

# Crowdsourcing Cultural Heritage

## UCL's Transcribe Bentham Project



# Transcribe Bentham

A Participatory Initiative

Dr Melissa Terras

Senior Lecturer in Electronic Communication, UCL Dept of Information Studies

Deputy Director, UCL Centre for Digital Humanities

[m.terras@ucl.ac.uk](mailto:m.terras@ucl.ac.uk)

# Crowdsourcing Cultural Heritage

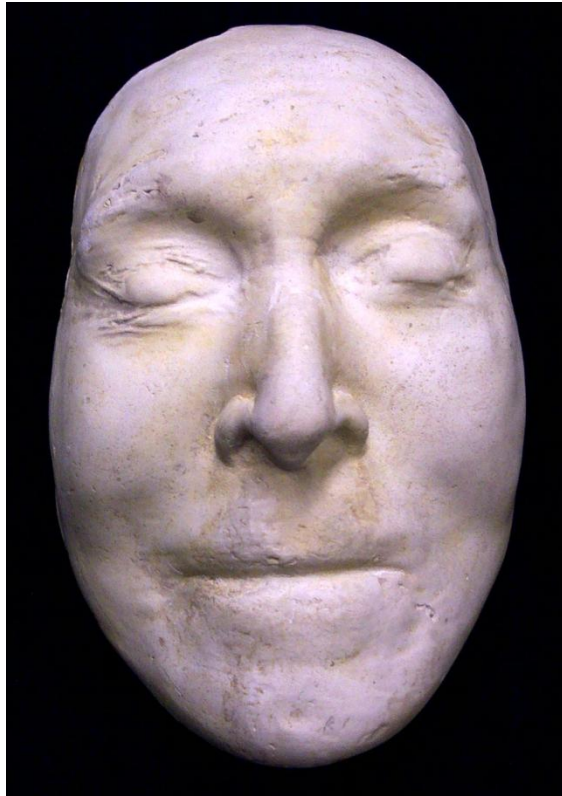
- Bentham and UCL
- Crowdsourcing
  - History and Ideas
  - Heritage and Culture
  - Features and Issues
- Transcribe Bentham
- Potentials and Problems

## Jeremy Bentham (1748-1832)



- Jurist, philosopher, and legal and social reformer
- Leading theorist in Anglo-American philosophy of law
- Influenced the development of welfarism
- Advocated utilitarianism
- Animal rights,
- Work on the “panopticon”
  
- Not founder of UCL, but...
- 60,000 folios in UCL Sp. Collections
- Auto-ikon

# The Bentham Project



- <http://www.ucl.ac.uk/Bentham-Project/>
- Since 1959
- “aims to produce a new scholarly edition of the works and correspondence of Jeremy Bentham”
- twenty six volumes of the new Collected Works have been published
- Previous AHRC grant catalogued the manuscripts
  - <http://www.benthampapers.ucl.ac.uk/>

### **Sheet:**

Box number	155
Folio number	097
Category	Correspondence
Recto / verso	recto

### **Associated people:**

Creator	Sir Edwin Chadwick
Recipient	162
Penner	Sir Edwin Chadwick

### **Dates:**

Dates appearing on page	1853
-------------------------	------

### **Headings:**

Sub-headings	Benthamiana / Notes to the London University College with the Library on Jurisprudence presented to them / 1853
--------------	---

### **Page:**

Number of pages	2
-----------------	---

### **Notes:**

Notes	Draft
-------	-------

# Des Browne MP's expenses

[Start reviewing pages](#)



Labour MP for  
Kilmarnock &  
Loudoun

[Guardian politics  
profile](#)

## Documents

Document	Total reviewed	Progress
<a href="#">Additional Costs Allowance 2004/05</a>	12 of 12 pages reviewed	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<a href="#">Incidental Expenses Provision 2004/05</a>	19 of 148 pages reviewed	<div style="width: 12.8%"><div style="width: 12.8%;"></div></div>
<a href="#">Additional Costs Allowance 2005/06</a>	16 of 16 pages reviewed	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<a href="#">Incidental Expenses Provision 2005/06</a>	19 of 222 pages reviewed	<div style="width: 8.6%"><div style="width: 8.6%;"></div></div>
<a href="#">Additional Costs Allowance 2006/07</a>	15 of 15 pages reviewed	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<a href="#">Incidental Expenses Provision 2006/07</a>	29 of 195 pages reviewed	<div style="width: 14.9%"><div style="width: 14.9%;"></div></div>
<a href="#">Additional Costs Allowance 2007/08</a>	24 of 24 pages reviewed	<div style="width: 100%;"><div style="width: 100%;"></div></div>
<a href="#">Communication Allowance 2007/08</a>	28 of 37 pages reviewed	<div style="width: 75.7%"><div style="width: 75.7%;"></div></div>
<a href="#">Incidental Expenses Provision 2007/08</a>	102 of 182 pages reviewed	<div style="width: 56.1%"><div style="width: 56.1%;"></div></div>

## Recent investigations

[eatmypoverity](#)

[gdw](#)

[anon-16048](#)

[anon-15556](#)

[chopper](#)

First 80 hours: 20,000 volunteers, 170,000 pages read.

Currently: 26, 717 volunteers, 220,965 pages read. 237,867 to go

# Crowdsourcing

- neologistic portmanteau of “crowd” and “outsourcing”
- coined by Jeff Howe in a June 2006 *Wired* magazine article “The Rise of Crowdsourcing”
  - Group intelligence
  - Cheap computers + large crowds = useful
  - “It’s not outsourcing; it’s crowdsourcing.”

## Technology and crowd-based research

- Often those outside established institutions that have taken the lead in exploiting new technologies
  - Science in the 19<sup>th</sup> century
  - Classics, maths, black studies, astrophysics, oral history, women's studies, contemporary history... all started outside established curricula
- Prizes for technological innovation
- Metal detectors/archaeology
- Binoculars/ ornithological fieldwork
- Cassette Recorders/ life history, oral history, language
- Telescopes/ astronomical research

## Crowdsourcing tasks

- The harnessing of online activity to aid in large scale projects that require human cognition
- Basic to complex tasks
  - Is this round or square? (yes/no)
  - Is this tag correct for this image?
  - Can you correct the OCR on this page?

## Crowdsourcing: Potentials for heritage institutions

- Achieving goals even with limited resources
- Achieving goals faster
- Build new virtual communities and user groups
- Involve and engage the user community with collections
- Utilising the knowledge, expertise and interest of the community
- Improving the quality of data/resource (e.g. corrections), more accurate searching
- Adding value to data (e.g. by addition of comments, tags, ratings, reviews).
- Making data discoverable in different ways f (e.g. by tagging).
- Gain insight on user desires by asking and then listening to the crowd.
- Demonstrating the value and relevance of the institution in the community
- Strengthen and build trust and loyalty of collection users
- Encourage a sense of public ownership and responsibility
- Holley, R. (2010) “Crowdsourcing: How and Why Should Libraries Do It?” D-Lib Magazine <http://www.dlib.org/dlib/march10/holley/03holley.html>

## Galaxy Zoo <http://www.galaxyzoo.org/>

- Online collaborative astronomy project
- Public assist in classifying millions of galaxies from digital photos taken by robots
- Released July 2007
- By August 2007 80,000 volunteers had classified 10 million galaxies
- To date, more than 60 million galaxies classified

# GALAXY ZOO

2

[Home](#) [How To Take Part](#) [My Galaxies](#) [Contact Us](#)

[Profile](#)

[Logout](#)



## Classify Galaxies

Answer the question below using the buttons provided.

**Is the galaxy simply smooth and rounded, with no sign of a disk?**




Smooth



Features or disk



Star or artifact

[Need help?](#) 

# Australian Newspapers Digitisation Program

<http://www.nla.gov.au/ndp/>

- In 2007 The National Library of Australia began to digitise out of copyright newspapers
- However the OCR quality of newsprint is poor
- Opened up the text to allow users to correct mistakes in the OCR
- 9000+ members of the public have so far corrected 12.5 million lines of newspaper text

## ELECTRONICALLY TRANSLATED TEXT


[Need help?](#)

[Keyboard Shortcuts](#)

 Save

 Save & Exit

 Cancel

 Undo Line

 Insert Symbol

Sicamor for South Hulkiston, Jervis

Bay...

THE SOPHIA JANE will

leave Jaques'» Coasting

Wharf, Sydney, for the above

Port, on the first Thursday in

August, and on the first Thura

day of every succeeding month, and will call as

usual at Wollongong and Kiama in \_ passing.

The steamer will leave South Huskisson on

**Steamer for South Huskisson, Jervis Bay.**



**THE SOPHIA JANE** will leave Jaques's Coasting Wharf, Sydney, for the above Port, on the first Thursday in August, and on the first Thursday of every succeeding month, and will call as usual at Wollongong and Kiama in passing. The steamer will leave South Huskisson on her return to Sydney, after sunset of the Sunday after her arrival.

Goods for the Steamer will be received at the store of Mr. A. Campbell, who will afford every facility to shippers from South Huskisson.

Freight, 25s. per ton; wool, 7s. per bale of the usual size; and grain 7d. per bushel. Apply to Captain Wiseman, on board; to Mr. C. T. Smith, Wollongong; or to

**MR. A. B. SPARK, Sydney.**

**Sydney, July 30.**

**1500**

Liverpo

The "SYDNEY MORNING HERALD" is Published intended to be inser

# Victoria and Albert Museum Crowdsourcing

<http://collections.vam.ac.uk/crowdsourcing/>

- Search the collections contains 140,000 images, selected automatically from the database
- Many images not the best view of an object
- Asking users to help find best crops of images
- 28375 images done in a year



Our progress



Your contribution



Choose image > Choose crop > Choose zoom level > Done

< Go back a step

Flask

Need to know more about this object?

No good image

Skip this object

Click the best crop:



















# Crowd sourced projects

- Picture Australia, National Library of Australia
  - <http://www.pictureaustralia.org/>
- Family Search Indexing
  - [http://www.familysearch.org/eng/indexing/frameset\\_indexing.asp](http://www.familysearch.org/eng/indexing/frameset_indexing.asp)
- Free BMD
  - <http://www.freebmd.org.uk/>
- Distributed Proofreaders (Project Gutenberg)
  - <http://www.pgdp.net/c/>
- Papyri
  - Project at Oxford to use Galaxy Zoo software to help in classification of documentary fragments
- Wikipedia
  - <http://www.wikipedia.org/>

# What do we know of Volunteers?

- Majority of work done by 10% of users
- Clay Shirky describes activity as 'cognitive surplus' time for social endeavours, rather than watching TV
- Personal interest
- Personal reward
- Community aspect
- Lot of interest from retirement community, and disabled and terminally ill individuals
- Many build up IT expertise as they volunteer
- “addictive”
- Help achieve group goal
- Like to be rewarded

# Successful Crowdsourcing

The Thing	The System	The People	The Content
 Clear goal	 Easy and Fun	 Acknowledge	 Interesting
 Big challenge	 Reliable and quick	 Reward	 New
 Progress	 Intuitive	 Team support	 Lots
 Results	 Options	 Trust	 History/Science

Rose Holley's checklist for crowdsourcing:

<http://www.dlib.org/dlib/march10/holley/03holley.html>

## Enter Transcribe Bentham

- 10,000 images of Bentham's manuscripts
- Ask user community to transcribe these
  - Provide plain text
  - Or “Markup” in rudimentary TEI
    - Underline, deletions, insertions
- Generate a “Knowledge Bank” of ideas from the transcripts
- Link with existing catalogue and transcripts
- Make material more accessible to scholars



# Plan

- Soft launch end of June
- Full launch early July
- In process of user testing and creation of system
- Two full time RAs working on this
  - One for user testing and promotion
  - One for user testing and technical aspects
- <http://www.ucl.ac.uk/transcribe-bentham/>

# User Interaction

- Involving users in the design process is key
- Currently recruiting for testers
- Will be working one to one with users
  - Established textual scholars from DH community
  - Members of the public
- Will open to Beta testing to find bugs
- Then onto full launch

**Navigation** Page

- [Transcription Desk](#)
- [Community portal](#)
- [Latest News](#)
- [Current events](#)
- [Discussion Forum](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

**Search**




## 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

- If you are a new user then you can [find out more about the project and how to get involved](#).
- [Watch a video lecture](#) by Project Manager [Prof Philip](#)

#### Existing users

- If you are a registered user you can [login here](#)
- [Community Portal](#)
- [Current Events](#)

#### Transcribe Bentham Right Now!

Transcribe Bentham is running [MediaWiki](#) version 1.15.1.

It has 9 articles, and 199 pages in total.

There have been 632 edits.

There are 12 [registered users](#), including 5 [administrators](#).

There are 44 uploaded files.

This information is correct as of 10:21 on May 28, 2010.

This page last edited by [RichardDavis](#) on 19/5/2010

#### Discussion Forum Info

[Which Markup Sh...](#) (6 Replies) (24 Views) (Thu 27th)

**Navigation**

- Transcription Desk
- Community portal
- Latest News
- Current events
- Discussion Forum
- Recent changes
- Random page
- Help

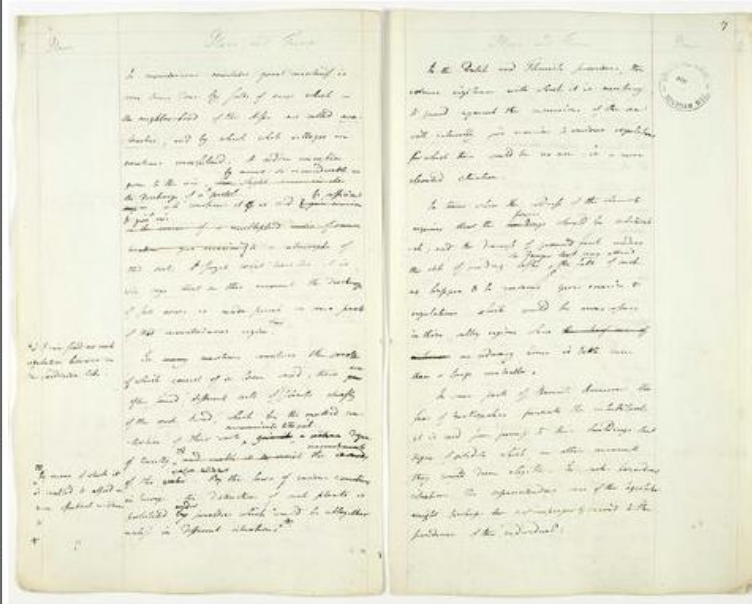
**Search**



Page

# JB/088/009/002

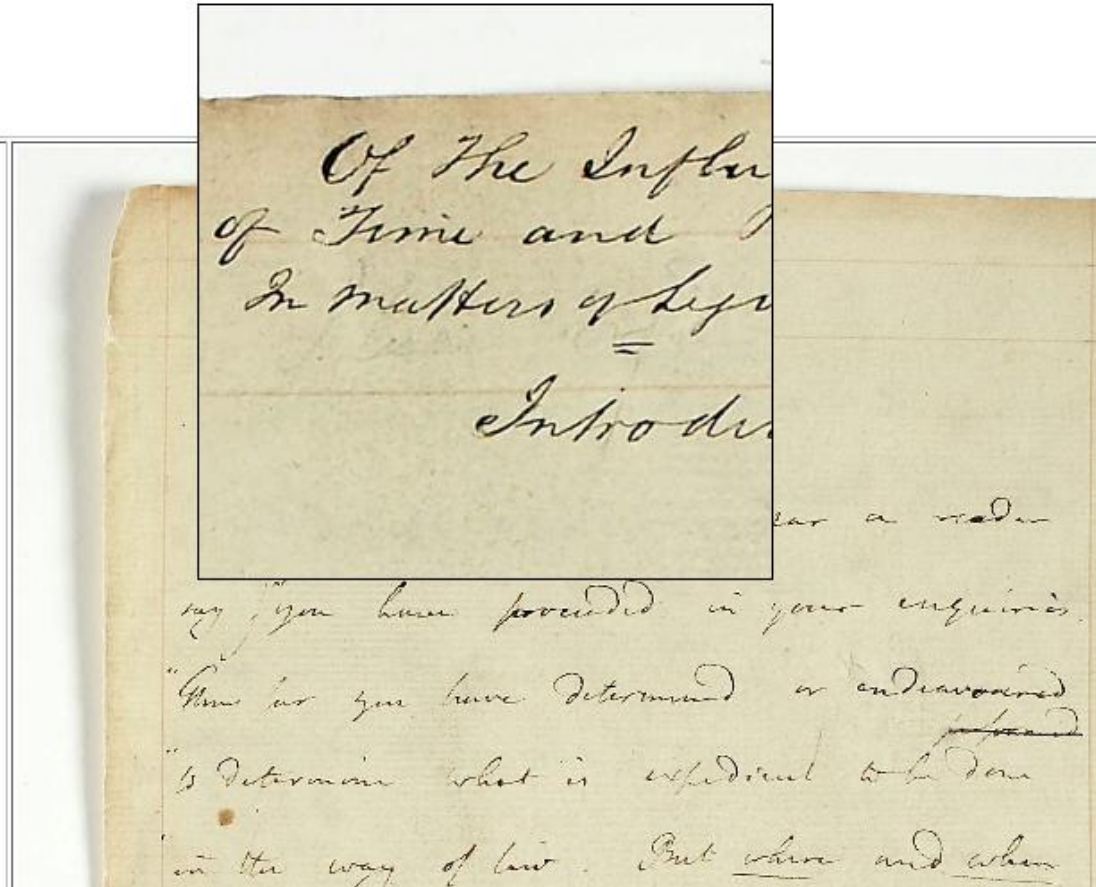
Identifier	{{{identifier}}} ⚠
Date	{{{date}}} ⚠
Box	088
Folio	009
Image	002
Category	Text sheet
Recto/Verso	Recto ⚠
Penner	{{{author}}}
Marginals	{{{author_marginals}}}
Corrections	{{{author_corrections}}}
Main Headings	{{{main_heading}}}
Main Heading Info	{{{main_heading_info}}}
Marginal Summary	
Numbering	
Titles	
Number of Pages	{{{num_pages}}} ⚠
Page Numbering	F27/F9 / F10 / F29/F11 / F12



# A Zoom View Candidate

Some notes at the bottom

```
Of the Influence<lb/>
of Time and Place<lb/>
in matters of Legislation<lb/>
=<lb/>
Introduction<lb/>
"Thus far then I think I hear a reader<lb/> say you
have
proceeded [?] in your enquiries<lb/>
"Thus far you have determined or endeavoured<lb/>
"to determine what is[?]<lb/>
in the way of law. But <u>where</> and
<u>when</><lb/>
```



**Navigation**

- [Transcription Desk](#)
- [Community portal](#)
- [Latest News](#)
- [Current events](#)
- [Discussion Forum](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

**Search**




**Toolbox**

- [Upload file](#)
- [Special pages](#)

## Transcribe Bentham Transcription Desk Forum

Hope you enjoy your stay




[Admin Control Panel](#) [MemCP](#) [Missed Posts](#) [Search](#) [Credits](#)

	Topics	Replies	Last Post
*   <a href="#">Project administration</a>			
*   <a href="#">Discussing Markup: TEI, HTML and Wiki Markup</a>	1	6	<a href="#">Which Markup Should We Re...</a> Thu 27th 2:20 pm RichardDavis
*   <a href="#">Discussion Forum Discussions</a> Discussions about the configuration or use of the discussion forum	1	4	<a href="#">Test Thread</a> Tue 13th 10:31 am MelissaTerras
*   <a href="#">General Project Discussions</a> General discussions about the project management that don't have a dedicated forum	2	4	<a href="#">Browsers &amp; transcription ...</a> Tue 18th 4:18 pm RichardDavis

## Issues and Outcomes

- Worst Case Scenario?
- Best Case Scenario?
- Is this task suitable to crowd sourcing?
  - Complex
- How can we gauge success?
  - Monitor and log user interaction
  - Report back on initiatives
- How can we reach a user community?

## Conclude

- Latest fad?
- Should provide input into cultural and heritage institutions, research, and projects
- Longer term outcomes
  - Sustainability
- Good to try these things!
- <http://www.ucl.ac.uk/transcribe-bentham/>