

Impact of crowdsourced transcriptions from the Library of Congress

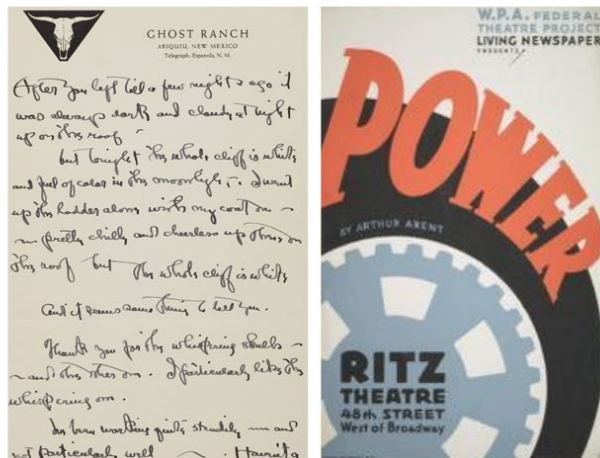
Lauren Algee

Senior Digital Collections Specialist
Digital Content Management Section

Abby Shelton

Digital Collections Specialist
Digital Content Management Section

SAA Research Forum, July 19, 2023



LIBRARY
LIBRARY
OF CONGRESS

BY THE PEOPLE

Dual goals: Collection enhancement & user engagement

Collection enhancement

- **Discovery**
Enabling page-level keyword search
- **Accessibility**
Easier to read for all people, as well as accessibility technologies
- **Datafication**
Opening computational research

User engagement

Close reading, primary source analysis, & forging personal connections to history & the Library



BergenCC
@BergenCC

In celebration of Women's History Month, the Sidney Silverman Library @bergencc sponsored a Transcribe-a-thon of the #ClaraBarton papers where volunteers transcribed about 40 documents on March 30, 2022 as part of the By the People project @librarycongress @Crowd_LOC



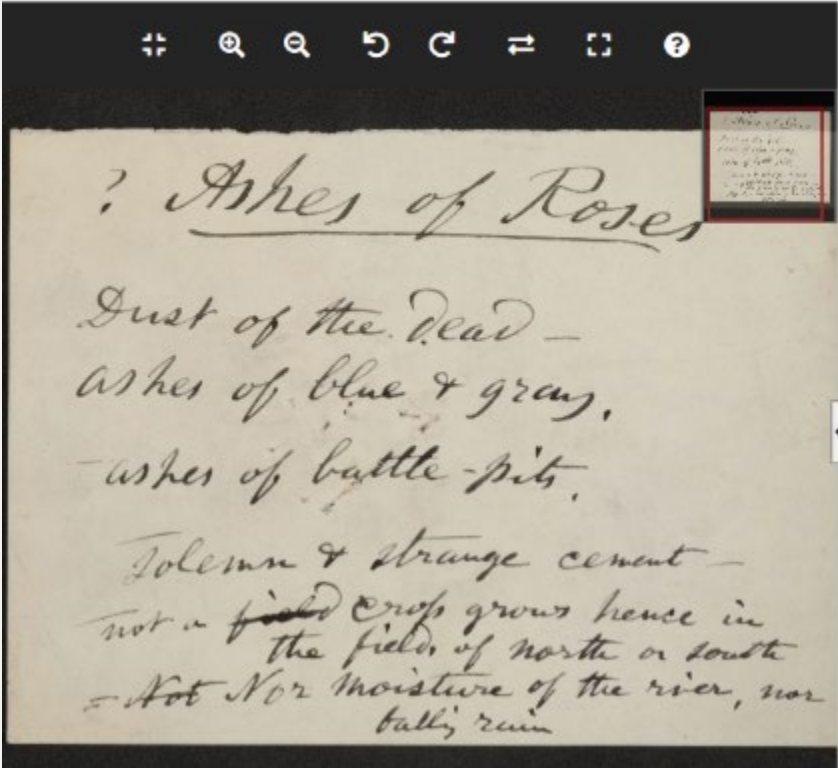
10:15 AM · Mar 31, 2022 · Hootsuite Inc.

Volunteers transcribe and review at crowd.loc.gov

Page 4 Go < > Fullscreen

View on www.loc.gov

Find a new page →



A snippet of a handwritten manuscript on aged paper. The text is written in cursive and includes the title "? Ashes of Roses" and several lines of descriptive text. A small thumbnail of the full page is visible in the top right corner of the snippet.

? Ashes of Roses

Dust of the dead -
ashes of blue & gray,
- ashes of battle-pits.

Solemn & strange cement -
not a ~~field~~ crop grows hence in
the fields of north or south
- Not Nor moisture of the river, nor
bally ruin

✓ Completed
This transcription is finished! You can read and add tags.

? Ashes of Roses
Dust of the dead -
ashes of blue & gray.
- ashes of battle-pits.
solemn & strange cement —
not a [field] crop grows hence in
the field, of north or south
= [Not] Nor moisture of the river, nor
falling rain

We publish completed transcriptions on loc.gov

Featherlite Pancakes

sift together

1 C flour

2 T B. Powder

1/2 t salt

2 T sugar

Mix

1 egg

1 1/4 C Milk

1/3 C peanut butter
melted

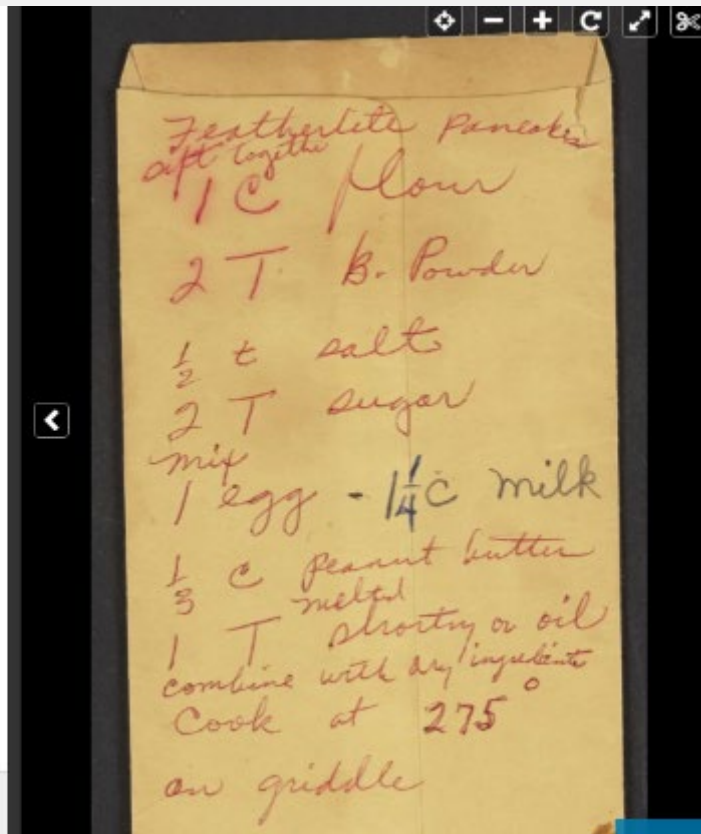
1 T shorting or oil

Combine with dry ingredients

Cook at 275°
on griddle

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

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



And package as datasets

Refine your results

- Available Online 7
- All Items 7
- Original Format 7
- Software, E-Resource 7
- Online Format 7
- Compressed Data 7
- Date 7
- 2000 to 2000 7
- Location 7
- United States 7
- Part of 7
- Library of Congress Online Catalog 7
- General Collections 7
- Selected Datasets 7
- U.S. Government Digital Resources 7
- Contributor 7
- By the People (Program) 1
- Anthony, Susan E. (Susan Browne) 1
- Bourne, Wm. Gland (William Gland) 1
- Catt, Carrie Chapman 1
- Gibson, Sarah (s) 1
- National American Woman Suffrage Association Collection (Library of Congress) 1
- Park, Rosa 1
- Stanton, Elizabeth Cady 1
- Subject 7
- Data Sets 7
- United States 7
- History 4
- 19th Century 3
- African Americans 2
- Civil Rights 2
- Suffrage 2



Software or E-Resource

SOFTWARE, E-RESOURCE

Dataset from Rosa Parks Papers

Rosa Parks:

This dataset is an export from the Library of Congress By the People crowdsourcing program (<https://crowd.loc.gov/>). It contains a combination of digital collections metadata, software-created text generated through a transcription and review process, and metadata representing the arrangement of the items in the By the People platform, as defined by the Concordia (<https://github.com/LibraryOfCongress/concordia>) instance on which By the People operates. Concordia is an open...

Contributor: Parks, Rosa - By the People (Program)

Date: 2021



Software or E-Resource

SOFTWARE, E-RESOURCE

Dataset from Carrie Chapman Catt Papers

This dataset is an export from the Library of Congress By the People crowdsourcing program (<https://crowd.loc.gov/>). It contains a combination of digital collections metadata, software-created text generated through a transcription and review process, and metadata representing the arrangement of the items in the By the People platform, as defined by the Concordia (<https://github.com/LibraryOfCongress/concordia>) instance on which By the People operates. Concordia is an open...

Contributor: National American Woman Suffrage Association Collection (Library of Congress) - Catt, Carrie Chapman - By the People (Program)

Date: 2020



Software or E-Resource

SOFTWARE, E-RESOURCE

Dataset from Wm. Gland Bourne Papers

Civ. War Soldiers: 'Disappointed but not dishonored'

This dataset is an export from the Library of Congress By the People crowdsourcing program (<https://crowd.loc.gov/>). It contains a combination of digital collections metadata, software-created text generated through a transcription and review process, and metadata representing the arrangement of the items in the By the People platform, as defined by the Concordia (<https://github.com/LibraryOfCongress/concordia>) instance on which By the People operates. Concordia is an open...

Contributor: By the People (Program) - Bourne, Wm. Gland (William Gland)

Date: 2020



Software or E-Resource

COLLECTION

Datasets from Branch Rickey Papers

Branch Rickey:

The data available here is from Branch Rickey, including reports, part of legendary baseball coach Rickey's Branch Rickey Papers available in the archival papers in a current scanning nearly 80 years. Rickey was a major league executive and outfielder. He is perhaps best known as the man responsible for bringing Jackie Robinson into major league baseball in 1947, breaking baseball's long-embedded color barrier. The Branch...

Contributor: By the People (Program)

Date: 2019

www.loc.gov/collections/selected-datasets/?fa=contributor:by+the+people+%28program%29

Why

- Valuing volunteers
- Library mission
- Iterative improvement

What

- Discoverability
- Accessibility
- Engagement

How

- Accessibility evaluation
- Web analytics
- Volunteer survey
- Reference data
- Researcher interviews

Researcher impact: Accessibility

Sept. 2022 screen reader assessment by LC Accessibility Team

“In general the content are accessible but would benefit from some changes to improve the experience of screen reader users.”

Result: Ticketed improvements that will impact loc.gov experience for all online text

The screenshot displays a digital interface for a manuscript collection. At the top, there is a button labeled "Cite This Item". Below it, a navigation menu includes a home icon and a search icon. A prominent heading reads "More Manuscripts/Mixed Material like this". The main content area features four vertical panels, each representing a manuscript item. Each panel contains a thumbnail image of a handwritten document, a set of three accessibility icons (a green square with a white 'h2', a purple square with a white right-pointing arrow, and a green square with a white document icon), and a text description. The descriptions are as follows:

- Panel 1: "Abraham Lincoln papers: Series 1. General Correspondence. 1833-1916: Edward Bates to Abraham Lincoln, Friday, April 05, ..." with a small orange icon of a person at a computer.
- Panel 2: "Abraham Lincoln papers: Series 1. General Correspondence. 1833-1916: Horace Binney to Abraham Lincoln, Friday, April 05, ..." with a small orange icon of a person at a computer.
- Panel 3: "Abraham Lincoln papers: Series 1. General Correspondence. 1833-1916: Montgomery Blair to William Grandin, Friday, April 05, ..." with a small orange icon of a person at a computer.
- Panel 4: "Abraham Lincoln papers: Series 1. General Correspondence. 1833-1916: Patrick Crowley to Abraham Lincoln, Friday, April 05, ..." with a small orange icon of a person at a computer.

Researcher impact: Discoverability

Clara Barton Papers

- 2x the visits and views on pages with transcriptions

Branch Rickey Papers

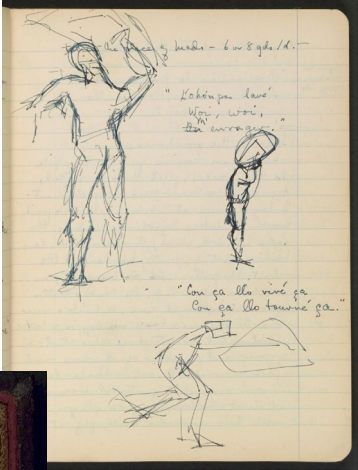
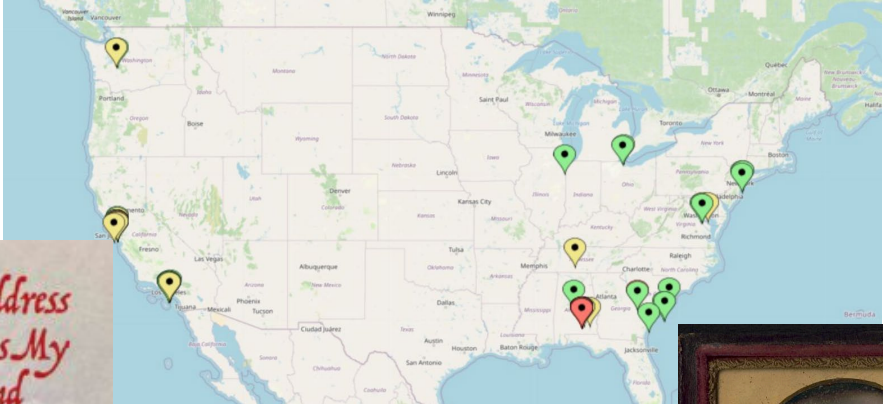
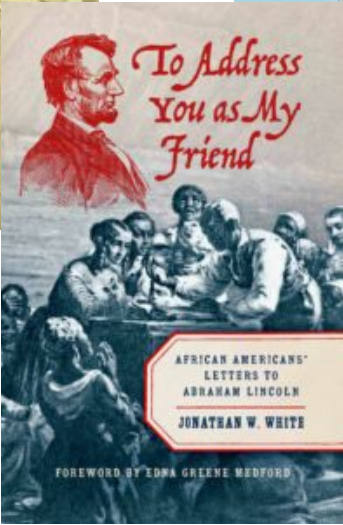
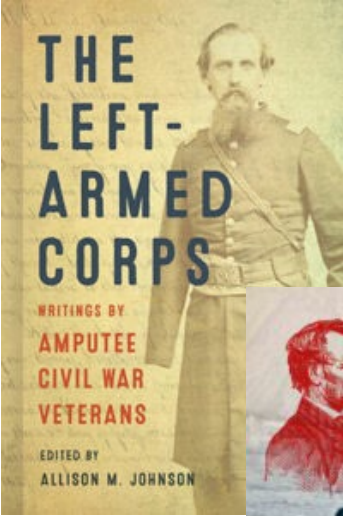
- 37% of successful search terms in transcriptions alone

Carrie Chapman Catt Papers

- 44% of successful search terms in transcriptions alone

Term	Visits	In Metadata	In Transcriptions
"fred rogers"	1		yes
"oklahoma a&m"	1		yes
"paul martin"	1		yes
"paul pettit"	1		yes
"roger bannister"	1		yes
"steve carlton"	1		yes
"transcribed and reviewed by volunteers p	1		yes
"branch rickey" "hank aaron"	1	yes	yes
"branch rickey" "willie mays"	1	yes	yes
1940 girls baseball league	1		yes
1944 world series baseball	1	yes	yes
1947 brooklyn dodgers	1	yes	yes
1971 pirates	1		yes
1st pan american games, nicaraguan baseba	1		yes
albert hall	1	yes	yes
albert robinson	1	yes	yes
all american girls baseball league	4		yes
all-american girls professional baseball lea	1		yes
american football	1		yes
arizona	1		yes
arkansas	1		yes
art red swanson	1		yes
athlete	1	yes	yes
august 16 1949 letter	1		yes
babe ruth	1		yes
balcena	2		yes

Researcher impact: Story-telling from data



Volunteer impact

Volunteers said they gained:

- A sense of purpose, satisfaction or accomplishment
- New knowledge of history
- New skill or ability
- A greater sense of well-being
- An enjoyable & relaxing hobby

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.

6:29 PM · Dec 10, 2018 · Twitter for Android

Volunteer-identified motivations:

Making collections accessible

Giving back

Interested in campaign topics

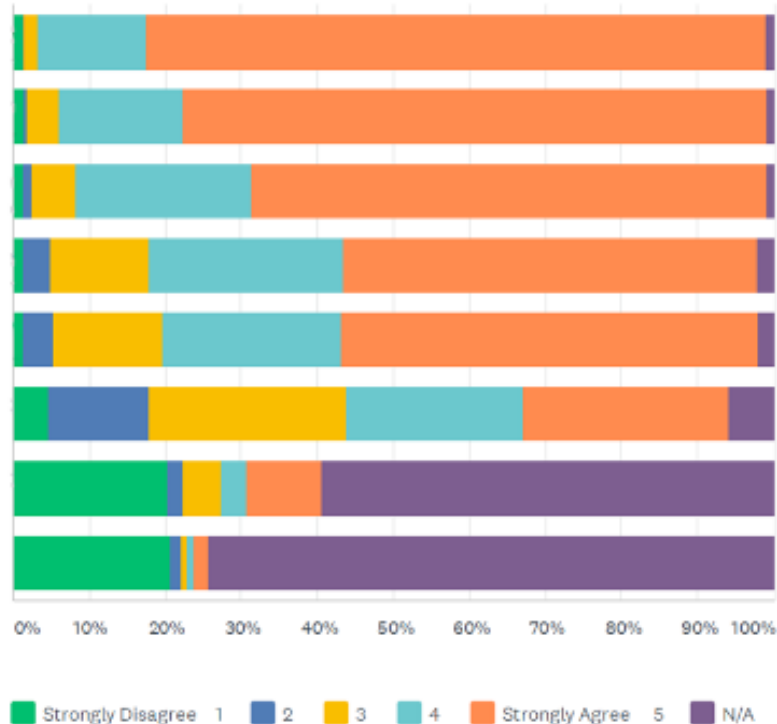
Improve skills

Supportive community

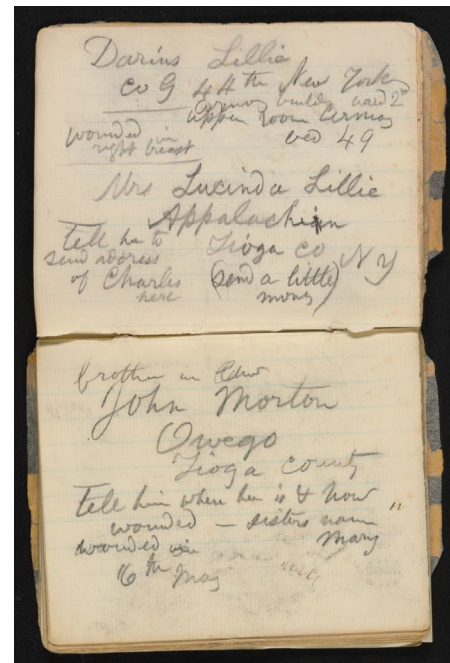
Meeting transcription goals

Service hour requirements


Classroom use



Volunteer impact: Collection connection



Volunteer impact: Community



YOUNGSTOWN
STATE
UNIVERSITY

NEWS CENTER / NEW YSU CLUB TRANSCRIBES NEARLY 7,000 HISTORICAL DOCUMENTS...AND COUNTING

New YSU club transcribes nearly 7,000 historical documents...and counting

MAY 28, 2021


PRINT | SHARE

Julie Centofanti faced a bit of a dilemma.

As a freshman member of Youngstown State University's Honors College, the Biology major was required to complete at least 60 hours of voluntary service, a task made even more daunting in the midst of a pandemic that nearly overnight switched the world into isolation mode.

So, she did what anyone her age would do: Googled.

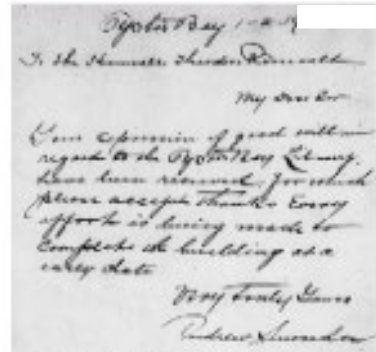
A search for "Ways to volunteer virtually" unveiled a relatively unique venture via the Smithsonian and Library of Congress - transcribing and digitizing millions of pages of hard-to-read historical and scientific documents.



Julie Centofanti

"I know I'm a Biology major, but I love learning about the past," said Centofanti, a graduate of South Range High School. "I found the historical aspect of these documents so interesting, and eventually I couldn't stop typing and reading."

So Centofanti recruited fellow students and organized the YSU Transcribing Club, which over the course of the past academic year logged more than 800 hours of work and transcribed nearly 7,000 historical documents, from letters to Teddy Roosevelt to notes of early 20th-century astronomy experiments.



Oyster Bay 1-2-19
To the Honorable President Roosevelt
My dear Sir
Your expression of good will in regard to the Oyster Bay Library have been received for which please accept thanks. Every effort is being made to complete the building at a early date
Very Truly Yours
Robert Roosevelt

This letter is the first document in January 1900 is a copy of the nearly 7,000 documents that YSU students have transcribed. The letter goes to the Oyster Bay Library on Long Island, N.Y., is in the Library of Congress and was written to Roosevelt when he was governor of New York.

Try it out at **crowd.loc.gov**

Learn more

[blogs.loc.gov/the-signal/category/
by-the-people-transcription-program/](https://blogs.loc.gov/the-signal/category/by-the-people-transcription-program/)

And please keep in touch!

Email

lalg@loc.gov

crowd@loc.gov

Twitter

@Crowd_LOC

Newsletter

<https://go.usa.gov/xsyzc>

Github

[github.com/LibraryOf
Congress/concordia](https://github.com/LibraryOfCongress/concordia)