Having a Nosh with SNAC
Dina Herbert
National Archives and Records Administration
Association of Jewish Libraries, June 18, 2018

Introduction
<slide 2>
What is SNAC? SNAC is an acronym standing for Social Networks and Archival Context. It is not a social network like Facebook, Instagram, or LinkedIn, although staff affectionelty call it the Facebook for Dead People. Rather, it demonstrates the separation of the description of corporate bodies (organizations), persons, and families, along with their socio-historical contexts, from the description of the historical resources that are the primary evidence of their lives and work. Essentially, SNAC allows researchers to trace the human experience in a new and robust way: following people and groups and then moving out to resources created by, about, and related to them. The SNAC framework virtually connects social networks, branches of family trees, and ever-evolving organizations. Therefore, SNAC makes explicit what has been usually implied in archival description: the social, professional, and intellectual networks within which the lives and work of the people documented in historical resources took place.

A key objective of SNAC is providing users (such as researchers, librarians, archivists, and historians) with convenient, integrated access to historical collections held by multiple private and public archives and libraries around the world. This effort also sets the stage for a cooperative program for maintaining information about the people documented in the collections. The descriptions in SNAC help resolve the challenge of an archival diaspora by offering search and discovery through one central system of description. This has the possibility to drastically change the way research about historical figures and events occurs.

<slide 3>
Before we go further, let's define basic concepts used in this presentation. First, *description and cataloging*. This is the application of standard content rules to devise a *finding aid*. For many who are unfamiliar with archival processes, the finding aid serves as a surrogate which describes the physical and subject content of any type of cultural heritage collections. These collections are not cataloged in a traditional library classification system since the materials are usually unpublished. The finding aid will serve as the basis for all information about the collection, from background about the creators to access points and physical control.

Authority control is the process of establishing the preferred form of a heading, such as a proper name or subject, for use in a catalog, and ensuring that all catalog records use such headings. Once established, the form is usually recorded in an *authority file* for future reference, along with cross-references from other forms of the heading, to ensure consistency.

Name authorities and authority files establish preferred or authorized forms of proper names (person, corporate body, and family names), then file and store the information about the names

in systems for indexing and retrieval. Many of us are likely familiar with the Name Authority Cooperative Program, or NACO, based out of the Library of Congress.

These combined activities create an atmosphere for consistent access points in information systems, with more robust retrieval during searching.

For those of us with a more traditional library, rather than an archival, background, authority control is similar to cataloging. One cannot create a proper catalog and access to a physical book without having a standard for authority control of names or subjects.

SNAC: Research and Development
<slide 4>
The vision for SNAC began in its research and development phase in 2010 and depended on two description standards:

> 1) *Encoded Archival Description (EAD):* an encoding standard for machine-readable finding-aids developed in the late 1990s, and,
> 2) *Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF)*: a standard for encoding and exchanging authoritative information about the context of archival materials, especially the creators.

The archival community needed to develop and adopt EAD and EAC-CPF. Both coding standards facilitate the exchange of authorized descriptive data, and together create a new environment of linked but separated description. SNAC leverages both EAD and EAC-CPF to create a network of contextual description in one searchable resource.

SNAC History
<slide 5>
In 2013 the project moved forward in its realization for an international archival description cooperative. The main players were the leaders in the field:

> Daniel Pitti (SNAC R&D and Cooperative Planning Project Director),
> Laura Campbell (retired Chief Information Officer, Library of Congress),
> Anne Van Camp (Director of the Smithsonian Institution Archives),
> Don Waters (Program Officer, Andrew W. Mellon Foundation),
> Clifford Lynch (Director, Coalition for Networked Information),
> David Ferriero (Archivists of the United States),
> Pamela Wright (Chief Office of Innovation, National Archives and Records Administration),
> John Martinez (Office of Innovation, National Archives and Records Administration)

The United States National Archives and Records Administration (NARA) committed to host the cooperative. (However, due to several factors, NARA is no longer the primary host but rather the

federal government lead and the University of Virginia Libraries now hosts the cooperative.) The California Digital Library (CDL), part of the University of California, committed to host the technological infrastructure, to be developed and maintained by the cooperative community as an open source undertaking.

Then, SNAC ingested the bulk of the data available in WorldCat and a substantial portion of the data available in EAD-encoded finding aids in the U.S. and the U.K. There were 2.25 million WorldCat records, nearly 200,000 EAD-encoded finding aids, 300,000 British Library authority records, and so much more. Remember, this still represents only a fraction of the data available globally. Thus, the Cooperative continues to employ batch processing and ingest methods to collect large sets of data as they are made available.

Rationale for SNAC
<slide 6>
It was evident that there is a usefulness in developing a resource discovery tool that was able to create archival authority records, using machines to extract the CPF entities found in EAD finding aids. That served as a linking device for all the collections that referenced such entities. The SNAC project began as an investigation into the possibility of creating such a tool.

Background
<slide 7>
Stepping back, we also must address how archivists arrange and describe their materials. Traditionally, archivists keep the description of creators with the descriptions of their papers and collections. This example from the National Archives Catalog shows the Person Authority Record for Abraham Joshua Heschel and a screen shot of the three records he is referenced in held by NARA.

<slide 8>
With SNAC, we can still see the collections linked within the record and it will still take you to the finding aid at the holding institution.

<slide 9>
Not only does SNAC take us to the original collections of the entity, but also to the archival collections of others where Heschel has contributed. In these instances, SNAC treats the "entity" as a subject of other archival, library, and museum collections; they are separate from the created entities. It also further allows the collections to create more connections of who corresponded with whom and so forth.

SNAC Records
<slide 10>
Each SNAC record contains a section dedicated for display of links between the SNAC description and the finding aids describing the archival materials.

The SNAC Record:
- Enables sharing, discovery, and display of standardized information in an electronic environment
- Facilitates discovery of contextual relationships amongst record-creating entities
- Facilitates linking descriptions of creators to descriptions of records and other entities in context
- Other metadata: geography, subjects, occupations, dates, languages, etc.

Among SNAC's strongest features is its ability to link between entities based on contextual relationships using EAC-CPF relationship types. This example shows the relationship links between Henrietta Szold and Hadassah, designated by the <founderOf> and <foundedBy> relationship type.

<slide 12>

Exploiting the linking power of EAC-CPF, we can search and locate one entity, then travel to other entity descriptions knowing the context of relationships. In this example, Jacob Rader Marcus's and Irving Berlin's relationship was in the context of their correspondence. And from the links to collections, we can easily find descriptions of resources by and about them, no matter where they are physically located in a brick and mortar archive.

<slide 13>

As information specialists, we strive to offer new ways to view, understand, and interpret relationships. SNAC generates connection and radial graphs for each and every person, family, and organization in the research portal. Here, this visualization from Yeshiva University's network provides another path for exploration and discovery.

Where is SNAC Now and Where are We Headed?
<slide 14>

Currently, SNAC is about half way through a second grant cycle funded by the Mellon Foundation. The University of Virginia Libraries heads up the administration of the cooperative. The full and part time staff at SNAC include Principle Investigator Daniel Pitti (UVA), Deputy Director Ivey Glendon (UVA), Developers Robbie Hott and Joseph Glass (UVA/IATH), NARA Liaisons Jerry Simmons and Dina Herbert, and retired Library of Congress CIO, Laura Campbell.

We meet annually as a cooperative and there are several committees to ensure the smooth operations as we grow. Committees include operations, editorial standards, technical standards, communications, and training. We offer SNACSchool remotely and in person for new editors and have trained over sixty people in the cooperative.

As we grow, sustainability in on our mind. Staff have been actively working on creating a business model that will enable the success of the cooperative without having to rely solely on outside funding. We are also hoping to garner more international groups to join the cooperative. Lastly, we are working with ArchivesSpace and Family Search to integrate their resources into our systems so that we can more easily ingest new resources and enable genealogists to get the most out of SNAC.

The technology continues to be updated. We plan a complete redesign of the research and education tool to provide a better and more up-to-date version of the public interface. The new interface is intended to expose more of the data in the records to users in an intuitive and aesthetically pleasing manner, as well as provide the means to access sources and metadata for the assertions made in SNAC.

<slide 15>
As SNAC grows please continue to follow us. We hope the next stage of our development will prove useful to different types of researchers and institutions.