Transforming Online Mail with Embedded Semantics

CAMILLE TYNDALL WATSON
DIGITAL ARCHIVIST
STATE ARCHIVES OF NC
Where We’ve Been...

- EMCAP, 2007-2009
- Email Records Review Panel, 2009
- Executive Orders 150 and 18, 2009
- Mimosa Nearpoint, 2009-2014
Where We Are…

- Email retention reduced to 5 years
- State IT has become centralized
- In the process of moving to Office365
- Facing the challenges of a move to cloud storage for email
Where we’re going now...
Transforming Online Mail with Embedded Semantics (TOMES)

- NHPRC State Government Electronic Records Grant
- 2015 – 2017
- Partnership between State Archives of NC, Utah State Archives, and Kansas State Historical Society
- Advisory group includes Cal Lee (UNC-Chapel Hill), Chris Prom (University of Illinois Urbana Champaign), and staff from the Library of VA
What should we keep?
Building Relationships

- Department of Information Technology
  - “Tagging” accounts by function
- Government Records Section, SANC
  - Reaching out to State Agencies’ CIOs through records analysts to understand organizational structures
- Office of State Human Resources
  - Include in “tagging” of accounts
The Blazed Trail

- EMCAP
- ePADD
- Stanford NER
- GATE
- Minezy
The Product, After the Process

- Development of State Government specific NLP libraries for use in the processing of email accounts containing public records
- An MPLP approach to the arrangement and description of email
- The ability to identify materials that should be reviewed for PII before release to public
Questions?

Camille Tyndall Watson
Digital Archivist
State Archives of NC
Camille.tyndallwatson@ncdcr.gov
(919) 807 - 7359