

# Social Networks and Archival Context Project (SNAC)

EAC-CPF: Moving Forward with Authority

National Archives

Washington, D.C.

9 August 2010

# SNAC Overview

- Funding and Timeline
- Project Team
- Data Contributing Institutions
- Project Objectives and Rationale
- Methods, Processing, and Products
- Early Results and Demonstration
- Final Thoughts
- Questions?

# Funding and Timeline

- National Endowment for the Humanities
- A Preservation and Access, Research and Development grant
- Two-year project
- May 2010-April 2012

# Project Team

- Daniel Pitti (PI) and Worthy Martin (Institute for Advanced Technology in the Humanities, University of Virginia)
- Adrian Turner and Brian Tingle (California Digital Library, University of California)
- Ray Larson (School of Information, University of California, Berkeley)

# Data Contributing Institutions

- EAD-encoded finding aids
  - Library of Congress (918+)
  - Online Archive of California (13,932+)
  - Northwest Digital Archive (5,160+)
  - Virginia Heritage (8,390+)
- Authority records
  - Library of Congress: NACO/LCNAF (3.8M personal names; 900K corporate names)
  - Getty Vocabulary Program: Union List of Artist Names (293K personal and corporate names)
  - Virtual International Authority File (intersection with NACO/LCNAF)

# Project Objectives

- Further the ongoing process of transforming archival description using advanced technologies
- By facilitating the separation of the description of people from the description of records
- Using EAC-CPF
- Goal: enhance the economy and effectiveness of archival description to enhance access and understanding for the users of archives, libraries, and museums

# Rationale for Separation

- Authority control of forms of names
- Flexible description
- Cooperative authority control
- Integrated access to cultural heritage
- Biographical/historical resource
- Social/historical context (social-professional networks)

# Methods and Processing

- Extract EAC-CPF records from existing EAD-encoded archival descriptions
  - Extracting both creators and referenced CPF names
- Match EAC-CPF records against one another and against existing authority records (ULAN, VIAF, LCNAF)
  - Enhance EAC-CPF by normalizing entries, adding alternative entries, titles (VIAF), and historical data (ULAN)
- Create a prototype historical resource and access system
  - Historical data and social-professional networks
  - Links to archive, library, and museum resources (by and about)



# Open Source Products

- Software used in the deriving of EAC-CPF records from EAD-encoded finding aids will be made available (XSLT)
- Software used in the matching, merging, and interrelating of EAC-CPF authority records and ULAN, VIAF, and LCNAF authority records will be made available (variety of code)
- Software for prototype historical resource and access system will be made available (XTF)

# Early Results and Demonstration

- Extracting of EAC-CPF records underway
- Matching-merging preliminary work nearly completed and processing will begin soon
- Preliminary resource and access system (primary focus on project use, for now)
- Library of Congress finding aids processing completed
  - 16,215 non-unique EAC-CPF records derived from 918 finding aids (17/finding aid)
  - Approximately 10,290 unique records after match-merge processing (11/finding aid)
- Brief Demonstration

# Final Thoughts

- For more information
  - <http://socialarchive.iath.virginia.edu/>
- Next steps
  - OAC finding aids EAC-CPF extraction
  - Match and merge processing (EAC-CPF to EAC-CPF and EAC-CPF to ULAN, LCNAF, and VIAF)
  - First public release of the historical resource and access system: late fall 2010

# Final Thoughts

- Pending IMLS proposal
- Social Networks and Archival Authority Records (SoNAAR)
- If funded, SoNAAR will expand on SNAC
  - 1 million OCLC WorldCat MARC archival descriptions (~400K with biographical/history notes)
  - National Archives and Records Administration (70,000 authority records)
  - 77,000 finding aids from 7 consortia and 17 repositories
  - Additional processing on timelines and social networks
  - Columbia University will be a fourth partner

# Questions?