NEH WHITE PAPER

Transcribing and Linking Early American Records with Scripto and Omeka S
Grant # HAA-256158-17
Roy Rosenzweig Center for History and New Media, George Mason University

Contents

1. Project Summary ................................................................. 1
2. Project Origins and Goals...................................................... 2
3. Project Activities, Team & Participants................................. 5
4. Project Outcomes............................................................... 8
5. Project Evaluation and Impact............................................. 9
6. Project Continuation and Long-Term Impact.........................10
1. Project Summary

This grant sponsored work aimed at increasing the availability, usability, and sustainability of digital cultural heritage collections. Those efforts revolved around community-sourced transcriptions: enabling a broad user base to generate and share transcriptions of historical documents. Between 2017 and 2020, Staff at George Mason University’s Roy Rosenzweig Center for History and New Media created an updated an improved version of Scripto, an open-source transcription tool, designed to work on its Omeka S platform.

Project staff designed the improved Scripto tool using a project already implementing the existing version of the tool within its collection. The Papers of the War Department, 1784-1800, an innovative digital archive that served as the test case for the original Scripto and which has an established and enthusiastic user base, served as the test case for the development of the updated version of Scripto. Using that project as the test case required staff to first migrate its collection to Omeka S, a process which had the added benefit of stabilizing the archive and ensuring its long-term sustainability.

Finally, staff produced documentation guides and publications to support other organizations in their efforts to develop community transcription projects with Scripto and Omeka S.

The Scripto tool is available for download at the Omeka website, <omeka.org/s/modules/Scripto/>.

The upgraded PWD resides at <wardepartmentpapers.org/s/home/page/home>.

Documentation for the Scripto Transcription module and “Transcribe” theme are available at <omeka.org/s/docs/user-manual/modules/scripto/>. 
2. Project Origins and Goals

The idea that user communities can be harnessed for input and volunteer assistance from students, visitors, researchers, scholars, and hobbyists to solve problems, vote on favorite things, collaborate, and contribute knowledge has circulated within academic communities for years. In the internet age, facilitating community participation and connecting online audiences is even easier. Once known as “crowdsourcing” and now more commonly referred to as “community sourcing,” it is an approach with the potential to both increase user engagement and to conduct work that would be difficult or prohibitively expensive to undertake in more traditional ways.

THE ORIGINAL SCRIPTO

Despite the promise of community sourcing, many cultural heritage institutions and universities have no easy way to engage their publics with their digital collections, as platforms are not designed to integrate easily with their existing content management systems. Staff at George Mason University’s Roy Rosenzweig Center for History and New Media (RRCHNM) have been working on addressing that challenge since 2010, when the Center received a NEH Office of Digital Humanities Start-Up Grant to develop Scripto, a free and open-source transcription tool designed to enable and encourage online public engagement with humanities themes found in digital collections.

One major advantage for choosing Scripto over other platforms for participatory projects was that it was built for common content management systems, including RRCHNM’s own free and open-source software, Omeka. With over 150,000 downloads and thousands of published digital projects, Omeka is a reliable and popular collections-driven content management system. Scripto became one of Omeka’s more popular plugins, encouraging scholars to connect with their communities by facilitating this type of intellectual work.

Developing a tool like Scripto required a project to use as the test case to inform both the plugin’s code base and the overall design of the user experience. The original Scripto used the Papers of the War Department 1784-1800 (PWD), a large and unique digital archival project that has resided at George Mason since 2005-2006, as its original test case.

THE PAPERS OF THE WAR DEPARTMENT ARCHIVE

PWD proved an ideal test case for the original Scripto. Its collection recreates more than a decade and a half of files that were destroyed in the fire that burned down the War Office in autumn of 1800. That fire consumed all the documents stored in the office; for nearly two centuries, historians believed that the first surviving document from the early War Office was a November 1800 letter in which then-Secretary of War Samuel Dexter lamented that “All the papers in my office [have] been destroyed.”

The destruction of the archive represented a terrible loss for scholars. The War Department played an outsized role in the history of the early federal government. It was the nation’s single largest consumer of fabric, clothing, shoes, food, medicine, building materials, and other supplies; it administered the first public pension program for widows and orphans of the War of Independence;
and it managed relations with indigenous peoples on the frontier. During this period, the War Department accounted for seven of every ten dollars in the Federal budget, excluding service on the debt. In the country’s early years, the War Department was (along with the Post Office) the only contact most Americans had with their new federal government, making the files of the War Office the young nation’s first “national archives.” For decades researchers who visited the National Archives in search of letters and documents from this period were informed that fire had destroyed the collection in its entirety.

Fire had indeed destroyed all of the files stored in the building housing the War Office. But in the 1980s Ted Crackel, a historian of the early years of the Republic who focused on the new nation’s military, began locating copies of some documents and correspondence produced by scribes and assistants in the War Office. Clearly copies of many of the files of the War Office had survived.

In 1989 Crackel proposed the idea of reconstituting the files in the archive by locating and scanning the documents that would have been in its collection on the night of the fire. Crackel began the PWD project in 1993 and served as its original Editor-in-Chief. In the following years he and his staff visited hundred or archives, scanning nearly 50,000 documents. The collection was too large for a letterpress edition, and the original staff began exploring the possibility of a CD-ROM version. When Crackel left the project to become Editor-in-Chief at the Papers of George Washington, RRCHNM applied to adopt the project in the summer of 2005, and began work on its completion in Spring 2006 under the direction of new Editor-in-Chief Christopher Hamner. RRCHNM adopted the project with the intent to present the collection, then numbering some 43,000 items, as a website with document images and a search engine.

**MAKING PWD SEARCHABLE**

Given the scope of materials involved and the limitations on funding support, the editors opted initially to index the documents with basic metadata that would allow a search engine to locate them. With that understanding, Hamner, Assistant Editor Ron Martin, and nearly a dozen graduate research assistants from Mason’s History Department began the multiyear process of systematically reviewing the documents and entering the information into a database.

By 2010, that basic metadata made the documents visible to the search engine, provided the researcher had a good idea of what she was looking for. Armed with a date, sender, or recipient, the search engine could locate a specific document in the corpus. For researchers with less concrete terms, the index proved less usable. A scholar searching for items related to pensions, the Whiskey Rebellion, the Quasi-War with
France, or any of dozens of events, ideas, and themes, could now identify those sources with the
search engine. Funding for the indexing effort ended in June 2013, leaving the PWD with a two-
tiered index for the entire collection.

Though the metadata collected in the index made the archive vastly more accessible, PWD staff
understood it as a partial solution. Ideally, the archive would offer not just high-resolution images of
the documents but transcriptions of them. Full transcriptions would make documents available to
the search engine, and would make individual documents easier for visitors to use. Given the size of
the archive, however, and the nature of the documents—every item in the collection is handwritten,
making automation unworkable—such a task was prohibitively expensive.

HARNESSING THE POWER OF THE COMMUNITY

The staff at the PWD lacked the capacity to transcribe any significant part of the collection. But
there were early indications that a community of researchers and scholars might volunteer to help
with the task of transcribing. Most scholars using the archive make rough transcriptions of
documents to use in their work. PWD received numerous emails from researchers suggesting
improvements to the index based on those researchers’ own work. A dedicated group of volunteers
transcribing documents had sprung up organically around the online archive.

In an effort to tap the efforts this pool of volunteers, RRCHNM proposed and built an open source
transcription tool to allow users easily to submit their transcriptions and that knowledge back to the
archive. That tool, the original Scripto, would allow PWD (and eventually other digital archival
projects) to harness upon the knowledge and efforts of the thousands of interested researchers,
scholars, and students who work with these materials. Gradually, users’ aggregate efforts would
increase access to and usefulness of the entire archive, without requiring massive time- and labor-
intensive institutional effort.

Figure 1:
The original Scripto
interface
The original version of Scripto debuted in 2011 and enjoyed significant success with PWD. More than 2,700 individuals successfully transcribed some 1,800 items from the PWD collection, many of them multi-page documents. The debut of the community sourced transcriptions also increased user interest in the site and enthusiasm for its collection. And the user-supplied transcriptions, in turn, increased the ease with which documents could be accessed and used, widening the audience for the archive itself.

Omeka continued to develop in new directions following its debut in 2008. In 2012 staff at RRCHNM began work on the next generation of the software, Omeka S. Omeka S features a completely new code base with a data model that fully integrates linked open data principles, satisfying the needs of larger institution-based users.

UPDATING SCRIPTO

The development of Omeka S had important implications for Scripto, as well. Major software revisions create a need to redesign the platform’s accompanying extensions. This grant sponsored the redesign of the Scripto tool to work on Omeka S. Redesigning a complicated plugin such as Scripto required a project already implementing the existing version of Scripto with a large digital collection to serve as a test case. PWD again provided an ideal case study for the upgrade. Its infrastructure was custom-built in-house at RRCHNM in the mid-2000s and lacked a seamless upgrade path to a standardized system. To ensure the long-term stability and sustainability of the PWD archive, staff at RRCHNM elected to move the project to Omeka S, where it could then serve as a test case for an updated and redesigned version of the Scripto tool.

As part of the effort to increase the functionality and usability of the new Scripto tool, staff at RRCHNM used the redesign as an opportunity to develop guides and documentation to aid others developing community transcription projects. Those new materials, including user guides and screencasts that walk users through the new tool, debuted with the new Scripto.

3. Project Activities, Team & Participants

PROJECT STAFF

The staff brought decades of experience working in the digital humanities to the project. Dr. Sheila Brennan wrote the original grant and served as Primary Investigator and Project Director until her departure from RRCHNM in 2018. Brennan brought years of experience directing a dozen digital projects. Jim Safley served as Lead Web Developer, bringing a decade and a half of experience designing projects at RRCHNM—indeed, Safley was the web developer for both PWD and the original Scripto. Safley migrated all the data in PWD to Omeka S and led development of the Scripto plugin and of the Omeka S installation for the PWD redesign. Kim Nguyen served as Lead User Experience and Web Designer, and brought years of award-winning experience working on Omeka. Nguyen led the design process for all user interface and user experience elements of the Scripto module, as well as the visual elements of the PWD theme and generalized “Transcribe” theme. A
trio of doctoral students a George Mason’s History PhD program worked as Project Managers and ran user testing during development. Megan Brett, Jessica Dauterive, and Alyssa Fahringer guided the project, coordinated staff efforts over three years, and led the project’s outreach efforts. Dr. Christopher Hamner, who serves as Editor-in-Chief of PWD, was initially written into the grant as a project advisor. Following Brennan’s departure, Hamner took over as Primary Investigator. Hamner has served as editor of PWD since the project arrived at RRCHNM in 2005-2006 and served as an advisor on the original Scripto grant. Years of experience working with members of the project staff on PWD made for a seamless transition.

PROJECT TIMELINE

Work on the project closely followed the timeline outlined in the original grant proposal.

The work proceeded simultaneously in three areas: staff completed the process of updating and improving Scripto for use with Omeka S; migrated PWD to Omeka S, providing a path for its long-term preservation and enabling its use as the test case for the improved Scripto tool; and produced documentation guides and publications to support other organizations in their efforts to develop community transcription projects with Scripto and Omeka S.

Lead Developer Safley oversaw the work of migrating, stabilizing, and upgrading PWD to Omeka S, work that involved moving than 43,000 digital items, along with the metadata and associated image files for each item—nearly 200 gigabytes of information in total. With the migration complete, Safley and Web Designer Nguyen focused their efforts on the redesign of the new Scripto module. That redesign was based on feedback from groups of established PWD transcribers, and solicited input from the user community at multiple points, including a full review of the module prototype. The updated module offers a number of new features based on user input, including options that allow admin users to more easily view and review transcriptions; to track the progress and mark the status of transcriptions; to track who is transcribing what across all Scripto projects within an Omeka S installation; to roll back revisions when necessary; and to import and sync transcriptions and updates with appropriate resource elements within items in Omeka S. The alpha version of the new Scripto module debuted in August 2018; the beta version followed in March 2019. The final version of the tool, incorporating minor fixes and improvements discovered in beta testing, launched in January 2020.

Work on the redesign of the PWD site unfolded alongside the work on the Scripto module. Nguyen developed a new front end for the site that makes navigation simpler and more intuitive, and gives the site a cleaner, more contemporary look. Design and development of the reimagined site occurred in conjunction with Mason’s Assistive Technology Initiative to ensure accessibility for the largest possible number of users. The redesigned PWD Site launched in April 2019.

As the new Scripto module and PWD site redesign moved through planning, testing, and implementation, staff worked to create new outreach materials to increase the user base and to aid users in taking full advantage of the new features. One set of supporting materials is aimed at end users, and includes video screencasts that walk through the process of installing, updating, and
configuring Scripto; transcribing documents; and using the new Notes function. A separate set of support materials drafted by the project Lead Developer Safley provides technical documentation for experienced developers looking to modify or enhance the Scripto module for their own Omeka S projects.

**CHALLENGES**

As we progressed, project staff concluded that the biggest development challenges in moving Scripto from the original Omeka to Omeka S led directly to our biggest accomplishments. Three in particular stood out to the development team.

The first challenge stemmed from the need to create a multi-project setup. The original Scripto is, essentially, a one-project setup, where each item can be transcribed only once. That limitation was not suitable for Omeka S’s multi-site architecture. In response, we introduced a multi-project setup, where each item can be transcribed in many ways for many purposes. That in turned necessitated the establishment of a syncing process, whereby project owners can define a corpus of items and create Scripto items that exist parallel to their original.

The second significant improvement involved improving the transcription and review workflow. The original Scripto offered no review process to speak of, introducing the possibility of lower-quality transcriptions. To encourage and facilitate higher-quality work, we introduced a transcription and revision workflow, allowing transcribers to mark their work as “completed” and reviewers to mark it as “approved.” That locks the approved version to the specific MediaWiki revision and prevents under-reviewed content from being published.

The third improvement revolved around improving the overall user experience. The original Scripto featured a simple, impersonal interface. The updated Scripto provides more features that center around the individual user. Users can easily view their recent contributions, and can add individual pages to their watchlist. Users can also see which items in the project are new, which are in progress, and which have been approved, making it easier to determine what resources need transcription.

![Figure 2: The updated Scripto interface](image-url)
4. Project Outcomes

The redesigned PWD site launched in April 2019. The updated *Scripto* tool launched nine months later in January 2020. New user support materials accompanied the debut of the updated *Scripto*, aimed at both the transcriber community and at developers looking to modify or enhance the *Scripto* module on their own *Omeka S* projects.

The audience for the work sponsored by the grant as a whole includes several groups, with significant overlap between them.

One is the audience for the updated *Scripto* tool itself. They represent a varied group of users. *Omeka S* is aimed at larger institutions that require a single point of administration for installation, software upgrades, and the extension of functionality for all of the sites developed in the network. Updating the *Scripto* tool for use in *Omeka S* allows these larger institutions and projects to take advantage of the new functionality. Sites like *Historic New England* (<www.historicnewengland.org>) have already implemented the new *Scripto* on *Omeka S*, and we expect adoptions to continue as the module gains more visibility. (The COVID-19 pandemic, and its associated shutdowns and budget cuts, seem to have postponed some projects’ plans for implementation: for example, *Digital Maine* (<digitalmaine.net>) was exploring a move to *Omeka S* before the pandemic and problems with their physical plant put those plans on hold.) The Gilder-Lehrman Institute of American History recently indicated that they will be moving several of their collections, and the associated community transcription efforts, to the *Omeka S* platform for use with the new *Scripto* tool. To date, the new *Scripto* module has been downloaded from GitHub 245 times since the launch in January 2020. The total lifetime download count for the *Scripto* module since its debut in 2011 now stands at 346.

Work on the updated *Scripto* also focused on the user experience for the end users who transcribe and upload documents using the tool. They represent another audience for the work completed under this grant. Our design efforts in this area focused on giving the *Scripto* tool a more intuitive feel while simultaneously expanding its functionality. That community formed the primary audience for work on the *Scripto* interface and for the support materials and screencasts developed to walk users through the process of transcription using the new version of *Scripto*.

The work on *Scripto* undertaken under this grant utilized the PWD as its test case. The community of scholars, researchers, teachers, and genealogists who have been using the PWD archive for a decade and a half represents another audience for the work completed under this grant. Having PWD as the test case also brought a built-in audience of hundreds of volunteer transcribers who were already using the original version of *Scripto*. The redesign of the PWD site represents the most visible update for the user community, and provides a cleaner, more modern experience. Much more important, though less immediately apparent to end users, is the work done to stabilize and migrate the collection to *Omeka S*, work that ensures that the archive will be available well into the future.

We have received encouraging early signals that suggest that the visibility of the PWD collection itself continues to grow. National History Day, a national non-profit that sponsors a project-based annual competition for history students across the country, has selected PWD as one of seven...
spotlight projects for inclusion in its 2022 contest themebook. The 2022 NHD competition, “Debate & Diplomacy in History: Successes, Failures, Consequences” will put the redesigned PWD site in front of tens of thousands of middle- and high-school teachers, and will reach over half a million students nationwide.

Anecdotal data gathered during the pandemic also suggests that the quarantine, and the transition to online instruction in most states for at least parts of the 2020-21 school year, encouraged interest in both the PWD archive and in community-sourced transcriptions. The closure of physical archives and the impossibility of travel led teachers, professors, and researchers to seek out online archives that could be used for research courses and independent research projects.

The audience of scholars, researchers, teachers, and students who utilize the PWD archive benefit from the synergy that exists between the collection and the redesigned _Scripto_ tool. Community-sourced transcriptions make collections more visible to search engines and to the general public, building a larger user base; users who invest in the effort to transcribe historical documents increase their own interest in the collection as a whole, and help make even more documents visible. That synergy reflects one of the most meaningful long-term investments the work sponsored under this grant has made possible: the opportunity for broad, multi-faceted constituencies to engage in meaningful historical work.

**5. Project Evaluation and Impact**

Early indications of the long-term impact of the updated _Scripto_ tool are already visible. Institutions are already using the tool, or are making plans to adopt it, and the number of downloads _Scripto_ has received in the eighteen months since its release offers evidence of a robust user base. National History Day’s selection of the PWD archive as one of its spotlight projects for the 2022 annual contest will further raise the visibility of the redesigned site among classroom teachers and among middle- and high-school teachers, and suggests ways in which the redesigned, stabilized site will continue to find new audiences beyond academic historians.

The long-term impact of the stabilized and redesigned PWD archive itself is difficult to overstate. The archive has resided at GMU since 2005-2006, but the history of the project dates back even earlier. In the thirty-five years since Ted Crackel first proposed reconstituting the archive, PWD has become an invaluable resource for scholars of early America, cited in dozens of historical articles and monographs. The profile of the online PWD archive among scholars and researchers continues to rise as more become aware that this large and essential trove of records, long thought gone forever, has been reconstituted online. That makes the work on stabilization and sustainability of the archive completed under this grant of monumental importance. Migrating the collection to _Omeka S_ guarantees that this critical archive—which has no brick-and-mortar analog—will continue indefinitely into the future.
The precedent thus established magnifies the importance of the sustainability work conducted under this grant. Many early digital projects will soon enter their third decade, and those built on old software require (or will soon require) a path forward to long-term sustainability. Beyond its immediate direct use to historians of the early Republic, the work done on the PWD archive to ensure its long-term stability serves as a model for other digital projects that will soon encounter similar challenges.

6. Project Continuation and Long-Term Impact

The launch of the updated Scripto tool and its associated technical and user documentation marks a major milestone in the development of community-sourced transcription. The finished tool and its documentation will remain available for download on Omeka S.

The updated PWD archive will continue at GMU as it has for more than a decade and a half. Migrating the site to Omeka S under this grant stabilized the database and ensured the long-term sustainability of the project. Hamner will remain on as Editor-in-Chief of PWD.