

**Amigos Fellowship and Opportunity Award
Final Report: June 19, 2015**

Project Title: My Dear Wife: Crowdsourcing Transcriptions of Civil War Manuscripts

Awardee: University of Oklahoma Libraries, Western History Collections, Kristina L. Southwell

Summary: The primary objective of the project was to crowdsource the digital transcription of a Civil War diary and a set of letters from the Western History Collections. This was accomplished by November 3, 2014, via the project website: <https://transcribe.ou.edu> . The project's transcriptions and related files are permanently retained in our institutional repository, ShareOK, for public access: <https://shareok.org/handle/11244/14633> . An additional access point for the paired transcriptions and document images will be created on the University Libraries' new website when all components of the site (currently in beta) are completed in late 2015: <https://libraries.ou.edu> .

Team Members:

Proposal Team:

Karen Rupp-Serrano, Director of Collection Management and Scholarly Communication

Kerry Magruder, Curator, History of Science Collections

Brian Shults, Coordinator of the Digitization Laboratory

Jacquelyn Reese, Librarian, Western History Collections

Implementation Team:

Kristina Southwell, Associate Curator, Western History Collections

Logan Cox, Information Technology Analyst III, Repository Services

Brian Shults, Coordinator of the Digitization Laboratory

Jacquelyn Reese, Librarian, Western History Collections

Twila Camp, Library Web Services Manager

Prudhvi Raj Maraboyena, Graduate Student

Timeline:

5-15-2013: Amigos funding awarded

9-24-2013: Scanning of diary and letters complete

4-25-2014: Beta site with transcription tool complete

8-15-2014: [Transcription website](#) opened to the public

11-3-2014: All documents fully transcribed twice

3-13-2015: Reconciliation of all transcriptions complete

5-21-2015: Transcriptions and documents made available via [ShareOK](#)

Objectives and Results:

The project team accomplished the principal goal of crowdsourcing digital transcription of the diary and letters within the proposed two-year time frame. The transcription website was open for only 81 days, and during that time received 1,596 transcriptions from volunteers (two for each page). After reconciliation of the double transcriptions and removal of a few blank pages, the final transcription total is 787 pages. A total of 152 participants signed up to transcribe. Surprisingly, a single dedicated transcriber completed nearly half of all transcriptions received. Several other transcribers also submitted large numbers of transcriptions, but many signed up who completed no transcriptions. This was most likely because there were no pages left to transcribe toward the end of the project, when public interest in the project was at its peak.

The quality and consistency of transcription varied greatly among participants. Some transcribers clearly showed knowledge of Civil War history by correctly interpreting place names and military terms. Others had difficulty deciphering the cursive handwriting and abbreviations. Overall, the transcribers seemed quite conscientious about their work, and several even requested feedback on the quality of the transcriptions they had submitted. The speed with which the volunteers transcribed the materials indicates there is a great deal of public interest in working with historical documents.

The project resulted in improved access and readability of these special collections materials, which were previously only available in original cursive handwriting. It also increased public engagement with our special collections and helped commemorate the 150th anniversary of the U.S. Civil War. The specific objectives noted in the proposal and how our team accomplished them are as follows:

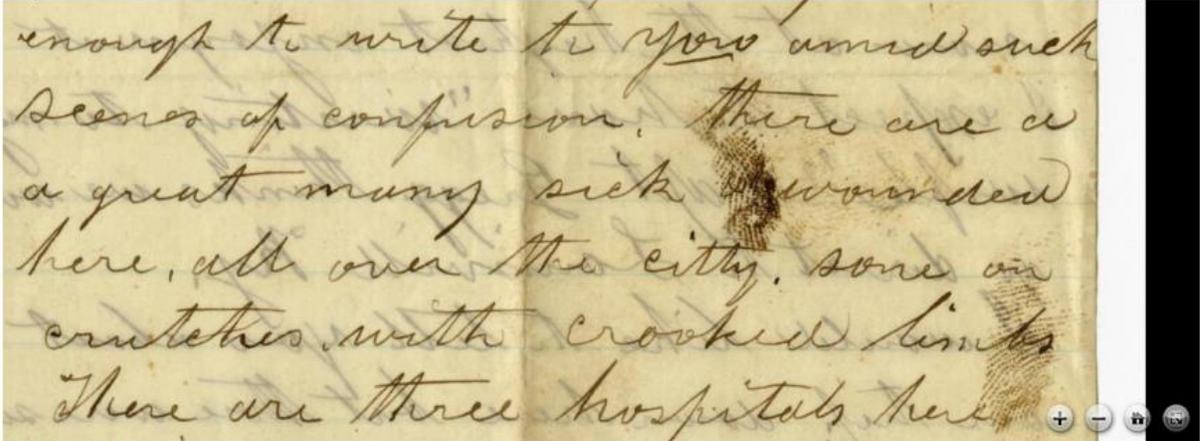
- **Apply Scripto for crowdsourcing.** [Scripto](#) was originally identified by the proposal planners as the tool that would be used for the project. Our implementation team opted not to use it partly because it is based on a wiki-like workflow that produces a single transcription of a document that can be edited and improved by other users of the site. The general workflow planned in our proposal was based on the [Old Weather project](#) in which each transcriber creates an independent transcription, then the transcriptions are compared as a validation step. Following that model, our project did not permit transcribers to edit each other's work and validation was instead performed by project staff. (The Old Weather transcription software, [Scribe](#), was designed for the transcription of data - in their case, ship's logs - so it was not an ideal fit for the letters and diary in our project, either.) In addition, the Scripto module for Drupal used Microsoft's [Zoom.it service](#) for image zooming. That service was deprecated while we were evaluating Scripto, and it was completely terminated on September 15, 2014. Scripto hasn't worked with Drupal or WordPress (our supported CMS platforms) since that time. So neither Scripto nor Scribe suited our project. As a result, our site was based on the Drupal content management system, with a custom Drupal module that managed the creation of transcriptions, and a variety of other Drupal add-ons including the [DeepZoom](#) module which provided similar features to the Zoom.it service. The resulting transcriptions were exported as text files.

Figure 1: The project's Drupal-based custom transcription tool enabled participants to enter their transcription text in the box below the image. The image could be enlarged and moved around to improve visibility.



Create Transcription

Image Status:
In Progress
Image File:



Collection Reference:
My Dear Mary
Transcription Text *

Keokuk Sep 26 1862 My Dear Mary: I am sitting in the shade of a tent trying to write you a few lines. I should like to say a great many things but how can I compose my mind enough to write to you amid such scenes of confusion. there are a great many sick & wounded here, all over the city. some on crutches with crooked limbs. There are three hospitals here I visited one yesterday, it appeared to be kept verry well. the floors were washed & rubbed clean every thing was neat airy & comfortable. We expect to start for St Louis tomorrow or some place south I do no know where There are two regiments here besides the 23 (from Des Moines Valley) They look like stout & rugged fellows, There was one regiment at Oskaloosa. They all are in the finest of spirits

- **Develop a compelling project website to recruit and retain volunteer transcriptionists.** As noted in our initial proposal, receipt of partial funding led to this portion of the project being truncated. The resulting [website](#) contained only the diary and letters to be transcribed, the transcription tool, a FAQ, user surveys, and the required OU Institutional Review Board documentation. This abbreviated version of the website was completely sufficient for the project's needs. We originally named the project "My Dear Wife" for the salutation used by Lt. Garrett to address his wife in his letters, but we opted for "Transcribing the Past: Civil War Manuscripts" to clarify the nature of the project, and to establish the URL "transcribe.ou.edu" as a hub for any future transcription projects. Future projects can use the title "Transcribing the Past" and indicate the subject matter in the subtitle.



Figure 2: This view of the transcription website (taken at the end of the project) shows the landing page where participants could choose to transcribe pages from *My Dear Mary* (the letters) or *A Soldier's Life* (the diary), with a navigation menu at left. A notice that no more transcriptions are needed was posted on the website when enthusiastic transcribers continued signing up for the project after all of the available pages were transcribed.

For recruiting transcriptionists we relied on advertising fliers distributed by e-mail, social media, hard-copy postings on campus, plus a [news release](#) and newspaper articles to publicize the project. (See article links in 4th bulleted objective, page 6.) We also created an optional Facebook group for interested participants to discuss the project among themselves, but it did not generate much activity since the transcribers were not collaborating or editing one another's work. The user satisfaction surveys were completed by only 9 users. They indicated overall satisfaction with the website but wanted more navigational options between images and ability to edit their work after submission. These user suggestions will be incorporated into the planning for any future transcription projects by our library.

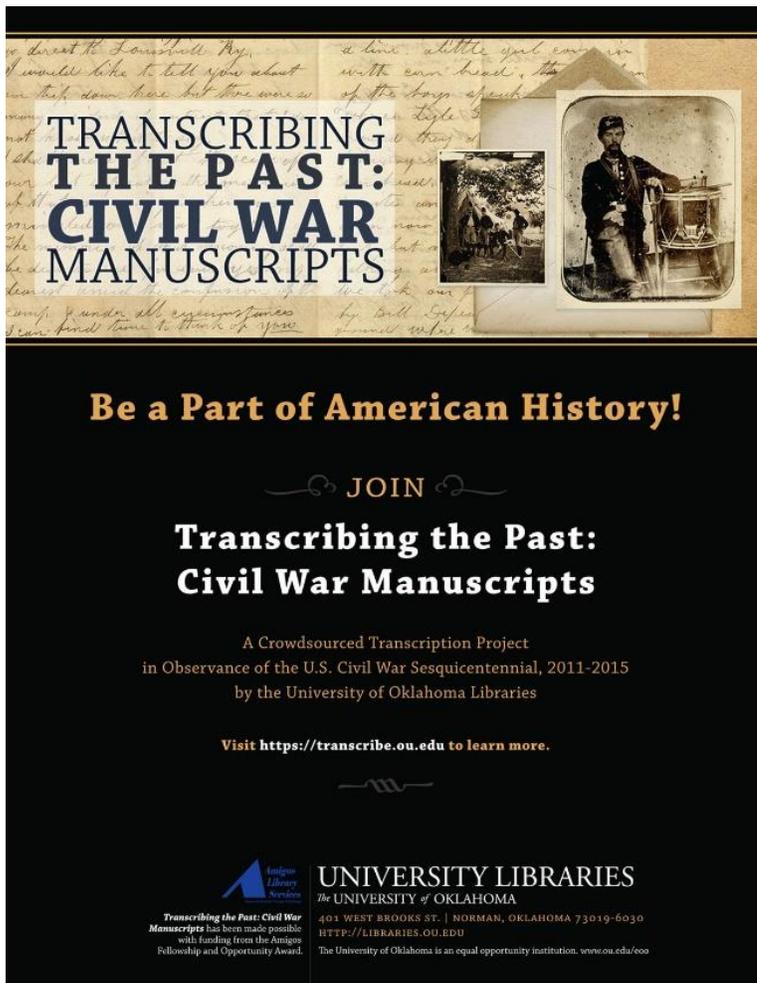
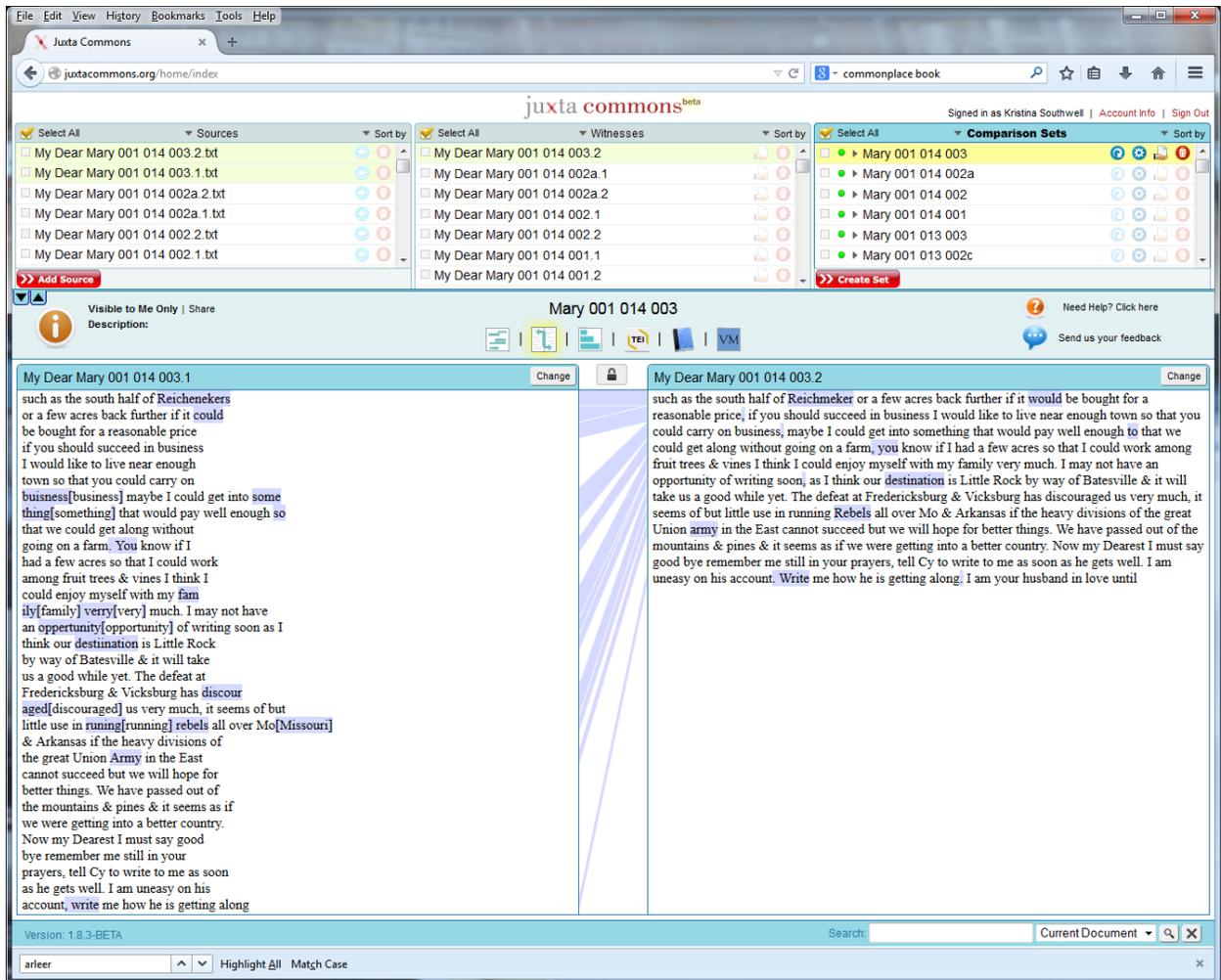


Figure 3: The advertising flier for *Transcribing the Past: Civil War Manuscripts* was distributed by e-mail, social media, and hard-copy postings on the Norman campus of the University of Oklahoma.

- **Triangulate transcribed manuscript materials.** We used the website [JuxtaCommons](#) to upload and compare the two transcriptions received for each page of the letters and diary. This free online tool highlights the differences between two transcription text files. We compared the transcriptions to the scanned images and chose the best version to edit and accept. This part of the project was more time-consuming than we expected due to the

many varied ways in which transcribers interpret handwritten words, capitalization, and punctuation. It required three months of full-time work. In the future we will consider allowing transcribers to edit and collaborate on transcriptions to minimize the amount of staff time spent on this activity.

Figure 4: This screenshot shows a sample view of the transcription reconciliation process using JuxtaCommons. The side-by-side comparison panes show each of the two transcriptions created by volunteers, with the differences highlighted.



- Promote and make freely available unique Civil War collections in observance of the Civil War Sesquicentennial.** The digital images of the letters and diary are freely available on the project website and our institutional repository, [ShareOK](#). Publicity for the project, including a [press release](#) and articles in [the OU Daily](#), the [Norman Transcript](#), and the [Southwestern Archivist](#) generated a surprising level of public interest in the project and OU's special collections. The response was so great that after the articles were published we continued to receive inquiries from hopeful transcribers even after all pages were finished. There was clearly a strong demand from the public for more materials to transcribe.

- **Incorporate the final product in our institutional repository to make it accessible to the public and promote the special collections of our institution.** The scanned letters and diary, along with their transcriptions, were uploaded to our institutional repository, [ShareOK](#), on May 21, 2015. They are freely accessible to the public for reading and research.

SHAREOK.ORG
A central, open-access repository of research by members of Oklahoma Higher Education

SHAREOK Home → The University of Oklahoma → OU - Library Special Collections → View Item

Transcribing the Past Project Data

OU Libraries

URI: <http://hdl.handle.net/11244/14633>
Date: 2015-05

Abstract:
Transcribing the Past was a crowd-sourced transcription project to commemorate the 150th anniversary of the U.S. Civil War. The project was made possible with funding from the Amigos Fellowship and Opportunity Award. All included files are property of the University of Oklahoma Libraries. They are made available under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International license.

[Show full item record](#)

Files in this item

Name	Size	Format	Description	View/Open
README.pdf	60.78Kb	PDF	description of data ...	View/Open
MyDearMary-Compil ...	2.018Mb	PDF	compiled transcripts ...	View/Open
Kroff-CompiledTra ...	745.8Kb	PDF	compiled transcripts ...	View/Open
MyDearMary-Images.zip	1.090Gb	Unknown	scanned letters	View/Open
Kroff-Images.zip	417.8Mb	Unknown	scanned diary pages	View/Open
MyDearMary-FinalT ...	534.0Kb	Unknown	transcription text ...	View/Open
Kroff-FinalTransc ...	133.2Kb	Unknown	transcription text ...	View/Open

Figure 5: The transcription files and scanned images of the diary and letters are freely available to the public for download at [ShareOK](#), which is a joint institutional repository shared by the University of Oklahoma and Oklahoma State University. The data include zip files of the full size diary and letter images, the individual text file transcriptions, plus a set of compiled transcriptions in PDF form for easy reading. The PDF compilations also contain background information on the writers of the diary and letters. The readme file provides information about the project, and summaries of the collections that were transcribed.

- ***Serve the needs of the scholarly and library communities through the project content and the development of a process which can be replicated.*** The general framework of our project can be replicated by other libraries to produce crowdsourced transcriptions, and the code for our customized transcription tool could be adapted for use by others. However, it was specifically tailored for our project to permit double-blind triangulated transcriptions, as well as anonymize all participants in order to comply with the requirements of our IRB study. Those unusual elements may not be needed by other libraries' transcription sites. Technology within and outside our library also changed so quickly during our 3-year planning and implementation period of this project that other libraries may not replicate our exact process using the same software - particularly if they use a different CMS. Regardless, we gained valuable experience in planning and executing a technology-centered project, and we hope others will be able to use what we learned in planning their own transcription sites. We do plan to continue offering manuscript collections for public transcription, and the experience gained during this project will be reflected in the processes we use to achieve this. In the meantime we are offering overviews of the project and what we learned from it to other librarians. Jacquelyn Reese and Logan Cox first presented the project results to OU's Digital Humanities Working Group on April 8, 2015. They delivered a second presentation to the [Central Oklahoma Archivists League](#) and the OU Collections Professionals group on June 17, 2015. Jacquelyn Reese is currently writing a scholarly article on the project that is planned for submission to *Archival Issues* in October 2015.

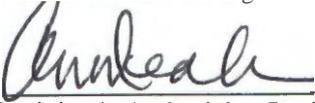
Amigos Fellowship Program

FINAL REPORT-FINANCIAL STATEMENT

List the expenditures for your Amigos Fellowship by type of expense and complete the institutional certification below.

Supplies	_____	
Services e.g., ILL, photocopying, database searching	\$4 555.50	_____
Telephone	_____	
Travel	_____	
Fees	_____	
Other (Describe) (Institutional overhead charges and capital expenditures such as computers or equipment and operating expenses are not eligible for funding.)	_____ _____ _____ _____	
Total Funds Awarded		\$5,290.00

I certify to the best of my knowledge and belief that the data above is correct and all outlays of funds were made in accordance with the terms of the Amigos Fellowship award.

Signature of Recipient's Authorizing Institutional Representative  Date 6/19/15

Print or type:

Name Andrea Deaton

Telephone 405-325-4757

Address Office of Research Services, 201 Stephenson Parkway, Suite 3100, Norman, OK 73019