

**CoSA – NARA Great Ideas Exchange Day
August 15, 2018**

Case Study: Crowdsourcing the Alabama WWI Service Records

Background

The Alabama Department of Archives and History (ADAH) is a state agency with responsibility for the state archives, records management, and the state history museum. The agency is housed in a facility dedicated in 1940 as the Alabama World War Memorial Building. In the years immediately after World War I, the ADAH actively collected wartime materials including manuscript collections, artifacts, federal and state publications, and government documentation of Alabamians' service.

As the centennial of World War I approached, the ADAH developed a range of commemorative activities including acquisitions, exhibitions, symposia, a reader's theater, and the creation of K-12 curricular resources. It also prioritized the digitization of World War I-era materials. Among these were 111,000 index cards containing compiled service records of Alabamians who served in uniform, as well as some civilian employees of military bases in the state.

The service records were chosen for both their genealogical value and their potential as the subject of a crowdsourcing project. In recent years, the ADAH had successfully implemented small-scale crowdsourcing to identify Montgomery residents in a series of mid-twentieth-century photographs. Having observed the success of crowdsourcing projects for the 1940 U.S. Census and the Freedmen's Bureau records, and desiring to develop a project that would engage Alabamians with the history of World War I, we decided that the service records were a logical, if ambitious, next step.

Method and Scope

Staff and onsite volunteers scanned the 111,000 cards over a period of eighteen months in 2016 and 2017. They were added to the ADAH's digital collections site (<http://digital.archives.alabama.gov/>) as compound objects grouped by the initial letter of the surname within each county. With only minimal metadata captured during digitization, searchability was extremely limited.

The ADAH lacked sufficient IT resources to develop our own robust crowdsourcing platform, as some states had done. We proposed a partnership with FamilySearch making use of their technology, but their own scheduling priorities precluded reaching an agreement in time to meet our needs. In late 2017 the ADAH was testing a homemade system using Google Forms, with unsatisfactory results, when we came across FromThePage (<https://content.fromthepage.com/>), an open-source platform developed by Brumfield Labs.

FromThePage had been publicly available for six years and was used extensively by manuscript repositories for free-text transcription of diaries, correspondence, and similar materials. The platform did not support the transcription of structured data fields, which we needed for the World War I records. Brumfield Labs enthusiastically embraced our suggestion to enhance FromThePage's capabilities. In November 2017 we entered a two-part agreement for the development of 1) a user interface for structured data fields and an updated project management portal; and 2) the ability to pull

images from our CONTENTdm collections in real time and then push metadata from FromThePage to CONTENTdm using the International Image Interoperability Framework (IIIF).

The Friends of the Alabama Archives committed to fund the software enhancement. Recognizing the benefit of this undertaking for all potential users of the open-source code, we invited other state archives to share some of the costs. The state archives (or their friends groups) of Indiana, Virginia, Missouri, and Maryland contributed financially. The Council of State Archivists (CoSA) served as a fiscal agent for administering the multi-state support and signed the services agreement as an additional client. The software enhancement was completed and successfully tested by March 2018. Without the time or expertise to stand up our own server for running the code, we opted to subscribe to a hosting solution provided by Brumfield Labs. The Alabama Genealogical Society joined as a partner at this stage of the project and contributed half of the cost of the hosting service.

The default access setting in FTP is to make all of the projects from many repositories completely open to volunteers who can self-register and dive into transcription. As we approached implementation, the ADAH opted to close the project space on FromThePage so volunteers would have to request a login from Meredith McDonough, ADAH digital assets coordinator and manager of the project. We had ambitions of enlisting volunteers to adopt the records from their own counties, thereby connecting Alabamians with the service histories of men and women from their corner of the state. This being new territory for us, we also wanted to ramp up the level of activity in a controlled way.

On Saturday, April 7, we hosted an orientation for a dozen invited volunteers. We introduced the platform and a printed user's guide, provided wi-fi and desktop stations so they could work, and then watched and listened. Their questions and feedback informed some changes in the user's guide as well as our field descriptions. Despite a few hiccups in the first half hour, they were soon clipping along and went home looking forward to doing more work.

By Monday morning, Meredith's inbox was inundated with requests from volunteers. An Alabama Genealogical Society tweet about the project had been shared by some of their peer institutions, and awareness took off. Social media posts further spread the word, and we added a project page to our Alabama History DIY site (<https://www.alabamahistorydiy.org/>).

Progress was rapid. Ten percent of the cards were transcribed in the first twelve days. At the two-month mark, 70,000 cards had been transcribed, and the entire collection was finished in three and a half months—much sooner than what we believed to be an optimistic goal of seven months. In total, 82 active volunteers donated 3,521 hours of work—the equivalent of one person working full-time for 88 weeks. Each card took an average of 1.9 minutes to transcribe.

Meredith is currently conducting a mix of random and targeted quality-control checks in the captured metadata, then migrating it to CONTENTdm before re-indexing the service records collection. The compound objects arranged by initial letter of the surname are replaced with individual cards, associated with newly captured metadata. Counties are updated as they become available, and Meredith expects to complete the process by year's end.

Outcomes and Lessons Learned

We are barely a month beyond the completion of the transcription phase of the project, and still evaluating how we might improve the process next time. We are tremendously pleased with several obvious outcomes. The project:

1. Created a meaningful World War I centennial commemorative program involving the public
2. Accomplished a massive indexing project more quickly than we could have in-house
3. Produced technology that can be adapted for other crowdsourcing projects
4. Established a base of “virtual volunteers” and formed relationships in a new community of users

Especially gratifying have been the many expressions of affirmation and appreciation received from volunteers, including this small sample:

1. “My background is in history and I love any chance to get involved in allowing the general public easier access to information from the past.”
2. “I have no ties to Alabama but I feel it is important for all of us to help all of us!”
3. “I am a displaced Alabamian living in St Louis, so nothing would make me happier, or more proud, than to be involved in this project.”
4. “I have more than once felt great honor (and a little choked up) when transcribing an old index card on which the service member is listed as having died in battle or of an illness . . . It's been a real surprise to me how affecting this has been.”

The passage of time will bring greater perspective on lessons learned, but a few are readily apparent.

1. A tremendous pool of interested and skilled volunteers exists who are ready to embrace projects such as this.
2. Their involvement is not based merely in a sense of magnanimity. The volunteers who committed hours at a time to working at home or in coffee shops were drawn into the tiny but powerful stories represented by these service records. Some shared with us that they peeled off the transcription work intermittently to do their own research on the places, units, and battles notated on the cards.
3. The notion that we could control the rate of growth of involvement, or focus energy on specific geographic parts of the state, was generally flawed. If we made the cards available and issued logins, the volunteers were going to forge ahead and take every opportunity possible.

Next Steps

We have identified additional records sets that are candidates for our next phases of crowdsourcing work, including World War II and Korean War service records. Licensure records and legislative directories are additional possibilities. Concurrently, we are continuing our conversations with FromThePage adopters to consider what additional improvements to the open-source code might be the focus of subsequent collaborative efforts. Already, the Library of Virginia has submitted code to improve performance on multi-page records. We are also interested in adding support for multi-record pages, such as state census records and ledgers. We welcome your input!