

Crowdsourcing and Participatory Archives Known Projects

*An annotated list of archival collections cataloging, indexing and transcription projects conducted by libraries and archives only. Initially compiled from the Wikipedia [List of Crowdsourcing Projects](#) then expanded as additional projects are discovered. **Compiled by Robert P. Spindler, January-May, 2014.***

[Africa Through A Lens](#): UK National Archives project to improve knowledge of colonial period Africa photographs. The CO 1069 series is a diverse collection of images with a rich variety of content. In many instances we know little about the people or contents of the photographs and this is one of the reasons why we have published them online and asked people to comment and share their knowledge.

[Ancestry.com World Archives Project](#): With a few keystrokes, you could bring a family together. You can play an important role in helping people find their family history. And all it takes is a computer, some basic software we provide and a little of your time. The Ancestry World Archives Project is thousands of volunteers from around the world with a passion for genealogy and a desire to help others discover their roots. Details and people long ago forgotten are found and stories are preserved for generations to come. And you can be a part of it in no time.

[Ancient Lives](#): The collective of academics and “citizen scientists” at the University of Oxford is working to identify and transcribe papyri belonging to the Egypt Exploration Society and several other institutions.

[The archIve](#): The archIve is the [National Archives of Australia’s] pilot and experimentation into the potential of large scale transcription crowdsourcing to improve access to records. Staff have looked closely at other crowdsourcing sites on offer and attempted to build on their knowledge and techniques, to provide a site that could be used as a large scale platform for a variety of transcription crowdsourcing projects... The project that the archives is undertaking is to digitise consignment lists [accession records] and then make them available for transcription by the public. Once transcribed they become searchable and the items within them can be found more easily...At present the site offers just over 800 lists for the public to transcribe. Some of these are typed and some handwritten.

[California Digital Newspaper Collection](#) In August 2011 the California Digital Newspaper Collection implemented crowdsourced OCR text correction of its digitized historical newspapers, some published as early as 1846 (California statehood 1850). CDNC is a project of the Center for Bibliographical Studies and Research (CBSR) at the University of California, Riverside. The CDNC is supported in part by the U.S. Institute of Museum and Library Services under the provisions of the Library Services and Technology Act, administered in California by the State Librarian.

[Citizen Archivist Dashboard](#) leads to an ever-changing series of projects developed by the US National Archives and Records Administration, beginning with users transcribing 1940 Census data. Several simultaneous projects are represented.

[Civil War Diaries & Letters Transcription Project](#): Help improve access to handwritten documents held by The University of Iowa Libraries.

[Cultura](#), specifically [IPSA or Imaginum Patavinae Scientiae Archivum](#) uses CULTURA's innovative adaptive services and interactive environment to offer genuine user empowerment and unprecedented levels of engagement with digital cultural heritage collections and communities.

[Dickens Journals Online](#) DJO's aim is to launch, by the time of the Charles Dickens bicentenary in February 2012, a [complete](#) online edition of Dickens's weekly [magazines](#), *Household Words* and *All the Year Round*. Readers have access to high-quality facsimile downloads of each weekly number, which show a fully searchable transcript of the text of each page, which has been patiently corrected by means of the [OTC project](#), and with the help of our volunteers.

[DIY History](#) is a project from the University of Iowa that calls upon the public to help transcribe historical documents. It began as a project to help transcribe American Civil War diaries and grew to include every type of historical document, most notably more than 300 years of handwritten cookbooks from the Chef Louis Szathmary Culinary Collection.

[FamilySearch Indexing](#) is a volunteer project which aims to create searchable digital indexes for scanned images of historical documents. The documents are drawn primarily from a collection of 2.4 million microfilms made of historical documents from 110 countries and principalities. Volunteers install free software on their home computers, download images from the site, type the data they read from the image into the software, and submit their work back to the site. The data is eventually made publicly and freely available at [familysearch.org](#) (the world's largest nonprofit genealogical organization) for use in genealogical research. Over one billion historical records have been transcribed to date.

[The Field Book Project](#): Smithsonian Institution. Participation in the [Smithsonian's Transcription Center](#) has also fostered excitement about these materials and created new opportunities for using field books as the content becomes increasingly text searchable. As more field books are transcribed, information within them can be utilized in new ways: reconstructing historical ecologies, clarifying specimen's provenance, and re-discovering localities.

[Frankenstein/Shelley-Godwin Archive](#): The Shelley-Godwin Archive will provide the digitized manuscripts of Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin, and Mary Wollstonecraft, bringing together online for the first time ever the widely dispersed handwritten legacy of this uniquely gifted family of writers. The result of a partnership between the New York Public Library and the Maryland Institute for Technology in the Humanities, in cooperation with Oxford's Bodleian Library, the S-GA also includes key contributions from the Huntington Library, the British Library, and the Houghton Library. In total, these partner libraries contain over 90% of all known relevant manuscripts. [This project currently is not crowdsourcing since selected institutional transcribers were used in the rollout. They anticipate crowdsourcing when the technology is built.]

[FromThePage](#) is free software that allows volunteers to transcribe handwritten documents online. It is used by the San Diego Natural History Museum to transcribe the Laurence M. Klauber Field Notes, by Southwestern University to transcribe the Mexican War diary of Zenas Matthews, and others. The platform integrates with the Archive.org CMS, and the transcriptions can include semantic mark-up for

indexing and annotation. Users can either host the platform on their own servers by getting the source code from GitHub, or have hosting provided by the FromThePage organization.

George Eastman House/Clickworker: The project involves the photo-tagging and cataloging of more than 400,000 images from George Eastman House International Museum of Photography and Film.....To meet the needs of this expansive project, Clickworker leverages its global crowd of more than 115,000 to efficiently tag and catalog the museum's vast collection of images from around the world. [Looks like results would be available in the Online Collections [here](#).]

Historic Cambridge Newspaper Collection. In March 2011, the Cambridge Public Library in Cambridge, Massachusetts launched a digital collection of historic newspapers that implements crowdsourced OCR text correction. The freely accessible and keyword searchable database contains newspapers dating back to 1846 when Cambridge was established as a city. The Historic Cambridge Newspaper Collection is a project of the Cambridge Room, the Cambridge Public Library's Archives and Special Collections, and is supported by funding from the Community Preservation Act.

Koorie Annotation System: A Monash University project with the Koorie People of Victoria concluded that the community wanted to be able 'to add their own stories and versions of other stories to records held in public archives and other institutions'. The 'Koorie Annotation System' now enables the People to 'comment upon the inaccuracies or limitations of institutional records, to contribute family narratives which expand upon or give context to institutional records and to present their version of events alongside the official one'.

Liljenquist Archive: In 2010 the Library of Congress received, as a donation, the Liljenquist family's collection of photographs of the American Civil War. With most of the people and photographers unidentified, the Library posted the images on Flickr and sought the public's help in providing information: "Please let us know if you recognize a face from your family, a regiment, or a photographer's painted studio backdrop!"

Map Warper: New York Public Library's project to improve description and geospatial referencing for their city maps collection. Essentially, a digital map interface with a "go back in time" function, allowing you to compare changes in the urban landscape over the years... The Map Warper is used heavily by Map Division staff, and also by members of the public, who are invited to create an account and learn to work with our maps through a video tutorial.

Metadata Games: A suite of games developed at Dartmouth College that can quickly gather valuable tags while offering fun for players. The games are an opportunity for the public to interact with cultural heritage institutions in ways they may not have otherwise. Cultural heritage institutions with digital collections that have little or no metadata will likely benefit the most from using Metadata Games. Our hope is that with Metadata Games, cultural heritage institutions will gain useful data for their collections, assist scholars to analyze their collections in novel and possibly unexpected ways and increase engagement with the community at large.

Old Weather is a web-based effort to transcribe weather observations made by Royal Navy ships around the time of World War I. These transcriptions will contribute to climate model projections and improve a database of weather extremes and will be of use to Historians in tracking past ship movements and the stories of the people on board. [Uses Zooniverse technology.]

Picture Australia, National Library of Australia: Picture Australia harvests digital images from over 50 cultural heritage institutions both in Australia and overseas. Since January 2006 the public have been encouraged to add their own digital images to Picture Australia via a relationship set up with Flickr. Any of the contributing institutions can then request permission from the public contributors to upload the contributed pictures into their own collections, while the same pictures also remain visible in the central Picture Australia pool. [looks like collection development, not cataloging]

Project Gutenberg: The purpose of Project Gutenberg Literary Archive Foundation is to encourage the voluntary creation and distribution of electronic books. Proofreading and error correction for OCRed books.

ReCaptcha Project: Places words that were OCRed (optical character recognition) incorrectly or were unable to be OCRed, in 'captcha' windows on thousands of websites. When a sufficient number of users have transcribed the word, the system gathers that data and compares it in order to statistically determine the correct transcription (Parkes 2013, 66–69).

Remember Me: US Holocaust Memorial Museum project seeks to identify what happened to child holocaust survivors by presenting photos of children from relief agencies.

<http://rememberme.ushmm.org/pages/about-the-project>

San Francisco Participatory Archives Group [Blog], re: Prelinger motion pictures. Last blog post in 2012.

Shakespeare Quartos Archive: A collaboration of the Bodleian Library of Oxford University, University of Maryland College Park and the Folger Shakespeare Library. Its aim is to include high-resolution facsimiles of every quarto edition of Shakespeare's plays printed before 1642...Joined to this archive is a new set of tools that allows side-by-side image comparison, text overlays, exhibits, tagging, and user annotation.

Smithsonian Institution Projects: The [National Postal Museum](#) is using 160 vetted volunteer researchers to document its artifacts. At the National Museum of American History, the [Access American Stories](#) app invites visitors to describe 180 of America's most evocative historical objects, helping the visually impaired to "see" the collections.

Transcribe Bentham is a crowdsourced manuscript transcription project launched in 2010. It is run by University College London's Bentham Project, in partnership with UCL's Centre for Digital Humanities, UCL Library Services, UCL Learning and Media Services, and the University of London Computer Centre. Transcribe Bentham will digitise 12,500 Bentham folios, and, through a wiki-based interface, allow volunteer transcribers to take temporary ownership of manuscript images and to create TEI-encoded transcription text for final approval by UCL experts. Approved transcripts will be stored and preserved, with the manuscript images, in UCL's public Digital Collections repository. The project makes available,

via a specially-designed transcription interface, digital images of UCL's Bentham Papers collection — the unpublished writings of the philosopher Jeremy Bentham, which run to some 60,000 manuscript folios — which volunteers are encouraged to transcribe. The transcripts are intended to contribute to the Bentham Project's production of the new edition of *The Collected Works of Jeremy Bentham*, and will be uploaded to UCL's digital Bentham Papers repository, widening access to the collection. Manuscripts can be viewed and transcribed by signing-up for a transcriber account at the Transcription Desk, via the Transcribe Bentham website. Media coverage has included a feature article in *The New York Times*, and a broadcast on Deutsche Welle radio. The project was shortlisted for the 2011 Digital Heritage Award, and received an Award of Distinction in the Digital Communities category of the 2011 Prix Ars Electronica. The open-source code for the *Transcribe Bentham* transcription tool is available for reuse and customisation.

Trove a digital newspaper project that's run by the National Library of Australia. Since that project launched in 2009, some 30,000 volunteers have corrected 40 million lines of scanned and digitized text. This is the first library project in the world that has undertaken crowdsourcing on a large scale. The leader of the project Rose Holley is an advocate of using crowdsourcing to help libraries and archives expose and improve digital resources and has written articles about the Australian newspaper achievements^[11] and tips for libraries on how to crowdsource effectively.^[12]

UK MPs Expenses Scandal, The Guardian: Tagging subject content of records/archives.

In June 2009 there was controversy over UK MP expenses. This created huge public interest and outrage and the result was that 2 million documents relating to MP expenses were made publicly available as PDFs. The Guardian newspaper built a website in a matter of days that enabled the public to view, read and mark documents they felt should be exposed and investigated. An article¹² was published by the Guardian outlining the lessons learnt about crowdsourcing.

War Office Archive: A project of the Center for History and New Media at George Mason University, this collection [unites](#) copies of the lost files in a [digital archive](#) that reconstitutes this invaluable historical resource. Call for public transcription is present on the website.

What's On the Menu? New York Public Library We're transcribing our historical restaurant [menus](#), dish by dish, so that they can be searched by what people were eating back in the day. It's a [big job](#) so we need your help!