

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

Since 2010, the Bentham Project at University College London has been supported by a hard-working community of volunteers. With *Transcribe Bentham*, our efforts to involve the public in the production of a scholarly edition have enhanced the accessibility of the Bentham papers and helped to disseminate the results of an established research project outside the confines of academia.¹ But engaging with the volunteer community has not been without its challenges. Today, I would like to talk through our experience of drawing on the work of volunteers, how we currently support our users and how we hope to strengthen our relationship with our community in the future.

The Bentham Project is responsible for editing the writings of the British philosopher Jeremy Bentham.² Bentham was born in 1748 and by the time he died in 1832, he had written on a huge range of subjects including legal theory, crime and punishment, education, religion, democratic reform and sex. His most notable contribution to philosophical thought was undoubtedly his role in formulating the doctrine of utilitarianism. In one of his earliest published writings, he explained that ‘*it is the greatest happiness of the greatest number that is the measure of right and wrong*’.³ Bentham was also able to ensure that he would remain a subject of curiosity by using his will to decree how his body should be preserved and displayed after his death. His Auto-Icon sits in the corridors of University College London,

The author can be contacted at louise.seaward@ucl.ac.uk. She wishes to thank Professor Philip Schofield and Dr Tim Causer for their comments on this piece. The READ project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 674943. The tranScriptorium project received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement No. 600707.

¹ Transcribe Bentham: <http://blogs.ucl.ac.uk/transcribe-bentham/>

² The Bentham Project: <http://www.ucl.ac.uk/Bentham-Project>

³ J. Bentham, *A Comment on the Commentaries and A Fragment on Government*, ed. J. H. Burns and H. L. A. Hart, London, 1977 (*The Collected Works of Jeremy Bentham*), p. 393.

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

free for anyone to visit (or take a selfie with!).⁴ The Bentham Project seeks to recognise and explore Bentham's enormous impact on philosophical thought. The first edition of Bentham's works was published between 1838 and 1843 but this version was incomplete and poorly edited. The Bentham Project was founded in 1958 with the intention of producing a new scholarly edition of Bentham's published writings and unpublished manuscripts. The Project has completed 31 volumes and the final edition is expected to comprise 80 books.

For decades, the work of the Bentham Project was primarily confined to small group of expert scholars. This all changed in 2010 with the launch of *Transcribe Bentham*. A short-term grant from the Arts and Humanities Research Council allowed the project to test whether untrained volunteers could complete tasks usually undertaken by trained researchers. We are happy to report that this experiment proved successful and collaboration with the public is now an integral part of the work of the Bentham Project. Over the past five years, volunteers have transcribed close to 15,000 manuscript pages at a high level of accuracy.⁵

The digitisation of the Bentham papers is ongoing and volunteer transcribers will soon have access to every one of the some 75,000 sheets of Bentham's writings held at University College London and the British Library. The digital preservation of Bentham's papers is thus an important legacy of the *Transcribe Bentham* initiative. Users can access these digitised manuscripts at the Transcription Desk, a platform constructed by the University of London Computer Centre on the basis of a Mediawiki framework.⁶ This is collaborative open-source

⁴ For more on Bentham's Auto-Icon see <https://www.ucl.ac.uk/Bentham-Project/who/autoicon>

⁵ The transcribers have now surpassed 15,000 manuscript pages. See <http://blogs.ucl.ac.uk/transcribe-bentham/2016/01/29/progress-update-23-to-29-january-2016/>

⁶ To access the Transcription Desk: http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

software, which is also familiar to the public. Once they register for an account, users can browse the manuscripts for a specific subject or date, or simply pick a random page.⁷

There are many other crowd-sourcing projects which allow volunteers to make a contribution to research. Most famously, the Zooniverse platform has attracted more than 1 million users by hosting an array of citizen science projects such as counting galaxies, finding fossils and most recently, transcribing the papers of renowned artists.⁸ *Transcribe Bentham* is notable because it asks volunteers to undertake a particularly difficult task. Transcribers first have to decipher Bentham's handwriting, which becomes increasingly illegible in his later years as his sight was fading. They also have to deal with the revisions Bentham made to his papers, where words are added, deleted and changed in complicated sequences. The volunteers must then encode their transcripts using TEI mark up, the standard format for presenting electronic texts. The users tag features like paragraphs, marginalia, headings, additions and deletions using a toolbar. The use of TEI means that these transcripts can be stored, read and searched electronically. It also enhances the likelihood of their preservation by ensuring that they can be easily converted into different formats.

Each submitted transcript is checked by someone working at the Bentham Project.⁹ Editors need to see that the text has been accurately transcribed and that the TEI mark-up has been applied appropriately. If we are satisfied that no further improvements could be made by further crowd-sourcing, the manuscript is locked. This editorial check is subjective but

⁷ To browse the list of available manuscripts: <http://www.transcribe-bentham.da.ulcc.ac.uk/td/Manuscripts>

⁸ Zooniverse : <https://www.zooniverse.org/>

⁹ T. Causer and M. Terras, 'Many Hands Make Light Work. Many Hands Together Make Merry Work': *Transcribe Bentham and Crowdsourcing Manuscript Collections*, in *Crowdsourcing our Cultural Heritage*, ed. M. Ridge, Farnham, 2014, pp. 57-88 (p. 80-4).

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

necessary. It ensures that the transcripts form a suitable basis for further editorial work, it provides a means of measuring the progress of transcription and ensures that individual volunteers have feedback on their contributions. Locked transcripts are converted into XML files, stored on a shared drive and uploaded to our open access digital repository.¹⁰

Since the project began, volunteers have worked on an average of 54 manuscripts per week. This rate has increased in recent years as new Bentham material has been uploaded to the Transcription Desk. Across November 2015, users were producing an average of 81 transcripts per week. This compares favourably with a single researcher, who could probably transcribe around 50 to 60 manuscripts in the same period.¹¹ By providing guidance, support and recognition we hope that we will encourage our existing volunteers to keep transcribing and continue to draw in new participants.

The Transcription Desk has a number of features designed to support our users. Transcribers can access guidelines on the process of transcribing and adding TEI mark-up.¹² Examples of Bentham's handwriting and external links to palaeography tutorials are also provided. The surveys that we have conducted suggest that our volunteers are interested in history and Bentham and like the feeling of contributing to a scholarly project.¹³ In order to maintain this interest, it is vital that users receive feedback – both about their individual transcriptions and about the project more generally. Once a transcript has been checked, the *Transcribe*

¹⁰ The Bentham Collection digital repository: http://digitool-b.lib.ucl.ac.uk:8881/R&?local_base=BENTHAM

¹¹ T. Causer, J. Tonra and V. Wallace, 'Transcription maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*', *Literary and Linguistic Computing* (2012), 119-37 (130-1).

¹² Transcribe Bentham Guidelines: http://www.transcribe-bentham.da.ulcc.ac.uk/td/Help:Transcription_Guidelines

¹³ T. Causer and V. Wallace, 'Building a Volunteer Community: Results and Findings from Transcribe Bentham', *Digital Humanities Quarterly*, 6, 2 (2012): <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

Bentham editors leave feedback on the volunteer's user page. Weekly blogposts keep volunteers up to date with the project, as do tweets and Facebook updates. The work of the transcribers will also be credited in forthcoming volumes of Bentham's collected works.

Crowston and Fagnot's work on virtual collaborations has suggested that there are three levels of contribution to such projects – initial, sustained and meta.¹⁴ *Transcribe Bentham* does not yet have highly engaged meta-contributors but it does have a committed corps of contributors at the second level, sustained or super-transcribers. The majority of the transcription on *Transcribe Bentham* has been carried out by around 26 volunteers. These super-transcribers have completed 96% of the finished transcripts.¹⁵ This striking statistic demonstrates how far *Transcribe Bentham* is dependent upon a small cluster of users. This is why we need to make improvements which will keep our existing transcribers happy and hopefully attract new participants. Uploading new and interesting material to the Transcription Desk, such as Bentham's personal correspondence, is one way we have been able to encourage users to take part more regularly.

When we asked our users what stops them from transcribing more, they cited the difficulty of Bentham's handwriting, the complexity of the TEI mark-up and a lack of time available to contribute.¹⁶ One strand of the READ project will focus on making *Transcribe Bentham* simpler for the users. We are in the process of developing a what-you-see-is-what-you-get

¹⁴ K. Crowston and I. Fagnot, 'The motivational arc of massive virtual collaboration', *Proceedings of the IFIP WG 9.5 Working Conference on Virtuality and Society: Massive Virtual Communities* (2008):

<http://floss.syr.edu/sites/crowston.syr.edu/files/MotivationalArc.pdf>

¹⁵ According to latest project figures from November 2015. For more on the super-transcribers see Causer and Wallace, 'Building a Volunteer Community' <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>

¹⁶ Causer and Wallace, 'Building a Volunteer Community' <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

toolbar for the Transcription Desk where the TEI tags will be hidden from the user. READ will be working on an e-learning app to help users understand Bentham's handwriting. In collaboration with our colleagues in the *tranScriptorium* project (the forerunner to READ), we have already begun to integrate Handwritten Text Recognition technology into *Transcribe Bentham*.¹⁷ The University of London Computer Centre has developed a client called TSX, which allows users to transcribe Bentham with the assistance of HTR technology.¹⁸ This is an open-source platform which can be used and adapted by other institutions. Users can ask TSX to suggest individual words and lines, or provide an auto-generated version of the entire transcript which they would then need to correct. The READ project aims to improve the accuracy of these word predictions and enhance the functionality of the TSX interface. HTR technology thus has the potential to facilitate the regular participation of new and existing users who are hesitant about reading Bentham's handwriting. It could also be good for our engaged super-transcribers who could tackle the more difficult manuscripts which are too complex to be read accurately by a computer at this time.

Bentham requested that his manuscripts be publicly exhibited after his death alongside his Auto-Icon.¹⁹ Thanks to the dedication of our volunteers, *Transcribe Bentham* is fulfilling Bentham's wish to make his manuscripts accessible. Images and transcripts are preserved and made available in a digital repository. *Transcribe Bentham* spreads awareness of Bentham's ideas and allows the public to take part in a scholarly project.²⁰ As an example, a

¹⁷ TranScriptorium project: <http://transcriptorium.eu/>

¹⁸ TSX platform: <http://www.transcribe-bentham.da.ulcc.ac.uk/TSX/>

¹⁹ J. Bentham, 'Auto-Icon; Or, Farther Uses of the Dead to the Living. A Fragment. From the MSS of Jeremy Bentham', 1832 (unpublished).

²⁰ For discoveries made by the volunteer transcribers, see T. Causer and M. Terras, 'Crowdsourcing Bentham: Beyond the Traditional Boundaries of Academic History', *International Journal of Humanities and Arts Computing*, 8, 1 (2014) 46-64.

The Crowd, the Volunteers and the Supertranscribers. Building and supporting an online user community for the Bentham Edition

Louise Seaward, Bentham Project, University College London

cookbook of Bentham's own recipes has recently been produced on the basis of volunteer transcripts.²¹ Crowd-sourcing has also increased the efficiency of the Bentham Project. Without the existence of *Transcribe Bentham*, Bentham's papers would not be completely transcribed until at least 2081.²² If the current rate of transcription continues, this task could be finished by 2035. We have many reasons to be pleased with the development of *Transcribe Bentham* but it is important to understand that running a crowd-sourcing project is not easy. Careful management is required to recruit and retain volunteers. Sufficient funding is also needed for staff time, digitisation and technical work. Nevertheless, we are hopeful that *Transcribe Bentham* has demonstrated the viability of crowd-sourcing a complex task and integrating the public into a long-established scholarly project. We look forward to seeing our connection with our volunteers deepen as we head into the READ project.

²¹ The Bentham cookbook: https://www.ucl.ac.uk/Bentham-Project/publications/Bentham_Cookbook

²² T. Causer, K. Grint, A.-M. Sichani and M. Terras, 'The Quality and Cost-Effectiveness of Crowdsourced Transcription', *Forthcoming article*