

NDNP Awardee Interim Performance Report

NEH Award Number: PJ-50038-08

NDNP State: Washington

Submitted By: Shawn Schollmeyer, Marlys Rudeen

Report Date: March 31, 2012

1. Describe any changes that you have made or that you anticipate making in the project work plan or methodology from the award proposal submitted to the competition.

Minor modifications have been made to the 2010-2012 work plan to improve workflow efficiency and accommodate the OCR processing and metadata creation that is now being done in-house rather than submitting for processing by our vendor, Content Conversion Specialists (CCS), in Europe. By processing the content locally we are able to avoid the waiting periods for shipping drives to Romania and packages delayed by weather or international customs. However there has been a considerable learning curve to use the CCS docWorks application and a few metadata conversion issues that we needed to resolve. Microfilm evaluation had originally been created in Excel spreadsheets, stored for the processing team to access via "the cloud" as a Google document, then converted to an Access database used by the docWorks program. To minimize formatting problems created by converting Google formats to Access, we now store the data in Microsoft SkyDrive for better compatibility with docWorks' Windows-based software. The current version requires the Access database to compare the METS files against the original film evaluation data, a critical step in verifying data at the end of the processing cycle. We are still experiencing some issues with the conversion and are currently working with the vendor to resolve the issue.

In order to improve the learning curve for several different document processing stages and quick onboarding of any new staff and volunteers, careful documentation is being created. In some cases the vendor was not able to provide enough specific information for the NDNP workflow and project handling for remote and local processing data. Export QA and verification was not covered since there is so much variation between the grantees. This information is being compiled from various resources including other NDNP grantees at the Universities of Virginia and California, Riverside who also use the same application.

After the deliverable files have been exported from the docWorks program, a few changes have been made to streamline the QA check before running LC validation and verification programs. A few steps have been eliminated in favor of a docWorks recommended process using an out-of-box file comparison tool and their comparison file generator script instead of an earlier DOS command line tool. This has been further improved upon by using an easier to read file comparison application called Beyond Compare 3. This cuts down a considerable amount of time in reviewing any differences between the Access database and the METS file.

2. Please describe any selection or film acquisition/evaluation activities during the performance period.

Newspaper title and film selection is complete for the Washington NDNP grant project and the number of scanned images remains at 75% due to the changes in staffing over the summer. No additional reels will be scanned for this grant cycle so we can prioritize current deliverables. The remaining content has been identified as a title chosen for the third grant cycle, the Seattle Post-Intelligencer.

3. Describe any specific award activities that have taken place between collaborating institutions in your state.

As part of the Washington State Library and University of Washington partnership, an NDNP office is maintained in UW's Suzzallo Library for our UW student staff. We were able to secure a larger space in the basement of the building this past fall after the Maps department moved to a different floor. This area is much bigger than the work area we had before and allows more than two people to work at a time. This is important in that we are interested in opening this project opportunity to students at the UW Information School through a non-paid, for-credit internship program known as "Directed Fieldwork" (DFW) or an unpaid, for-credit internship for the undergraduate students in the Informatics program. We are currently working with a student volunteer in addition to these other options as a way to promote newspaper preservation and digitization projects. We've had a positive amount of interest in the program and would like to test it this summer. If we are awarded a third grant cycle we would launch the program solidly in the fall quarter.

Our current UW student staff are a tremendous asset to the NDNP program. Claire Imamura has been with us since the early microfilm evaluation days and will graduate this summer. She is very detailed and production oriented, with a great knowledge of the docWorks tool. April Martin graduated last year and we fear we may lose her to a full time job soon. She is a passionate researcher and librarian and brings great energy and questions about the NDNP program. I hope to find more students like Claire and April through the internship programs and anticipate that we have a pool of potential candidates for promotion of NDNP and if needed, interested library students to fill their shoes if they should find other opportunities in librarianship.

Our UW staff contacts have been very supportive, filling me in on access to UW networks and resources. Glenda Pearson and Jessica Albano are a wealth of information. I have been meeting quarterly with Glenda to keep the UW team updated on the progress of the program. After an exciting discovery that we have an even earlier start date to the Ellensburg Dawn title than previously thought, the UW and WSL cataloging staff are aligning MARC records to reflect the print copies of the 1893 *Reformers' Dawn*, Volumes 1-3, held in UW Special Collections department. April was thrilled when her research revealed the copies she had seen references to and now we have the opportunity to add these to the *Chronicling America* catalog and ensure the records are consistent across agencies, including Library of Congress.

4. Describe specifically your progress in conversion and/or vendor selection (attach documentation as desired).

Our vendor, CCS, has been very good at making themselves available and assisting me where possible in learning the nuances of the docWorks tool. Most of the documentation however is oriented toward the broad scope of materials they provide support for and far less is provided specifically for the NDNP program due to the fact that each of their NDNP customers has a variation in workflow. With this though, they have provided phone support every two weeks, online communication tools for issue tracking and as of September 2011, a U.S. based systems engineer who is in a much nearer time zone.

We found a few problems last summer with the generation of JP2 files that were resolved in a February upgrade, but as with many upgrades, some new problems were introduced. We have been working closely with the CCS support team to resolve the issues and maintain an environment as close to we had before the upgrade. The software is very configurable and sometimes the smallest script change may disrupt the output of files. As mentioned in the answer to question one, the conversion of our metadata from evaluation spreadsheets to a

docWorks accessible database has been a challenge to a fairly straightforward process of export and batch verification. Once we have this resolved with the vendor, most of the system issues we've been experiencing will also be resolved.

5. Do you have any questions or comments regarding the use of the tools provided by LC (Digital Viewer Validation)?

The DVV tool provided by LC has been very easy to use and reliable. I do find the DOS command line less user friendly, but I've had to use both now and the results are consistent with the data in the batch.

6. Provide an update on your planned schedule for delivery of digital assets to the NDNP repository. Include anticipated number of files to be delivered and planned means of delivery or transfer. (Keep in mind the delivery deadlines noted in the Cooperative Agreement Document, but also take into consideration your own resources and workflow plans. LC requests that you plan to deliver new data monthly, with one batch per hard drive. Batches that have been submitted and require rework should be shipped to LC as soon as fixes are complete (send more than one hard drive at a time, if needed)).

The last reported plan to deliver 25% of grant deliverables by August 2011 was unfortunately delayed by several factors but we are now on schedule for a delivery of over 32,000 pages by the end of April 2012. The workflow change and learning curve of the docWorks software program; a significant reduction of WSL staff over the past year; the three month vacancy of the Washington NDNP coordinator; and issues surrounding a docWorks software upgrade have been challenges that we are surmounting. Two WSL staff members working part-time, two University of Washington students, and our two volunteers were able to keep the processing going over the summer and fall while I came up to speed on NDNP requirements and docWorks software beginning full time in late November. The processing team kept working with minimal guidance and built up a high rate of content ready to be exported into batches as soon as I began the program. Several batches, named after Washington rivers, are ready to be exported and allow us to catch up quickly. I resubmitted our first batch, **batch_wa_american**, at the end of December 2011. The second batch, **batch_wa_bumping** has been submitted to Library of Congress and pending approval. We now have two more batches that will be shipped in the next two weeks and then account for approximately 32,000 pages. These batches will include the following titles:

- *Wenatchee Daily World
- *Labor Journal (Everett)
- *The Lynden Tribune
- *The Spokane Press
- *The Vancouver Independent

We will continue toward a goal of another batch of 10,000 pages every two-three weeks to catch up on the previous delays and complete 100,000 pages by the end of the grant cycle. I have been reviewing the steps of the docWorks process and working with CCS in finding ways to streamline processing and make many of the steps easier. These improvements along with the UW student staffing plans mentioned in question three will speed up the overall processing time from import to export and verification.

7. State briefly the status of any other digitized newspaper activities in your or collaborating partners' institutions (non-NDNP-funded).

With the greater visibility of our papers on the Chronicling America website, we have been receiving more phone calls and emails regarding small projects around the state from museums, publishers and archives. The furthest along in a newspaper digitization project is the Spokane Valley Heritage Museum, celebrating their centennial this year, who will be promoting the content in their searchable database. Their questions were based around image standards and ways to collaborate with the state library. They had also contacted the state archives regarding image formats to use and guidelines for their volunteer with a professional camera.

After a Grays Harbor county publisher for local paper, *The Vidette*, contacted us wanting to know how she could use volunteers to have their paper searchable on the Google Online Newspaper Site, I realized it would be beneficial to have some guidelines and standards available in a quick and easy to use format. This activity will focus on image standards and basic metadata that organizations should capture for best preservation and tracking. It will start as a handout and lead to a more detailed document for the level of interest and development of a project. Ideally, these standards will enable us to collate collections and allow easier database migration as collections are developed and shared.

8. Please provide copies of or URLs for any public presentations or publications on topics related to your participation in NDNP (or describe future plans for such).

As we get more titles into the Chronicling America database we are eager to promote usage of the content in broader regions of our state. We are planning two main tours of the eastern side of Washington and the Washington Library Media Association conference to promote usage within our schools. The list below is only a preliminary itinerary which will be solidified in the next few months. I will begin introducing myself and our new Washington NDNP energy to librarians at our Washington Library Association conference next month and begin contacting key libraries across the state.

I will also be partnering with a Washington State Library sister program, Washington Rural Heritage, which focuses on our regional history through images and historic events. WRH often receives questions regarding the digitization of local newspapers and we are already planning on ways we can collaborate through common standards and preservation goals. My message will focus on ways we can encourage future preservation programs in regional organizations and develop high quality standards across the state, based on NDNP guidelines, even after the program ends.

- *Spring Tour, (TBD 2012) Southeast: Walla Walla (Whitman College)
- *Autumn Tour, Northeast: (TBD 2012) Spokane, Colville, Omak
- *NDNP Conference, September 26-28, 2012 Washington, DC
- *WLMA, October 11-13, 2012 in Yakima
- *Everett (Labor Journal) & Puget Sound libraries, informal over several dates

9. Describe any follow-up issues or questions you would like convey to the NDNP program committee.

Bringing the image and metadata processing in-house was an ambitious step that brought with it a host of challenges for staff. Continued state budget problems and staff changes exacerbated them. But the process has also allowed staff to learn so much about the structure of the data and the workflow used by vendors that we believe the experience has been extremely valuable.