessed February 20, 2013) (paywall).

Marmoy, C.F.A. 'The "Auto-Icon" of Jeremy Bentham at University College London'. *Medical History* 11, no. 2 (1958): 77–86.

Milne, A. Taylor, ed. *Catalogue of the Manuscripts of Jeremy Bentham, in the Library of University College*. 2nd Edn. London: Athlone Press, 1962.

O'Sullivan, Luke and Catherine Fuller, eds. *The Correspondence of Jeremy Bentham: Volume 12* (*Collected Works of Jeremy Bentham*). Oxford: OUP, 2006.

Pitkin, H.F. 'Slippery Bentham: Some Neglected Cracks in the Foundation of Utilitarianism'. *Political Theory* 18, no. 1 (1990): 104–31.

Quinn, Michael, ed. *Writings on the Poor Laws: Volume 1* (*Collected Works of Jeremy Bentham*). Oxford: OUP, 2001.

Quinn, Michael, ed. *Writings on the Poor Laws: Volume 2* (*Collected Works of Jeremy Bentham*). Oxford: OUP, 2010.

Robinson, Peter. 'Why Digital Humanists Should Get Out of Textual Scholarship'. Social, Digital, Scholarly Editing conference, University of Sasketchewan, July 11–13, 2013. http://www.academia.edu/4124828/SDSE_2013_why_digital_humanists_should_get_out_of_textual_scholarship (last accessed July 30, 2013).

Schofield, Philip. *Bentham: A Guide for the Perplexed*. London: Continuum, 2009.

Schofield, Philip. *Jeremy Bentham: Prophet of Secularism*. London, 2012. http://discovery.ucl.ac.uk/1370228/ (last accessed March 22, 2013).

Schofield, Philip, Catherine Pease-Watkin and Cyprian Blamires, eds. *Rights, Representation and Reform: Nonsense upon Stilts and Other Writings on the French Revolution* (*Collected Works of Jeremy Bentham*). Oxford: OUP, 2002.

Schofield, Philip, Catherine Pease-Watkin and Michael Quinn. *Of Sexual Irregularities, and Other Writings on Sexual Morality* (*Collected Works of Jeremy Bentham*). Oxford: OUP, 2014.

Semple, Janet. *Bentham's Prison: A Study of the Panopticon Penitentiary*. Oxford: OUP, 1993.

Stephen, Leslie. *The English Utilitarians: Volume I – Jeremy Bentham*. London: Duckworth and Co., 1900.

Weinberger, David. 'Crowdsourcing Transcription'. Too Big to Know blog. http://www.toobigtoknow.com/2012/09/04/2b2k-crowdsourcing-transcription-2/ (last accessed February 20, 2013).

Zou, Jie Jenny. 'Civil War Project Shows Pros and Cons of Crowdsourcing'. *Chronicle of Higher Education*, June 14, 2011 (updated June 21, 2011). http://chronicle.com/blogs/wiredcampus/civil-war-project-shows-pros-and-cons-of-crowdsourcing/31749 (last accessed February 7, 2013).