SOCIAL NETWORKS AND ARCHIVAL CONTEXT

ICA /ALA Meeting. Mexico City, 2017

Daniel Pitti, University of Virginia | Kelly Spring & Jimmy Zavala, University of California, Irvine
Session Overview

- Description and history of SNAC Cooperative (Daniel)
- Governance (Kelly)
- History Research Tool (Jimmy)
- Editing User Interface (Kelly)
SNAC Cooperative
Archival Records

• Record: Linguistic, symbolic, or graphic information represented in any persistent form, on any durable carrier, by any method, by an Agent in the course of life or work events and Activities.

• Records *document* (are evidence of) people living and working together

• The lives and work of people, and the records created by them constitute a vast network of interrelated people and documents: a social document network
Description of Archival Records

- Archival descriptive dominated by using one provenance-based apparatus that combines/intermixes context description and record description: the finding aid
  - Creators are documented in detail
  - Many individuals and groups documented in the records are also referenced in description (as access points or informally)
- **Archival description documents interrelations among people and records**
- A vast social-document network connecting the past to the present to the future
- To date, interrelations have been *implicitly* rather than *explicitly* documented
- SNAC makes the social and document interrelations explicit
Rationale for Cooperative

• The corporate body, person, or family (CPF) document in the holdings of one repository is often documented in the holdings of another repository
• Research, identify, and describe the CPF entity once, and shared by all: economy of cooperatively curating the data
• Link the people descriptions to one another and to record descriptions: build the social-document network
• An international, Internet-based “linked archival authority system”
• Research economy
  – Integrated access to distributed historical resources
  – Expand the context and understanding of the records by revealing the social networks
Cooperative

- Cooperative host: University of Virginia Library
  - Administration and governance coordination
  - Technology infrastructure
  - Cooperative members ...
<table>
<thead>
<tr>
<th>Cooperative Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Institute of Physics</td>
</tr>
<tr>
<td>American Museum of Natural History</td>
</tr>
<tr>
<td>Archives, National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore, India</td>
</tr>
<tr>
<td>Archives nationales de France</td>
</tr>
<tr>
<td>Brigham Young University</td>
</tr>
<tr>
<td>California Digital Library</td>
</tr>
<tr>
<td>Cecilia Preston (individual scholar)</td>
</tr>
<tr>
<td>George Washington University</td>
</tr>
<tr>
<td>Getty Research Institute</td>
</tr>
<tr>
<td>Harvard University</td>
</tr>
<tr>
<td>Indiana University–Purdue University Indianapolis</td>
</tr>
<tr>
<td>Jane Addams Papers (documentary editing)</td>
</tr>
<tr>
<td>Library of Congress</td>
</tr>
<tr>
<td>Mojave Desert Archives</td>
</tr>
<tr>
<td>U.S. National Archives and Records Administration</td>
</tr>
<tr>
<td>New York Public Library</td>
</tr>
<tr>
<td>Princeton University</td>
</tr>
<tr>
<td>Smith College</td>
</tr>
<tr>
<td>Smithsonian Institution</td>
</tr>
<tr>
<td>Tufts University</td>
</tr>
<tr>
<td>University of California, Irvine</td>
</tr>
<tr>
<td>University of Miami</td>
</tr>
<tr>
<td>University of Nebraska</td>
</tr>
<tr>
<td>– Library</td>
</tr>
<tr>
<td>– Walt Whitman Archive (documentary editing)</td>
</tr>
<tr>
<td>University of North Carolina, Chapel Hill</td>
</tr>
<tr>
<td>University of Oregon</td>
</tr>
<tr>
<td>University of Virginia</td>
</tr>
<tr>
<td>Utah State Archives</td>
</tr>
<tr>
<td>Yale University</td>
</tr>
</tbody>
</table>
History: Research & Demonstration 2010-2015

- From R&D to Cooperative Program
- R&D Partners
  - University of Virginia, Institute for Advanced Technology in the Humanities
  - University of California, Berkeley School of Information
  - California Digital Library (University of California)
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 ...
  – Reveal the social-document network implicit in the description and ...
  – Lay the foundation for an international cooperative for centrally maintaining the biographical/historical descriptions
Data Sources

- 2.25M 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
- U.S. National Archives and Records Administration 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 (Virtual International Authority File (VIAF))
  - Enhance EAC-CPF by normalizing entries, adding alternative entries, titles, “same as” links (VIAF)
- Create a prototype historical resource and access system
  - Social networks in which people lived and worked
  - Integrated access to distributed archival resources
  - Access to other resources by and about (publications, other artifacts)
The Identity Resolution Challenge

- Different names for the same person
- Different people with the same names
- Names are weak identifiers
- A challenge for computers
- A challenge for people

- Identity must be based on evidence and as much as possible: names, dates, places ...
Extraction Results

• Original Source Records: 6,719,064
  – 4,653,365 Persons
  – 1,868,448 Corporate Bodies
  – 197,251 Families

• Merged Records: 3,741,262
  – 2,466,425 persons
  – 1,077,588 corporate bodies
  – 197,249 families
Establishing the Cooperative: 2015-2019

- Funding: Mellon Foundation 2015-2019
- Two primary objectives
  - Social
    - Administration
    - Community governance
    - A growing team of trained editors
    - Business model
  - Technological
    - Transform R&D platform into maintenance platform
    - An editing interface for manually creating, revising, linking descriptions
    - A History Research Tool for researchers
    - Collaborative Ingest Tool for importing large sets of data
Current Focus of Work

• Developing capacity to increase membership
• Training editors, training trainers of editors:
  – Real and virtual classroom
  – Self-paced online modules
• Collaborative Ingest Tools
  – Enable new data contributors to assist in improving the quality of the data before import
  – Enable new contributors to assist in Identity Resolution
  – Enable new data contributors to integrate persistent identifiers into local description of archival records
Going Forward: 2017-2019

- Continue both the social and technological development, including making both the editing and research interfaces available in other languages
- As we build capacity, recruit new members
- Increase members of all kinds, including international members
SNAC Cooperative Identities

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

Smart people

MARC21, EAD, TEI, Local formats

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

Smart algorithms

EAC-CPF
The SNAC Cooperative Community
The SNAC Cooperative Community

Principles

- Voluntary & open
- Democratic control
- Economic participation
- Autonomy & independence
- Education, training, information
- Cooperation with cooperatives
- Concern for community
The SNAC Community Structure

Operations Committee

Technology Infrastructure
Editorial Policy & Standards
SNACSchool
Communications
Technology Infrastructure
Technology Infrastructure Highlights

- History Research Tool (HRT) transition
- HRT Advanced Search Feature
- User Interface development for maintenance system
History Research Tool transition

Social Networks and Archival Context

SNAC History Research Tool early prototype
Phase II Goals

- Feedback and development
- Batch data refinement
- SNAC API integration
Technology Infrastructure Phase II Projects
Editorial Policy & Standards
Editorial Policy & Standards Highlights

- Defined roles and permissions
- Developed editorial practice
- Determined ownership policy
Constellation Ownership Policy
Phase II Goals

- Demographic classification
- Function & Occupation policy
- Geographic names
- Subject terms
Editorial Policy & Standards Projects Include

- Alphabetical Index
- Metadata
- FAQs
- SCM
- Master List
- Training
SNACSchool
SNACSchool Highlights

● Developed training program
● Created SNACSchool Lite
● Hosted live training events
Phase II Goals

● Enhance training material
● Effective training methods
● More training!

SNACSchool
SNACSchool Projects Include
Communications
Communications Highlights

- Established regular information exchange
- Evaluated existing channels of internal communication
- SNAC Cooperative central portal
Cooperative Members Portal

This is a member-only website for the use of the SNAC Cooperative. It is managed by the University of Virginia Library and the SNAC Cooperative Program.

Please click on any section of the menu to view available information about the Cooperative.
Phase II Goals

● Develop the “SNAC brand”
● Cooperative portal content
● Outreach efforts
Communications Projects Include
History Research Tool Demonstration
SNAC R&D Technical Highlight: Front-end history research tool
Social Networks & Archival Context
Paz, Octavio, 1914-1998

Biographical notes:

Related names in SNAC
- People
- Families
- Organizations

Links to collections
- Archival Collections
- Related Resources
- Related External Links

Constellation Options
- Collector Locations
- Detailed View
- Similarity Assertions
- Constellation History
- Source Summary

Visualizations
- Connection Graph
- Radial Graph
- Constellation JSON

Constellation Information
- Permalink:
  - http://521.net/art/991680w90x114r1
- Ask @:
  - v000tr1
- SNAC Id:
  - 17230-00
Guillén, Jorge, 1893-1984

Dates:
Birth 1893-01-18
Death 1984-02-06

Nationality: ES
Language: Spanish; Castilian

Biographical notes:

Jorge Guillén (1936-1984) was a Spanish lyric poet. He went into exile in 1938 and settled in the United States in 1940.


His opposition to Franco during the Spanish Civil War led him into self-imposed exile. While in the US, he taught at Middlebury, Wellesley, Harvard, Yale, Berkeley, and Ohio State. He died in 1984 at age 90 in Naples, Italy, and was survived by his second wife and his children.

Related names in SNAC:

People
Families
Organizations

Related collections:

Archival Collections
Related Resources
Related External Links

Constellation Options

Information
Collection Locations
Detailed View
Simularity Assertions
Constellation History
Source Summary

Visualizations
Connection Graph
Radial Graph

Export As
Constellation JSON
EAC-CPF XML

Constellation Information

Permalink:
http://n2t.net/ark:/99166/w63n2s8m
Ark ID:
w63n2s8m
SNAC ID:
44215511

Subjects:
Editing Interface Demonstration
Social Networks and Archival Context

Eliot, T. L. (Thomas Lamb), 1841-1936

Bigot de Morogues, Marie, 1786-1820

Ehrenberg, Herman, 1816-1866

Brooks, Alfred H. (Alfred Hulse), 1871-1924
Welcome, Kelly, this is your dashboard. The left column shows the most recently published Constellations across SNAC and provides searching. The second column shows Constellations you have checked out for editing and those you are currently editing in other browser windows. It also provides a button to create a new Constellation object. The last column shows constellations you have out for review and other information depending on your permission level.

- **Create New Constellation**
  - Use the button below to create a new constellation.
  - [Create New Constellation]

- **Locked for Review**
  - You have the following Constellations that have been sent for review. You may unlock them here to continue editing them.

- **Locked Edits**
  - You have the following Constellations open in a browser for editing. You may unlock them by clicking below, however, your changes are not automatically saved.

- **View Recently Published**
  - Browse interesting and recently published Constellations from all of SNAC.
  - Winema National Forest (Agency: U.S.)
  - Great Britain, Royal Air Force
  - United States. National Advisory Commission on Civil Disorders
  - Voice of America (Organization)
  - Bahá’u’lláh, 1817-1892
  - Hendrix, Jim, 1942-1970

- **In Progress Edits**
  - Continue where you left off on one of the following Constellations.
  - [No Constellations Available]
Welcome, Kelly, this is your dashboard. The left column shows the most recently published Constellations across SNAC and provides searching. The second column shows Constellations you have checked out for editing and those you are currently editing in other browser windows. It also provides a button to create a new Constellation object. The last column shows constellations you have out for review and other information depending on your permission level.

Quick Search Results

- Paz, Abel
- Paz, Octavio, 1914-1998
- Paz-Soldán, Carlos Enrique, 1885-
- Paz Remolar, Ramón.
- Trio da Paz
- Paz, Ireneo, 1836-1924
- Paz, Suni
- Paz Soldán, Mariano Felipe, 1821-1886
- Paz Estenssoro, Victor, 1907-2001

Create New Constellation

Use the button below to create a new constellation

Locked for Review

You have the following Constellations that have been sent for review. You may unlock them here to continue editing them.

In Progress Edits

Continue where you left off on one of the following Constellations.

Locked Edits

You have the following Constellations open in a different browser window.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ArchivalResource</td>
<td>[30] h. 28 x 22 cm o menores.</td>
</tr>
<tr>
<td>http</td>
<td><a href="http://www.worldcat.org/oclc/430603058">http://www.worldcat.org/oclc/430603058</a> [View]</td>
</tr>
<tr>
<td>Relation Entry</td>
<td>Guillén, Jorge 1893-1984. [Miscelánea sobre Octavio</td>
</tr>
<tr>
<td>Role</td>
<td>referencedIn</td>
</tr>
<tr>
<td>Descriptive Note</td>
<td></td>
</tr>
</tbody>
</table>
Functions

You may perform any of the following actions on this constellation.

- Save & Continue
- Save & Dashboard
Sources

Display Name: Source 56: EAD from
http://oasis.lib.harvard.edu//oasis/deliver/deepLink?_collection=oasis&uniqueId=hou00077

URI: http://oasis.lib.harvard.edu//oasis/deliver/deepLink?_collection=oasis&uniqueId=hou00077
You may perform any of the following actions on this constellation:

1. Save & Continue
2. Save & Dashboard
3. Save & Send for Review

You may write an optional saving message here...
Social Networks and Archival Context

- Eliot, T. L. (Thomas Lamb), 1841-1936
- Bigot de Morogues, Marie, 1786-1820
- Ehrenberg, Herman, 1816-1866
- Brooks, Alfred H. (Alfred Hulse), 1871-1924
Session Review

- Description and history of SNAC Cooperative
- Governance
- History Research Tool
- Editing User Interface

Thank You!

snaccooperative.org