Social Networks and Archival Context

From Project to Program

CNI
Seattle
2015

Daniel V. Pitti - IATH, University of Virginia
Preface

• Context and Overview: Social Networks and Archival Context (SNAC) [DP]
• Research &Demonstration (DP)
• History Research Tool (BT)
• Cooperative (DP)
Archival Records

• Records are the by-products of people living and working as individuals, in organized groups, in families
• Records document people living and working
• People exist in social-professional-intellectual contexts, in relation to one another
• Records document these relations
Description of Archival Records

• Archivist describe *Records in Context*: Principle of Provenance
  • Records from one source kept together
  • “Original order”
  • Context includes the social context

• Archival descriptive practice has been dominated by using one provenance based descriptive apparatus in which context description and record description are intermixed in which
  – Creators are documented in detail
  – Many of the people documented in the records are also referenced in description

• Archival description documents interrelations among people and records

• A vast social-document network connecting the past to the present to the future

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 IATH
  – University of California, Berkeley SI
  – 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 biographical data
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 (archival standard) 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

• Original Source EAC-CPF Records: 6,290,472
  – Person: 4,337,808
  – Corporate body: 1,760,046
  – Family: 192,618

• Merged EAC-CPF Records: 3,471,304
  – Person: 2,252,413
  – Corporate body: 1,026,275
  – Families: 192,616

• Pending EAC-CPF Records: 295,128
  – Person: 219,769
  – Corporate body: 72,865
  – Family: 2494

• And a few more, for example, 2000+ from the Archives nationales, France
Over to Brian
SNAC Cooperative
Rationale for Cooperative

• For archivists
• Economy of sharing 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: build the social-document network
• 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
Human editors:
evaluate, verify,
add new
evidence &
create, edit, link ...

Smart people

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

Smart algorithms

MARC21, EAD, TEI,
Local formats

EAC-CPF
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

- Proposal to the Mellon Foundation underway
- Cooperative host: U.S. National Archives and Records Administration (NARA)
  - Administration and governance
- Technology infrastructure host: IATH, University of Virginia
- Launch of pilot cooperative July 2015
Inaugural Members of the Cooperative

- American Institute of Physics
- American Museum of Natural History
- George Washington University
- Getty Research Institute
- Harvard University
- Library of Congress

- National Archives and Records Administration
- New York Public Library
- Princeton University
- Smithsonian Institution Tufts University
- University of California, Irvine
- University of Miami
- Yale University
Some of the Key People

• Director: Laura Campbell (retired Associate Librarian CIO of the Library of Congress)
• Deputy director: John Martinez, NARA (administration lead)
• Jerry Simmons, NARA (governance lead)
• Worthy Martin, IATH (technology lead)
• And a host of others
Closing Thoughts

• Exceptionally complex: socially, intellectually, and technologically
• The social is the most challenging and complex: international community building
• Aspires to do what had never been done:
  • Build a cooperative international, Internet-based “linked archival authority system”
    – Build a vast social-document network connecting the past to the present to the future
• Contribute a significant component to the international humanities research infrastructure

Daniel V. Pitti - IATH, University of Virginia