Crowdsourcing manuscript transcription: the *Transcribe Bentham* project

Martin Moyle, Justin Tonra, Valerie Wallace
UCL (University College London)

m.moyle@ucl.ac.uk

LIBER 2010, Aarhus, 29 June – 01 July 2010
Overview

• About *Transcribe Bentham*
• The transcription interface
• Encouraging participation
• Expected outcomes
• Next steps
Transcribe Bentham

• A 1-year project (from April 2010) harnessing the power of crowdsourcing to facilitate the transcription of 12,500 Jeremy Bentham manuscripts.

• Crowdsourcing:

Taking tasks traditionally performed by an employee or contractor, and outsourcing them to a group of people or community, through an "open call" to a large group of people (a crowd) asking for contributions. [Wikipedia]
Project origins

• 60,000 manuscripts of the philosopher and jurist Jeremy Bentham (1748-1832) held in UCL Library
  – Fully catalogued (http://www.benthampapers.ucl.ac.uk)

• UCL Bentham Project
  – Producing a complete scholarly edition of Bentham
    • Began 1959; 26 volumes now published, from a projected 68
  – 20,000 Bentham manuscripts previously transcribed
    • To varying degrees of quality; no standard markup
    • The majority of the manuscripts are untranscribed and unstudied
Project aims (1)

- Digitise 12,500 previously unread Bentham manuscripts
- Create a public transcription interface, with appropriate training tools, enabling crowdsourced TEI-encoded transcription
- Promote the project to specific target communities of volunteer transcribers
- Retrospectively convert existing transcripts to TEI
Project aims (2)

- Develop a web-based ‘Ideas Bank’, based on the transcripts
- Carry out log analysis and a user study on public interaction with the project
- Roll out a generic TEI transcription tool, for use by other transcription projects and services
- Long-term digital curation of digitised MSS and TEI transcripts in the UCL Library Services repository
Project partners

- UCL Bentham Project
- UCL Centre for Digital Humanities
- UCL Library Services
- University of London Computing Centre
- Arts and Humanities Research Council
- Jeremy Bentham
  - “present, not voting...”
Project components – overview...
Interface design: some challenges

- Transcription is hard!
  - Legibility; additions, deletions, marginal notes...
- TEI markup is complex for beginners
- Quality assurance is expensive, but to demand high quality from volunteers would be unrealistic
- Wiki environment may alienate some participants
Technical challenges: steps taken

• Help and guidance in different formats (web pages, video tutorials), and aimed at beginners
• Users shielded from the underlying complexity
• Accurate transcription – no markup - is welcomed
  – Users can begin to add markup as confidence grows
• Site is being user-tested and soft-launched
• Digitisation focusing on earlier, more legible MSS
The ‘Transcription Desk’ (beta)

Transcribe Bentham

Welcome to the Transcription Desk [TEST SITE]

The Transcription Desk is the heart of a major online initiative to transcribe the manuscripts of Jeremy Bentham from the archives of University College London. The project is funded by the AHRC, and managed by UCL Bentham Project and UCL Digital Humanities, with the support of the University of London Computer Centre.

Researchers and students interested in the works of Jeremy Bentham are invited to assist us by using the Transcription Desk to type up the text of the manuscripts. For each manuscript to be transcribed, the Transcription Desk shows a digital image of the manuscript and an online text editor to enter and edit your transcription of the text.

New users

- Create an account
- New users can find out more about the project and how to get involved.
- Watch a video lecture by Project Manager Prof Philip Schofield

Existing users

- Registered users can login here
- Transcription Guidelines
- Discussion Forum
- Transcribe Bentham Blog

Discussion Forum Info

- On Languages (0 Replies) (0 Views) (Tue 1st 4:33 pm - Justintonna)
- Which Markup Sh... (7 Replies) (29 Views) (Tue 1st 12:08 pm - Justintonna)
- Browsers & tran... (4 Replies) (14 Views) (Tue 18th 4:18 pm - RichardDavies)
- Making transcri... (0 Replies) (0 Views) (Tue 18th 3:26 pm - Justintonna)
- Test Thread (4 Replies) (0 Views) (Tue 13th 10:31 am - MelissaTeras)

Transcribe Bentham Right Now!

Transcribe Bentham is running MediaWiki version 1.15.1
It has 11 articles, and 221 pages in total.
There have been 718 edits.
There are 13 registered users, including 5 administrators.
There are 62 uploaded files.
This information is correct as of 13:55 on June 11, 2010.

This page last edited by Justintonna on 11/6/2010.
this second country should in [delite his circumstances][delite] be as different from England as it is possible.[/delite] Personnels[/delite] as a strong a contrast with England as it is possible.[/delite] On this and future we seem to have candidate primitive[sic] habit[sic] in Personnels[/delite]

[delite] diversity of climate mixture of inhabitants)[/delite] Climate, [delite] such as [delite] favor, [delite] climate[sic] of the country, [delite] natural productions. [delite] People[sic] customs, religion of the inhabitants were circumstances[/delite] on which a difference in the point[/delite] of view can be grounded, as different as is[/delite] can be. Add to which that there are two[/delite] countries a transfer of this kind in question[sic] has already been made or attempted to be made[/delite] in reality to European [suggestions] any two other[sic] examples that could be mentioned.[/delite] The [delite] problem[/delite] [delite] mistaken question[/delite] [delite] would have been a [delite] question of speculation, in regard to these[sic] to every thing that happening, whatever may happen to be[/delite]

Made of mixture [delite] cadet dore[/delite] [delite] use and apply immediately [delite] immediately [delite] [delite] directly [delite] to practice.[/delite] Bengal then is the country[sic] I am [delite] which I still fix my eye on [delite] process[sic] [delite] Tool[sic] Bengal [delite] let us
...
TEI Toolbar

• Hiding complexity

• Line Break - Paragraph - Addition - Deletion - Unclear Reading - Unusual Spelling - Ampersand

```
20 <text>
21   <body>
22     <div>
23       <head>Chapter II. <lb/> Regard to be paid to subsisting institutions</head>
24       <p> Looking over the <del> above </del> examples above <lb/> given, we shall find <add place="supralinear"> reason for dividing </add>
25   </div>
26   </body>
27 <text>
```

Transcribe Bentham
A Participatory Initiative
this country should be [its circumstances] regard to the circumstances in question be as

different from England as it is possible, as strong a contrast with England as is possible.

Such an example we seem to have in Bengal [diversity of climate, mixture of inhabitants]

Climate, size of the country, natural productions, recent laws, manners, customs, religion, of the inhabitants, were circumstances on which a difference in the point

in question can be grounded, as different as can be. Add to which that between these two countries a transfer of the kind in question has actually been made or attempted to be made in reality, in regard to any two other examples that could have been chosen

the previous question would have been a more

question of speculation: in regard to these

every living whatever it may happen to be made is of immediate use, and applies immediately to practice. (Bengal then is the country to which I still fix my eye in prospect.) To Bengal then let us appeal—the principal share of our attention; not precluding ourselves from watching into every one of them

immediately

new and then for the sake of variety a turn at glance towards other
**Additions**

In its simplest form, the + button in the toolbar is used to mark a part of the text that was added to the manuscript after the surrounding text was written. The exception to this is marginal additions, which are described below. Highlight the addition and click the + button to surround it with `<add>`/`</add>` tags, as in the example below:

![Image of handwritten text with addition](image)

```
whatever <add> just </add> remark may
```

**Deletions**

Where a word or a sequence of words has been deleted in the manuscript, highlight the relevant text and click the - button in the toolbar. This will surround the text with `<del>`/`</del>` tags.

![Image of handwritten text with deletion](image)

```
artificial: <del>tables of it's population: </del> tables of the
```

Exercise common sense when deciding on the extent of deletions. Where the strikethrough does not physically cancel a punctuation mark that is apparently part of the deletion, you may assume that it forms part of the deletion. If in doubt about a particular example, you may email the editors at transcribe.bentham@ucl.ac.uk.
“Profiles” for registered users

Profiles for registered users

Justintonra 4576 points (Apprentice)

Friends


Personal information

Real name Justin Tonra
Location United Kingdom
Occupation Research Associate, Bentham Project.
Websites http://www.ucl.ac.uk/transcribe-bentham

Board

RichardDavis (private)
posted 32 days ago
Still no sign of Melissa or Valerie or here :) Maybe Oliver and Anna should be on it too? And Philip needs an avatar! Keep prodding them, the more we get on here the better. BTW I only just noticed this public wall/private message part of the system... looks promising.

RichardDavis posted 32 days ago
Long-term access/preservation via Library repository
Sourcing crowds

• Three target audiences
  – Schools
    • Teachers nationally, especially 16-18 year-old level
    • Local schools, building on UCL’s outreach links
  – Academics
    • Educators in palaeography, research methods etc
    • Scholars in economic and social history, digital humanities, etc
  – Amateur historians, enthusiasts, and general public
• Different communications strategies in place for each group
Encouraging participation

• Targeting each group involves a combination of activities
  – Workshops, classes and presentations; paid-for advertisements in relevant print publications (eg History Today); approaches to disciplinary and professional bodies (eg IHR); press releases...

• Careful planning required
  – Publication lead times; academic cycle; short project!

• Web 2.0 activity...
Outcomes and impact

- Stimulation of public engagement with scholarly archives and manuscript transcription
- Opening up Bentham’s thought to new audiences
  - Policy makers, media, public
- Creation of an open access, digitally-preserved resource for scholars
- Availability of a re-usable, user-tested transcription tool for future projects and services
- How do users interact with digital resources?
  - Quantitative and qualitative data to help best practice
Progress / next steps

• Digitisation began April 2010
• Transcription Desk (beta) in user testing
• Soft launch, ~20 testers, July 2010
• Official launch August 2010
• Publicity campaigns begin August 2010
• Final report and user study May 2011
Thank you

- http://www.ucl.ac.uk/transcribe-bentham
- m.moyle@ucl.ac.uk