

**Final Performance Report for National Endowment for the Humanities
Grant HD-50927-10
Crowdsourcing Documentary Transcription: an Open Source Tool
September 30, 2011**

Executive Summary

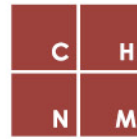
Between July 1, 2010 and June 30, 2011, the Center for History and New Media has completed the user testing and development of *Scripto* <scripto.org>, our open source tool for crowdsourcing documentary transcription. We have had a successful implementation of the tool with the *Papers of the War Department Project*.

Development of the Basic Tool:

Web Developer Jim Safley used the first six months of the grant period to create the *Scripto* library in PHP. This library allows for the connection of the basic elements of the transcription tool: the collection content management system, the image viewer, and the MediaWiki-driven transcription window. In January, Web Developer Jim Safley made the code for *Scripto* available for testing through GitHub <<https://github.com/chnm/Scripto>>. This site allowed any interested user to download the code, install it, comment on it, and contribute to development. As a result testers from Alberta, CA, Austin TX, Cambridge MA, and elsewhere downloaded the code and submitted a number of helpful development issues that eventually resulted in improvements to the functionality.

User Testing

Sharon Leon, Ken Albers (CHNM Web designer) and Jim Safley (CHNM Digital Archivist and Web Developer) conducted three rounds of user interface design testing for *Scripto* during the development of the basic tool. In August, once Safley had sufficiently mapped out the functional requirements for the tool, the *Scripto* team met with the editorial team from the *Papers of the War Department* to review wireframe drawings and possible layouts for the tool's functional web interface. Albers proposed a layout that was quite similar to the design mock-ups submitted with the initial grant proposal, which included a vertical split screen with the digital image to be transcribed in a left window and the editing/transcription window on the right. After interacting with this basic layout the PWD editors expressed their concern about the narrow image viewer and the ways that it would constrain a user's ability to view a whole line of script at once. This critical feedback resulted in the first revision to the tool's user interface.



In September, Albers, Leon, and Safley returned to the editors with a functioning mock-up of the tool, which included a horizontal split screen with the image viewer on the top and the editing/transcription window on the bottom. The PWD editors reacted much more positively to this interface because of the increased width of the image viewer window. They were also quite pleased with the ability to rotate, pan, and zoom the images. During this round of testing, the users offered suggestions about the location of the page navigation and the links to helper materials such as the style guide. Albers and Safley integrated this feedback into the subsequent build of the tools, which they then implemented with the administrative interface of PWD. This implementation represents the third iteration of the initial user testing. We will be collecting ongoing feedback from the editors over the course of the early part of 2011, in preparation for making the tool available to general users on the PWD site.

Papers of the War Department Implementation:

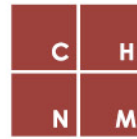
As we approached the public launch of *Scripto* in March, Project Director, Sharon Leon, authored a blog post to explain CHNM's considerations upon opening up a project to contributions from the community. The piece articulated four goals for the project, which remain central to our work:

- 1) Text contributed by the community is essential to improving the findability of individual documents;
- 2) The community-created transcriptions will provide editors with a first pass at transcription that can later be improved if resources permit;
- 3) The selection of documents for transcription by the community provides valuable insight into their interests in the archive; and
- 4) Opening the archive to community work allows us an opportunity to create community around historical materials.

These principals continue to guide our work as we engage with community transcribers and their contributions.

On March 17, 2011 CHNM launched the first implementation of *Scripto* with the *Papers of the War Department* project. This implementation involved a minor redesign of the PWD site to include access to the transcription functionality, and several new pages to support the process. With the redesign, users now have access to a sign-up page, a list of suggested documents for transcription, and a set of guidelines for their work. PWD's new functionality garnered immediate attention both in the digital humanities community and in the larger technology community. For example, the project received notice in a post by Curt Hopkins entitled "Crowdsourcing the Preservation of the U.S. War Papers" <http://www.readwriteweb.com/archives/crowdsourcing_us_war_papers.php> on *Read, Write, Web*, a leading source of technology news.

The enthusiasm of PWD users for the transcription project has exceeded our expectations. Between the launch of the functionality and June 30, we have had 240 users sign up for transcription accounts. Those users have begun transcription of 210 documents, each of varying numbers of pages, and have substantially completed the work on 175 documents. This work is proceeding at the pace that we expected, slowly but steadily. James Halabuk, the Digital History Associate responsible for



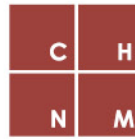
overseeing the contributions and editing the work, has engaged in 105 conversations about the documents and the process of transcription through the “Talk” feature of the interface. These interactions have resulted in improving the work of the transcribers and reassuring them that they have the support of the editorial team should they need it. Additionally, Halabuk has begun a series of blog posts for the *PWD* site updating the community on the project’s progress, such as “Three Months On—A Transcriber Spotlight” <<http://wardepartmentpapers.org/blog/?p=374>>, which introduces one of our most prolific contributors, Patricia Gerard. Finally, while site usage statistics are difficult to generate for the transcription elements of *PWD*, we can confidently count at least 130,000 pulls on the *Scripto* API, and roughly 3,000 views of individual user transcription records. This suggests healthy and steady return traffic to the site from our registered users.

Outreach

We have embarked on the process of doing outreach for *Scripto*. In September, we launched the tool’s website: <<http://scripto.org>>. This website is a central place of information about the tool, its development, and its user community. The site provides periodic updates from the designers and developers on the tools progress. Additionally, the site offers links to *Scripto*’s code base on GitHub, a repository for open source software <<https://github.com/chnm/Scripto>>. The availability of the tool on GitHub means that any interested user or developer can download the code for testing and implementation. Finally, we have secured a twitter account for the tool (@ScriptoTool), and currently have 31 followers. We expect this number to increase as the tool develops a user base.

In December 2010, *Scripto* received press coverage in the *New York Times* <<http://www.nytimes.com/2010/12/28/books/28transcribe.html>>. The article generally dealt with efforts of documentary projects to experiment with crowdsourcing, and did not refer to *Scripto* by name. We regret that the article failed to mention that NHPRC is a major funder of the tool, though it does include a link to the NHPRC website. We have tried, to no avail, to obtain a correction or an addition to the article on this matter. Despite the oversight, the article resulted in eight documentary editors volunteering to test *Scripto* with documentary projects and documentary editing courses. The volunteers include among others Cathy Moran Hajo of New York University and the Margaret Sanger Papers Project, R. Darrell Meadows of the Papers of the Kentucky Civil War Governors DDE and the Kentucky Historical Society, Amy Rudersdorf of the State Library of North Carolina, and Mark Shelstad of the Archives and Special Collections at the University of San Antonio Libraries. We are excited about the widespread interest in *Scripto* as contributing to one element of the larger process of documentary editing.

In the period after the launch of *Scripto* with *PWD*, we have approached outreach for *Scripto* in three main ways over the course of the last six months. First, we have provided periodic update through blog posts. On the *Scripto*, blog Sharon Leon and Jim Safley have made readers aware of the progress on the generalized code libraries. James Halabuk has handled all of the transcription related updates on the *PWD* blog. Second, we have established twitter accounts for both *Scripto* and *PWD*, which serve to amplify the reach of the blog posts, and to allow for interaction with followers. Finally, we have established a Google Group specifically for web developers who want to raise issues



about the code and functionality. Though the group is currently quiet, we expect it to serve as a site for discussion and support as testers begin to implement and modify the software.

In addition to these general and ongoing outreach efforts, CHNM Web Developer Patrick Murray-John discussed *Scripto* with the attendees at the Maryland Institute for Technology in the Humanities' API Workshop <<http://mith.umd.edu/apiworkshop/>> in February. Workshop participants included leading programmers and developers from Google Maps, Open Library, Data Unbound, Freebase, and digital humanities centers throughout the country. Murray-John's presentation allowed the participants to consider *Scripto* in the wide array of tools at their disposal to access and enhance existing streams of digital content.

While the *PWD* implementation has garnered the majority of the focus of outside attention for *Scripto*, the tool is also beginning to gain some traction as individuals and institutions consider their own projects. Thus, *Scripto* has been mentioned in a host of blog and forum posts from a wide variety of interested parties, including the European Journalism Centre <<http://www.ejc.net/resources/featured/scripto/>>, and in the Ideas Incubator of the Open Knowledge Foundation <<http://ideas.okfn.org/ideas/59/generic-transcription-tool>>. This warm reception suggests that there is a great interest from the digital humanities community at large in crowdsourcing transcription.

Overall, we have been pleased with the progress of the project, as we have completed all of our deliverables. We welcome the opportunity to collaborate with other members of the digital humanities community so that *Scripto* can become a useful part of the growing ecosystem of tools. Similarly, we welcome further input from and discussion with the NEH-ODH staff about ways to extend our work. Please feel free to contact me at any time, should you wish to visit the Center or to discuss developments in digital humanities more broadly.

Sincerely,

Sharon M. Leon
Director of Public Projects and Associate Professor,
Roy Rosenzweig Center for History and New Media
sleon@gmu.edu
703-993-3831