



## North Carolina Department of Natural and Cultural Resources

**ROY COOPER, Governor**

**SUSI H. HAMILTON, Secretary**

**TO:** Nancy Melley, National Historical Publications and Records Commission

**FROM:** Camille Tyndall Watson, Project Director, Digital Services

**SUBJECT:** NAR15-RG-50005-15—Transforming Online Mail with Embedded Semantics Collaborative Project

**DATE:** April 30, 2018

The Transforming Online Mail with Embedded Semantics Collaborative Project is a multi-state, multi-part collaborative initiative to identify state positions by function that are responsible for generating or receiving archival email, develop methodologies to move archival email accounts out of their native system, and to develop predictive coding and libraries to assist archivists to process those email accounts and make it available. The project team includes Camille Tyndall Watson, Jeremy Gibson, Jamie Patrick-Burns and Nitin Arora (NC), Elizabeth Perkes (UT), and Megan Rohleder (KS), with staff from the Library of Virginia contributing and offering their insight and experiences. The project also has an advisory committee—Cal Lee, University of North Carolina at Chapel Hill and Chris Prom, University of Indiana at Urbana Champaign.

**Process at least ten email accounts designated as containing permanently valuable correspondence based on Capstone roles.**

Staff continue to work on the export of email accounts from the McCrory administration; the large size of the files has proved challenging for Department of Information Technology (DIT) staff to pull down files from the servers. We will be working to DIT staff to come up with work arounds, and will potentially need to change the workflow for email account pulldowns from Office365's eDiscovery module as a result. We are also working with DIT and Department of Natural and Cultural Resources (DNCR) IT and Administration to get Jamie Patrick-Burns, our digital archivist, access to the eDiscovery module for further development of the workflow.

SANC staff have completed gathering position numbers for the data on Capstone positions collected from a series of three forms distributed to state agency Chief Records Officers (CRO) and Chief Information Officers (CIO). We have sent the list to staff from DIT and the Office of State Controller (OSC) so that they can begin developing and testing the modules to automate the "tagging" of accounts in Microsoft Outlook and BEACON, the state's personnel management tool.

Our storage purchase has been installed and is ready to store the emails that have been pulled down. We have also purchased Amazon services to use for processing space in order to test the feasibility of running the TOMES tool in the cloud.

Over the last six months, we have completed the core functionality of the tool. The three basic functions of the tool are:

- 1) extract mime emails from pst files ([tomes-pst-extractor](#))
- 2) convert mime emails into EAXS xml ([tomes-darcmail](#))
- 3) tag the xml representation of the email with semantics. ([tomes-tagger](#))

The constituent parts of the application are also hosted in its docker form as a series of public repositories on Docker Hub: <https://hub.docker.com/u/govsanc/dashboard/>.

As we move from the development phase to the deliverable phase, we are testing all aspects of the code, documenting code and testing deployment variables. Jeremy has been working on structuring the tool into its final deliverable form as a dockerized application. During the past six months Jeremy has also been testing and developing launch scripts on all modern deployment platforms: Windows 7, Windows 10, Linux, and MacOS.

The current tagging-related toolset is hosted on our GitHub account:  
<https://github.com/StateArchivesOfNorthCarolina/tomes-tagger>

**Publish findings, tools, and training pieces on a TOMES project website, through the CoSA Program for Electronic Records, Training, Tools and Standards (PERTTS) portal, social media platforms e.g. Twitter**

Camille presented the progress of the TOMES project at the Best Practices Exchange (BPE) Conference in Boston, MA in November 2017. The presentation, discussed the project's history, the process of identifying permanent email accounts, collaboration with DIT and OSHR to automate the tagging of identified "archival" positions, and progress on the processing tool. She also discussed final, expected deliverables for the grant, as well as the benefits and drawbacks of a Capstone approach and challenges we have encountered in developing the tool. She also introduced our concept of iterative processing of email accounts based on access demand.

In January, Camille was brought on as a member of the Mellon Foundation Email Task Force, and travelled to New York to attend the final review of the Task Force's report. The meeting was very useful, and the members, which included representatives from Stanford/ePADD, Harvard, and the Digital Preservation Coalition, were very interested in the TOMES project and the specific needs of state government archives when it comes to email. The Task Force put together a presentation proposal for SAA in Washington DC in August on 2018 that included Camille and Jeremy; the proposal was accepted and they are working on their presentation. Camille has also been included on Task Force proposal to facilitate a workshop at iPres 2018 in Boston regarding email archiving. Additionally, Camille will likely be travelling to New York again in the fall for a follow up meeting of the Task Force to discuss open source email preservation and processing tools.

In March, Camille and Jeremy presented on TOMES for a NAGARA webinar. The webinar discussed similar themes as the BPE presentation, but was focused more strongly on the tool and its functionality, and included a live demonstration of the tool. The feedback from the presentation was excellent, and NAGARA staff informed us that there was a higher than average participant retention rate for their webinars

### **Future Work**

Coding work on the TOMES tool has been completed, with all features having been built and debugged. Work for the last phase will focus on testing and component validation.

We will be spending the last part of the grant creating training and documentation for both internal, NC state government audiences and external audiences, including videos and software documentation. We will also be finalizing the workflow for the pull-down of accounts from eDiscovery, and working with DIT to test automated tagging of Capstone accounts. All documentation and training will be made available on the TOMES website and in the PERTTS portal when they are completed.

Additionally, Camille and Jeremy will be presenting on TOMES at SAA in August 2018, and Camille will potentially be attending Mellon Foundation Email Task force events in September, including facilitating an iPres workshop in Boston.