Crowdsourcing transcription at the Library of Congress, and how to engage virtual volunteers with your collections

Dr Victoria Van Hyning

By the People Community Manager
crowd@loc.gov; @Crowd_LOC; @VanHyningV
In this talk...

• Crowdsourcing—an overview
• Challenges, and solutions to better describing data at Library of Congress
• Library of Congress Flickr Project
• By the People transcription and tagging project
• How to get involved in crowdsourcing
Crowdsourcing for cultural heritage

“Cultural heritage crowdsourcing projects ask the public to undertake tasks that cannot be done automatically, in an environment where the activities, goals (or both) provide inherent rewards for participation, and where their participation contributes to a shared, significant goal or research interest.”

Crowdsourcing our Cultural Heritage, ed. Mia Ridge (Ashgate 2014), p.2
I'm a collaborator!
How can you get involved?

• Help a researcher find or create a crowdsourcing project
• Help an educator use crowdsourcing in their classroom
• Help someone use a crowdsourcing project to fulfill volunteer or community service requirements
• Host a transcribathon or a reviewathon at your institution
• Create a crowdsourcing project
• Volunteer on your own, with friends or a book group etc.
• Visit Wikipedia for a list of projects
Engage, inspire, and inform Congress and the American people with a universal and enduring source of knowledge and creativity.
Library of Congress Flickr Project

Celebrating its 12th year today (January 16, 2020)

Invites people to add information about images in LOC collections. These details are vetted and then added to https://loc.gov.

 Featherlite Pancakes

1 C flour
2 T B. Powder
5 t salt
2 T sugar
mip
1 egg - 1/4 C milk
1/3 C peanut butter
1 T shortening or oil
Combine with dry ingredients.
Cook at 275° on griddle.
By the People  [https://crowd.loc.gov/](https://crowd.loc.gov/)

Transcription and tagging of digitized collections (mostly handwritten but some typed) to make them:

- Keyword searchable
- Readable by individuals and accessibility technologies
- Available for research, such as large scale textual analysis
It’s demo time!

https://crowd.loc.gov/
New challenges for 2020: Spanish Legal Documents

- We hope to reach new audiences in the US and Spanish speaking countries.
- A collaboration with Hispanic division.
- Transcription of ~63,000 pages.
- No translation of these pages, but an opportunity for LOC to translate BTP instructions.
On the night of the 17th August a tornado passed over this place, laid the fences flat, rooted up trees blew down corn & done other damage. The next morning by daylight as I was putting up my fence 2 great wolves walked along unconcerned within 30 yds of me - I tried to scare them by taking off my hat & running towards them but they would not quicken their gait - there are the only ones I have seen - High winds seem to be common everywhere this year, so the wolves say - There has been a flock of wild turkeys around my house since I cut my oats and as I had not shot at them but twice. They were pretty tame. Seeing them last week I took out my gun which had a great charge in it that I intended to shoot a Bull with who had badly hooked Clark's horse / and as their heads all go up together when they see any one, I let drive & shot fine, but could not catch but 3. Had a dog or another person with me I would have [snared?] all - Prairie hens are very plenty this fall - they are all the color of a "gray duck" & their feathers lay closer to them than to barnyard fowls, with a trap a person could catch as many as they could eat during the winter. Partridges are too small game to shoot at - if you were only here we could live like Princes.

I do not know exactly where I shall go yet - but if no better chance [ce] offered

Transcribed and reviewed by volunteers participating in the By The People project at crowd.loc.gov
By the People by the numbers

Since October 24, 2018:

- 12,227 registered users
- 53,031 pages transcribed and awaiting review
- 44,668 pages completed
- 8,018 transcriptions published on loc.gov
Connecting in person and online

Be in touch @Crowd_LOC

Henry M. Rosenberg @DoctorHenryCT · 13 Nov 2018
It’s very rewarding to know I am contributing to preserving history.

Beth Graham @zagramham · 13 Nov 2018
I feel exactly the same way. A bonus is that I’m personally learning so much about U.S. history.

Christian Mobley @Mo... · 10 Dec 2018
Replying to @Crowd_LOC
I have been transcribing today to help keep my mind off of the chronic pain I am having. I enjoy doing the long letters and difficult ones.

Kristina Killgrove @DrKillgrove · 13 Nov 2018
My son is ridiculously excited to help the @librarycongress transcribe records from the Civil War. She’s learning about the “Left Handed Penmanship” contest for veterans who lost use of their right hands.

Check it out at crowd.loc.gov!

Show this thread

Kristina Killgrove @DrKillgrove · 10 Dec 2018
Replying to @Crowd_LOC
I’ve also been doing the Lefthanded Penmanship Contest transcriptions, since I’m both left-handed and interested in disease and disability in the past. Your project is great, and now I’m going to see what other citizen-STEAM projects I can get my girls involved in.
Campaign #11: The Man Who Recorded the World: On the Road with Alan Lomax

- Led by American Folklife Center
- Very different content from the previous Campaigns
- Launched September 4, 2019
- Featured all of January 2020 for a special “Challenge”—to move the number of completed pages up to 3000
External transcribathons and Challenges are a big help!

Mrs. Flaherty @RooseveltLib  Mar 20
The Black Experience class with @ms_nmb is transcribing some primary sources of Mary Church Terrell for the @Crowd_LOC to celebrate Women’s History and make learning history accessible for all! #KentRiderPride
In the classroom

Educators can use the project to help students build skills in:

• Close reading and examining historical context
• Building interpretive skills and primary source analysis
• Learning to read handwriting
• Typing, and navigating digital resources (digital literacy)
Transcribathons and reviewathons

https://crowd.loc.gov/resources/

Dr. Wendy A. Matlock @wam6000 · Nov 20
At tonight’s @englishkstate Literature Track Transcribe-a-thon, participants discovered the joys and difficulties of transcribing old handwriting and contributed to the Library of Congresses digitization work.
Concordia code: “With one heart”

• A reusable open source codebase for text transcription projects
  https://github.com/LibraryOfCongress/Concordia

• Read more: “With One Heart”: Agile approaches for developing Concordia and crowdsourcing at the Library of Congress”
  https://journal.code4lib.org/articles/14901
Thank you!

History Hub Discussion Forum: historyhub.history.gov/community/crowd-loc

Email: Crowd@loc.gov

Twitter: @Crowd_LOC

Github: github.com/LibraryOfCongress/Concordia

Newsletter: sign up at bottom of any By the People page

Be in touch!
Back-up slides if demo fails
Crowd.loc.gov

By the People
Volunteer to uncover our shared history and make documents more searchable for everyone.

LET’S GO!

Anyone Can Contribute

- TRANSCRIBE
  Type what you see on the page.

- REVIEW
  Register to edit and complete transcriptions.

- TAG
  Register to add tags and share what you find.

Anonymous users
Registered users
Suffrage: Women Fight for the Vote

The struggle for women's voting rights—considered the largest reform movement in American history—lasted more than seven decades, from 1848 to 1920. Determined women organized, lectured, petitioned, lobbied, paraded, picketed, and went to jail for daring to support suffrage. Explore the papers of leading suffragists such as Susan B. Anthony, Elizabeth Cady Stanton, Mary Church Terrell, Carrie Chapman Catt, and Anna E. Dickinson.

View Projects

Civil War

The Library of Congress holds one of the largest collections of Civil War materials in the world. This topic draws together materials from a range of Campaigns that reveal the impact of this bloody period in American history.

View Projects

The Man Who Recorded the World: On the Road with Alan Lomax
Campaign #11: The Man Who Recorded the World: On the Road with Alan Lomax

- Led by American Folklife Center (first non-MSS)
- Launched September 4, 2019
- Featured all of January 2020 for a special “Challenge”—to move the number of completed pages up to 3000
Bahamian Material
Wassary Dec. '39
E. & A. Lomax
How to transcribe

Instructions

Welcome to By the People

How to transcribe

How to review

How to tag

We transcribe to improve search functionality. Our goal is to make documents word-searchable in the Library of Congress website, which means typing transcriptions that can be read by that computer system, as well as by humans. Most handwriting and some typed text cannot be automatically and accurately translated into machine-readable text using current technologies -- that's why we need your help!

We ask you to transcribe a document roughly as it appears on the page. Preserve line breaks, except in cases where words are broken over two lines. Our main goal is to capture all of the words on these pages, so broken words are not helpful for search. Using line breaks to roughly capture the layout of the page helps reviewers (other volunteers) check transcriptions. These will be viewable beside the original images in the website, so anyone who is interested in the physical layout of the original document will be able to see it.

All in order:
Transcribe text in the order it appears.

Spelling:
Preserve original spelling unless the author seems to have made a minor error, such as writing “teh” instead of “the”. If a misspelling will impact the searchability of the document, use a tag to add the correct spelling. Example:

- An author wrote “Willa Kather” instead of “Willa Cather”. Transcribe Willa Kather, and tag “Willa Cather”.

Insertions:
Review

- Review can be a great place to start!
  - Build confidence
  - Get eyes comfortable with the handwriting
  - Sometimes reading what someone else thinks a page says can really help people gain confidence.
- If you’re hosting an event, review is a good place to start to get people up to speed with how the interface works. Work as a group using a large screen or projector to try to parse a few lines on a page.
Campaign #12 and a Challenge: Civil War POW diary

5 Day Challenge over Veterans Day

200 images fully transcribed and reviewed in 36 hours

Stretch challenge: review documents in another Civil War Campaign: 371 completed.