Transcribe Bentham
Dr Tim Causer, UCL Bentham Project

Transcribe Bentham is an award-winning crowdsourced transcription project, run by University College London's Bentham Project, in collaboration with UCL Library Services, UCL Learning and Media Services, UCL's Department of Information Studies and Centre for Digital Humanities, the University of London Computer Centre, and the online community.

The Bentham Project, founded in 1958, is producing the new scholarly edition of *The Collected Works of Jeremy Bentham*, which will replace the inadequate, incomplete and poorly-edited eleven-volume edition published between 1838 and 1843. Bentham (1748-1832) is one of the world's great thinkers, whose thought has had a profound impact and is still of contemporary significance. He is perhaps best known as the founder of the modern doctrine of utilitarianism, and Bentham's writings are important in the fields of economics and jurisprudence, and he was an influential critic of the theory of natural rights. He developed a theory of punishment and reward, advocated female suffrage and an international court of arbitration to ensure peace. He is perhaps best known for his panopticon prison scheme, which was given a central place in Michel Foucault's interpretation of the modern state.

Objectives

The new edition is based to large extent on edited transcripts of unpublished manuscripts, of which 60,000 folios are held by UCL Special Collections (c. 30 million words), and a further 12,500 by the British Library (c. 6 million words). Prior to the public launch of *Transcribe Bentham* in September 2010, at least 40,000 folios were untranscribed and their contents largely unknown, and thus our understanding of Bentham and his thought—of importance to anyone studying the eighteenth or nineteenth centuries—is incomplete. In short, the Bentham Papers are a source of enormous historical and philosophical importance, yet much of the collection remains unknown, let alone adequately studied.

In *Transcribe Bentham*, volunteers are asked to transcribe the material via a collaborative editing platform based on a customised MediaWiki, and encode their work in Text-Encoding Initiative XML. Volunteer-produced transcripts of the requisite standard have two purposes:

1. Completed transcripts will be uploaded to UCL's digital Bentham Papers repository and linked to the relevant image, thereby gradually opening up the undiscovered parts of the collection for study and research.

2. Making a valuable contribution to research: transcripts will also form the basis of new volumes of the *Collected Works*, by providing editorial staff with a head-start in the form of draft transcripts. This is an exciting prospect for the Bentham Project, and the humanities more generally: volunteer transcribers are active in the production of scholarship, rather than being merely the passive receivers of the fruits of humanities research. Volunteers will potentially make new discoveries, and are often transcribe manuscripts which have not been looked at since Bentham wrote them.

We consider 'engagement' in *Transcribe Bentham* to be a partnership between UCL and volunteers in order to meet these aims. This seems to have been taken to heart by the project's most regular volunteers; they take great care and pleasure in participating, and feel an evident sense of responsibility in being entrusted with the material and the crowdsourcing platform.

Contributors

As of 20 April 2012:

- 1,615 registered accounts
- 3,302 manuscripts transcribed or partially-transcribed (c. 1.6 million words, plus markup)
- 3,066 (93%) of these are complete
- Average of 38 transcripts per week (c. 19,000 words) from 8 Sept 2010 to present
- Since 6 January 2012, an average of 55 transcripts per week (c. 27,500 words).
- Only 304 (19%) registered users have transcribed material, and around two-thirds of these have worked on only one manuscript.

In common with other crowdsourcing projects, the majority of the work in *Transcribe Bentham* is carried out by a minority, fourteen 'super transcribers' responsible for working on 94% of all transcribed manuscripts. This is both a significant strength and weakness of *Transcribe Bentham*: a dedicated core group has been recruited who produce large volumes of work at an increasing pace, supplemented by a larger group of less frequent transcribers. On the other hand, the loss of one or more regular transcribers will be damaging, and the project has—thus far—been unable to engage the majority of those who registered accounts.

This may be owing partly to the nature of the material: transcribing the difficult handwriting, idiosyncratic style, and dense and challenging ideas of an eighteenth and nineteenth-century philosopher is more complex, esoteric, and of less immediate appeal than, for example, contributing to a genealogical or citizen science project. More importantly, however, it is clear from user surveys and feedback that to increase user recruitment and retention, the transcription process needs to be made more straightforward, specifically the XML encoding.

To stimulate participation, *Transcribe Bentham* utilises some game-like mechanics: a point system, rankings, and a leaderboard. Yet survey results suggest competition and recognition were unimportant in stimulating participation to volunteers, and of more influence was an interest in the material, in Bentham specifically, and a sense of public-spiritedness in opening

up the material to others. That said, the most active transcriber *was* partly motivated by recognition and competition, and others may have been subconsciously competitive: 'I'm usually a very non-competitive person', wrote one volunteer, 'but […] became inspired to do more when I find I'm losing my "position" [on the leaderboard] – it's ridiculous! But quite fun'. Crowdsourcing projects must be able to cater to a variety of motivational requirements.

Value

In terms of opening up the collection and widening access to it at a time of increasing research on Bentham by prominent scholars (e.g. Linda Colley, Richard Tuck, David Armitage), *Transcribe Bentham* is of great value to researchers and students. The project is also of value in having carried out an experiment in crowdsourced manuscript transcription.

Transcribe Bentham's value to the UCL's research will be more long-term, and will more than cover the cost of establishing the project. Significant staff costs will be avoided in producing future volumes of the *Collected Works*: if all relevant draft transcripts are available, we estimate this to be up to six months of a full-time researcher's time per volume. This will make the Bentham Project more competitive for grants to produce the volumes, increase the pace of publication, and having XML transcripts will make it more straightforward to switch to electronic publication, if required.

Main research outputs

- An in-depth examination of the development and methodology crowdsourced manuscript transcription, and how it can be incorporated into an editorial project: if Bentham's manuscripts can be successfully crowdsourced, then anything can!
- An examination of the motivations of the demographics and motivations of volunteers transcribers, contributing to the existing literature.
- Code for the 'Transcription Desk' released on an open source basis, for others to customise and re-use: http://code.google.com/p/tb-transcription-desk/
- UCL's free-to-access digital Bentham Papers repository, which will expand as time goes on: http://www.ucl.ac.uk/library/bentham
- At a later date: volumes of *The Collected Works of Jeremy Bentham*, which are based in part upon volunteer-produced transcripts

Other resources

- Bentham Project website: http://www.ucl.ac.uk/bentham-project
- *Transcribe Bentham* project website: http://www.ucl.ac.uk/transcribe-bentham
- *Transcribe Bentham* 'Transcription Desk': http://www.transcribe-bentham.da.ulcc.ac.uk/td/Transcribe_Bentham
- 'Transcription Maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*': http://llc.oxfordjournals.org/content/early/2012/03/28/llc.fqs004.short?rss=1&amp%3bssource=mfr