Transcribe Bentham!

The Bentham Project is a British Academy Research Project. Members of the project team describe both an innovative research method, and what's next in the 'Collected Works' pipeline.

THE BENTHAM PROJECT at University College London (UCL) has been working on the production of the new edition of *The Collected Works of Jeremy Bentham* for over 50 years.[1] Most people will know Bentham (1748-1832), the philosopher and reformer, as the originator of the Panopticon prison, and for requesting that his remains be preserved as an auto-icon (in which state Bentham is now displayed at UCL). But as well as being custodians of his corpse, UCL also owns a great deal of Bentham's corpus: UCL Special Collections holds around 60,000 manuscript folios (conservatively estimated at *c.* 30 million words), while the British Library has a further 12,500 folios (*c.* 6 million words). Many volumes of the *Collected Works* are – and will be – based upon edited transcripts of this material. Transcription is itself a very time-consuming task.

Thus far, 29 volumes of the new edition have been published, with several more at an advanced stage of publication (see page 37, and a total of around 25,000 manuscript folios have been transcribed. The majority of the Bentham Papers – a resource of great historical and philosophical importance – are therefore untranscribed and their contents largely unknown, and are otherwise only accessible to those able to visit UCL Special Collections. Our understanding of Bentham's thought –

together with its historical significance and its contemporary relevance – is therefore rendered at best provisional, and at worst a caricature.

Recently, and to help overcome these problems, the study of Bentham has moved firmly into the digital age. Since 2010, the Bentham Project has co-ordinated a pioneering and exciting public engagement initiative, which provides students, researchers, scholars, and the general public with access to Bentham manuscripts, and involves volunteers in research. Launched to the public in September 2010, the double award-winning 'crowdsourced' transcription project, *Transcribe Bentham*, recruits online volunteers from around the world to assist in the transcription of unpublished manuscripts written and composed by Bentham, on a range of subjects including economics, representative democracy, religion, law, convict transportation, and sexual morality (Figure 1).

At the heart of the project is a collaborative transcription platform known as the 'Transcription Desk', which is a customised MediaWiki – therefore instantly recognisable to the millions who use Wikipedia every day – programmed by our colleagues at the University of London Computer Centre. Volunteers are presented with a high-quality digital image of a manuscript, and a plain-text box into which they type their transcript (Figure 2).

Figure 1. *The Transcribe Bentham homepage.*

[1] See 'Thinking Around the Box: The Work of the Bentham Project', *British Academy Review*, 13 (June 2009), 46-50.
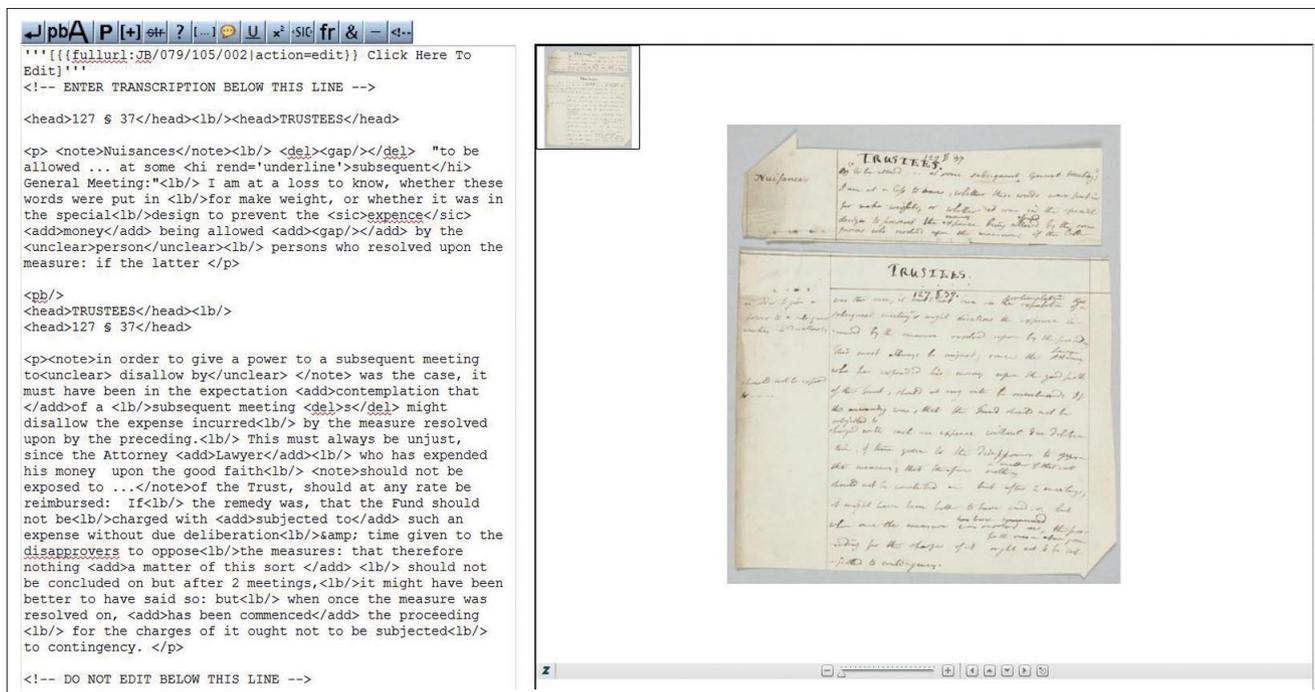
Figure 2. *The Transcribe Bentham online text editor.*

Transcribers are also asked to encode their work in Text Encoding Initiative (TEI)-compliant XML, using a specially-designed 'transcription toolbar' to facilitate the straightforward addition of mark-up to the text. (The TEI is a standard set of guidelines for scholarly text-encoding, and is important for preservation purposes: TEI XML files can be converted easily to any number of formats.)

Volunteer-produced transcripts are submitted to project staff who check them for textual accuracy and encoding consistency, and those transcripts judged to be complete have two purposes. First, they are uploaded to UCL's freely-accessible digital Bentham Papers repository, and linked to the relevant manuscript image to facilitate searching. Second, the transcripts will make a valuable contribution to research and scholarship, by forming a starting-point for future *Collected Works* (transcribers will be credited for their work in the relevant volumes). Volunteers frequently transcribe manuscripts that have not been read since Bentham composed them, and can potentially make new discoveries (see inset on page 36). *TB* is thus a partnership with volunteer transcribers who, rather than merely consuming the fruits of humanities research, play an active part in its generation.

'Crowdsourcing' is a term generally acknowledged to have been coined by the journalist Jeff Howe, in his 2006 article 'The Rise of Crowdsourcing' published in *Wired* magazine. Crowdsourcing generally involves seeking assistance in solving a problem from as wide a group of people as possible through an open call for volunteers.

In recent years, there has been a movement in cultural and heritage industries, and in academia, to trial crowdsourcing. A number of successful projects have harnessed the enthusiasm of volunteers to contribute to heritage, widen access to collections, and to help generate research. Some of the most successful include the National Library of Australia's digitised newspapers programme, in which genealogists and enthusiasts have corrected tens of millions of lines of Optical Character Recogntion (OCR) software-generated transcripts of digitised historic newspapers; *Dickens Journals Online* volunteers have done the same for the full run of both *Household Words* and *All the Year Round*, the periodicals edited by Charles Dickens; and, of course, *Galaxy Zoo* and its associated projects have involved innumerable 'citizen scientists' in generating scientific data for astronomical research.

There have also been attempts at crowdsourcing a more challenging task: the transcription of manuscripts. Such projects include *Old Weather*, which asks volunteers to transcribe ships' logs to map historic weather patterns; the American *Papers of the War Department* project; the *World Archives Project*, run by Ancestry.com; as well as open-source transcription tools such as *Scripto*, *T-Pen*, and *FromThePage*.

The manuscripts crowdsourced by these projects are, however, reasonably straightforward to decipher, and/or formulaic in layout, whereas many Bentham manuscripts cannot, with the best will in the world, be described as easy to read. *TB* was thus formulated as an experiment. Could volunteers – who may not have had any palaeographical training, or have encountered historical manuscripts previously – be able to read and decipher Bentham's handwriting, deal with compositional and structural features of the manuscripts and mark these up in TEI XML, and navigate both Bentham's often idiosyncratic style, and his dense and challenging ideas? In addition, would the work of volunteers be of sufficient quality to act

as a basis for the Bentham Project's editorial work, and to be uploaded to the digital Bentham Papers repository for public access?

More than two years after the project began, we are delighted to report that the answer to all of these questions is a resounding 'yes'. As of 7 December 2012, 4,728 manuscripts (*c.* 2.5 million words) have been transcribed or partially-transcribed by volunteers. Of these transcripts, 4,481 (94%) passed our quality control process, which is a real testament to the care and attention to detail paid by *TB* volunteers to their work, and is also an indication of no little skill on their part. The current rate of transcription is certainly impressive: since the beginning of 2012, volunteers have worked at a faster rate than a full-time researcher dedicated to transcription could be expected to, having submitted an average of 51 transcripts (*c.* 25,500 words) per week.

Over 2,500 people from around the world have registered with *TB*. Surveys of volunteers suggest that most are motivated to participate by an interest in history and/or philosophy, in Bentham, and by a general interest in crowdsourcing. Yet also notable are altruistic motivations: as one volunteer put it, beautifully encapsulating the ethos of the projct, *TB* 'is a literary form of archaeology', whereby 'instead of using a brush to uncover an object, you get to uncover historical information by reading and transcribing it. It leaves his legacy available for all to access.'

The following episode from Bentham's childhood is referred to briefly in John Bowring's biography of Bentham (volume X of the 1838–43 edition of the *Collected Works of Jeremy Bentham*), and the transcription of manuscript JB/027/026/004 has now confirmed where Bowring got the information from. This source describes Bentham's indiscretion in much more detail, and makes clear that Bentham considered it a formative experience in shaping his views on how animals should be treated.

I was about 4 or 5 years old when one evening I took it into my head to amuse myself with putting ear-wigs in the candle. I had no malice to the ear-wigs (poor insects): for my disgust at them if I had conceived any would have led me rather to have avoided them than to have handled them in that matter. But the writhings of their bodies, added to the little explosions made by the moisture of their juices composed a scene which amused my curiosity. A servant who had the charge of me, asked me what I thought the ear-wigs must suffer, and what I could have to say [for myself if any one that was stronger than I should] were she to serve me in the same manner. I was struck with remorse. I looked back upon the deed with horror: and from that time I have been nearly as attentive to the feelings of the brute part of the creation as of the human.
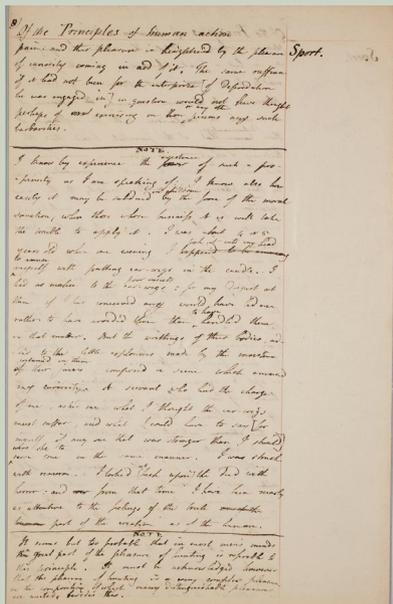
Manuscript JB/027/026/004, with a detail, right, showing the relevant portion of transcribed text.
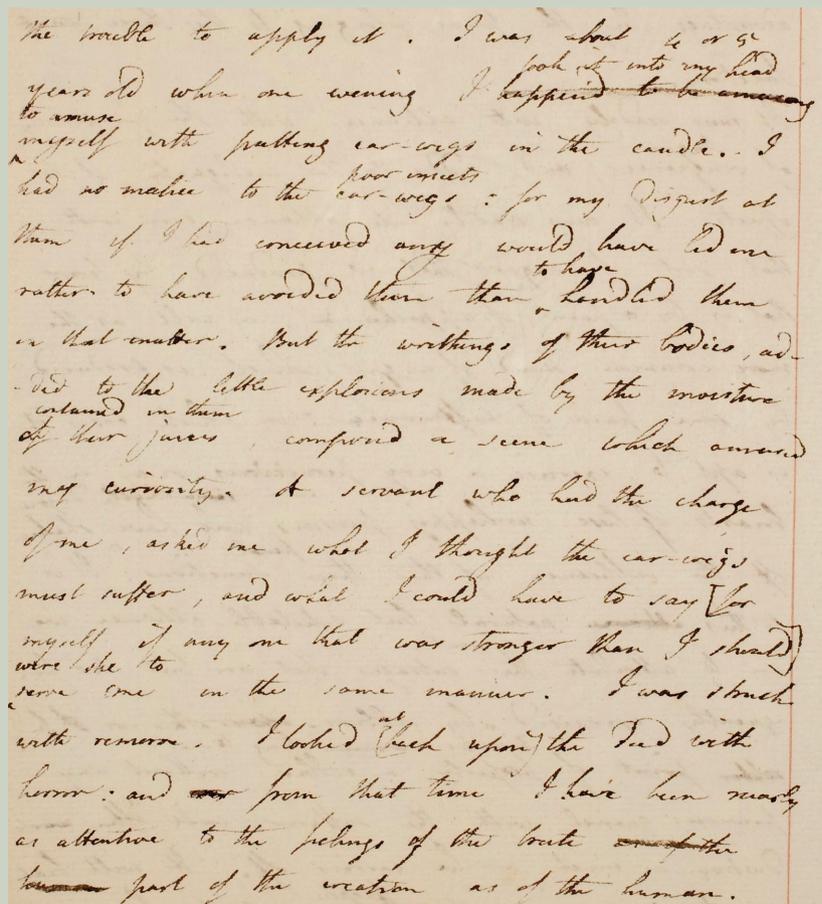
Yet, of all those who have registered with *TB*, only 318 have carried out any transcription, and most of these have only worked on one manuscript. The overwhelming majority of the work has been carried out by a core group of 15 'Super Transcribers', who submit contributions of an extremely high standard on a regular, ongoing basis. Transcribing Bentham manuscripts is always going to be of limited appeal, compared to correcting the text of historical newspapers, classifying galaxies, or transcribing ships' logs, but respondents to our survey of both regular participants and non-transcribers have offered suggestions as to what may have limited participation. Addressing these issues should enable us to increase our volunteer recruitment and retention.

Bentham's handwriting is a significant barrier to participation in itself (particularly for manuscripts written in the 1820s), while the TEI mark-up is seen as an unnecessary complication to transcription, and there are issues with identifying untranscribed material. So, while regular *TB* participants have overcome these obstacles, more work remains to be done to make participation straightforward, both for 'Super Transcribers' and those who may wish to dabble on a more irregular basis.

TB, and the work of its volunteers, has had a significant impact. It has featured in several international media outlets – including the *New York Times*, *Sunday Times*, Deutsche Welle World radio, and ORF1 radio (Austria). *TB* received an Award of Distinction in the 'Digital Communities' category of the 2011 Prix Ars Electronica, the world's foremost digital arts competition; in 2009, this same award was bestowed upon Wikileaks. The project was also shortlisted for the 2011 Digital Heritage Award, and recently came second in the *Knetworks* 'Platforms for Networked Innovation' competition. The code for the Transcription Desk is available on an open-source basis for download and re-use by others, and has been implemented by the Public Records Office of Victoria in Melbourne for their own transcription pilot project.

TB should now be able to build on its already considerable successes, because from 1 October 2012 the project is funded by a two-year grant from the Andrew W. Mellon Foundation's Scholarly Communications programme, as part of a wider scheme of work entitled the *Consolidated Bentham Papers Repository* (*CBPR*). Under this programme, we will digitise almost all of the UCL Bentham Papers collection, as well as all of the Bentham manuscripts held by the British Library. We will address the issues with the transcription interface that have been identified by volunteers; and, in perhaps the most useful planned innovation, for participants and potential participants alike, we will introduce a What-You-See-Is-What-You-Get transcription interface – volunteers will then be able to concentrate on transcription with the TEI mark-up being hidden and not obscuring the transcript as at present.

We believe that Bentham would have heartily approved of *TB*. The project is bringing a wider audience to the life and work of Bentham, who incidentally loved technology: his house had central heating, and he experimented with refrigeration, conversation tubes, and counterfeit-proof banknotes, amongst other things. *TB* has also demonstrated that the abilities of the 'crowd' should not be underestimated, and that complex tasks and material can be crowdsourced; if we can crowdsource Bentham, then surely anything is possible!

Perhaps most importantly, the ultimate fruit of *TB* will be a digital collection of enduring national and international historical and philosophical importance, accessible to all, and which will be created through a genuine partnership between scholars and the public.

To take part in Transcribe Bentham, you can create an account at the Transcription Desk. (www.transcribebentham.da.ulcc.ac.uk/td/ Transcribe_Bentham). You can also follow the project's latest progress at the Transcribe Bentham blog (www.ucl.ac.uk/transcribe-bentham), 'like' the project's Facebook page (www.facebook.com/TranscribeBentham), follow @transcribentham on Twitter, and contact the project by email (transcribe.bentham @ucl.ac.uk). The digital Bentham Papers repository is available at www.ucl.ac.uk/library/bentham

Transcribe Bentham is hosted by UCL's Bentham Project (a British Academy Research Project), and produced in association with UCL's Centre for Digital Humanities, UCL Library Services, UCL Creative Media Services, and the University of London Computer Centre. It was established under an initial 12-month grant from the Arts and Humanities Research Council. For two years from 1 October 2012, the project is funded by the Andrew W. Mellon Foundation, with the British Library joining the project consortium.

Latest editorial work on *The Collected Works of Jeremy Bentham*

Bentham's writings on religion and sexual morality

The next volume scheduled for publication will deal with Bentham's writings on sexual morality dating from the mid-1810s. This volume presents in authoritative form two substantial essays entitled 'Of Sexual Irregularities' and 'Sextus' (referring to sexual feelings as the sixth sense). Bentham argued that attitudes towards sexual morality in Britain had been based primarily on the Mosaic law and the teachings of St Paul, both of which condoned asceticism. More particularly, the story of the destruction of Sodom and Gomorrah provided the Biblical evidence for the view that God had issued an absolute prohibition against homosexuality, which was known as 'the crime against nature'. According to the Christian Church, sexual activity was sinful unless it involved one male and one female, within marriage, for the procreation of children. Bentham offered classical Rome and Greece, where certain male same-sex relationships were regarded as normal, as

alternative models of sexual morality. All that the term 'natural' meant in this context was 'approved by opinion' – no one sort of sexual activity was any more natural or normal than any other. The story of Sodom and Gomorrah did not condemn homosexuality, in Bentham's view, but gang rape. The key question was whether there was consent. No one, it was best assumed, would willingly consent to engage in a particular sexual practice – whether with a person of their own or the opposite sex, with more than one person, with an animal, with an inanimate object, with various parts of the body – unless they expected it to be pleasurable. Bentham argued that the Biblical record was rather more ambivalent than Christian apologists were prepared to recognise: the relationship between Jonathan and David in the Old Testament was clearly portrayed as homosexual, and he said there was evidence in the Gospels that Jesus was a practising homosexual. Bentham's argument was, in essence, that since sexual gratification constituted the most intense of pleasures, then there were no utilitarian grounds for condemning consensual sexual activity. Bentham was, moreover, persuaded by Malthus's argument that population growth tended to outstrip food supply. Hence, far from condemning non-procreative sexual activity, this was a reason for encouraging it, since it would help to reduce population growth. Bentham also has radical views on prostitution, infanticide, euthanasia, and sexual relationships between teachers and pupils. We expect that many people will reassess their view of Bentham in the light of this material.

Bentham's writings on political economy

In 2007 the editorial work began on two of a then projected four volumes of Bentham's writings on political economy, under a grant from the Economic and Social Research Council. Two volumes have indeed been edited, and significant progress has been made towards completion of a third, although the sheer amount of manuscript material has meant that a further three volumes at least will be required to complete this project. The works in the first two volumes were written between 1786 and 1795; they include previously unpublished texts, such as Bentham's marginal contents for his first systematic discussion of the subject, drafted in French for 'Projet Matière', and critical editions of texts drawn both from manuscript, such as 'Manual of Political Economy', and from printed works, such as *Defence of Usury* and *A Protest Against Law Taxes*. The second volume reflects Bentham's own focus, in the second half of 1794, on the subject of '*Πόροι*' (i.e. 'Resources' or 'Supply'). Under this head, he investigated the range of possible sources of supply for public expenditure, sought to rank them in order of preference, and drafted text for a range of proposals which he viewed as related branches of a single grand project, and which he intended to submit to government. In the event, of the nine proposals which he discussed, Bentham published only *Supply without Burthen*. It is hoped that its presentation in the context of the much larger project of which it was part will facilitate scholarship in the field of Bentham's political economy.