Social Networks and Archival Context

Fonds and Bonds

Austin

2014

Daniel V. Pitti - IATH, University of Virginia
Overview of SNAC

• Funding: NEH (2010-12), IMLS (2011-2014) and Mellon (2012-15)

• Partners
  – U.S. National Archives and Records Administration
  – University of Virginia
  – University of California, Berkeley
  – California Digital Library (University of California)

• Two complementary activity areas
  – Research and Demonstration
  – Cooperative Planning
R&D Objectives

• Demonstrate that data describing people in existing archival description can be used ...
  – To address the challenge of finding/discovering/locating/understanding distributed historical resources and ...
  – To lay the foundation for an international cooperative for centrally maintaining the collectively created EAC-CPF records
R&D Data Sources

- 2.2M WorldCat archival descriptions (MARC21)
- Nearly 190,000 EAD-encoded finding aids, primarily from US and UK, though some French
- 300,000 British Library authority records
- NARA authority records
- Agency descriptions from Smithsonian Institution Archives/New York State Archives
- And more …
Methods and Processing

• Extract/Assemble/Migrate EAC-CPF records from existing archival description
  – Extracting both creators and referenced CPF names
• Match EAC-CPF records against one another and against existing authority records (VIAF)
  – Enhance EAC-CPF by normalizing entries, adding alternative entries, titles (VIAF)
• Create a prototype historical resource and access system
  – Historical data and social-professional networks
  – Links to archive, library, and museum resources (by and about)
The Match/Merge Challenge

• Different names for the same person
• Different people with the same names
• A challenge for computers!!
• A challenge for people!!
Current Results – WorldCat + BL

- Original Source Records: 4,846,001
  - Person: 3,196,499
  - Corporate body: 1,480,638
  - Families: 168,864

- Merged Records: 2,632,320
  - Person: 1,568,481
  - Corporate body: 894,975
  - Families: 168,864

- 2M++ EAC-CPF between now and end of R&D
Prototype Historical Research Tool

• Quick Demonstration
  • http://socialarchive.iath.virginia.edu/
Rationale for Cooperative

• For archivists
• Economy of sharing our descriptive work
  – People are social beings and records document this fact
  – Same person is the creator of records and referenced in the records of others
  – Describe the person once and well; share the description
• Link the people descriptions to one another and to record descriptions …
• An international, Internet-based “linked archival authority system”
Rationale for Cooperative

• For research users
• Integrated access to distributed historical resources
• Expand the context and understanding of the records
  – Access to the social/professional/intellectual networks within which people lived and worked
SNAC Cooperative Identities

Human editors: evaluate, verify, add new evidence & create, edit, link …

Smart people

Sources: archives, libraries, museums, scholarly research projects …

MARC21, EAD, TEI, Local formats

Smart algorithms

EAC-CPF

Sparse

dense

certain

uncertain

evidence
SNAC Cooperative and ISNI

• Processing models similar: smart algorithms and smart people
• But differences
  – Domain: cross-domain versus archival domain
  – Quality of the source data: authority files versus record description, for now (quality very uneven)
• SNAC Cooperative aspires to become part of the ISNI network, some day
• “Collaboratively consolidating identities at universal scale”
Cooperative: Next Steps

• Final stages of planning
• Cooperative host: U.S. National Archives and Records Administration (NARA)
  – Administration and governance
• Technology infrastructure host: California Digital Library
• Launch of pilot cooperative in 2015