

SNAC (Social Networks and Archival Context) History

PRE-SNAC DEVELOPMENTS

The envisioning of the Social Networks Archival Context (SNAC) project, that began its R&D phase in 2010, depended on the creation of two other standards for use in the archival world that developed in the preceding 20 years. These include Encoded Archival Description (EAD) an encoding standard for machine-readable finding-aids developed in the 1990's, and Encoded Archival Context-Corporate Bodies, Persons, and Families (EAC-CPF) a standard for encoding and exchanging authoritative information about the context of archival materials. With the latter's release in March of 2010, the building of linked archival description systems allowing for the separation and independent maintenance of the corporate body, person, and family contextual data from the archival description of records became possible.

PROJECT RATIONALE

At this point, the usefulness of developing a resource discovery tool that was able to create archival authority (identity) records using machines to extract the CPF entities¹ mentioned in archival collections from EAD finding aids, and that served as links for all the collections that referenced such entities became evident. The SNAC project began as an investigation into the possibility of creating this tool.

SNAC COMPONENTS

Early successes with the technical aspects (data extraction and archival authority creation) of SNAC indicated the necessity of creating an international cooperative to serve as an administrative body in order to realize the full potential of this resource as a global archival description system. A history of SNAC therefore must describe both the technical and administrative elements that followed the initial R&D phase.

DEVELOPMENT

The Institute for Advanced Technology in the Humanities (IATH) at the University of Virginia, the California Digital Library (CDL), and the School for Information Science at the University of California, Berkeley were the initial partners involved in the research and development of SNAC. Beginning in 2010 with a two-year grant from the National Endowment for the Humanities (NEH) their work focused on technical aspects of archival data extraction and manipulation. The CDL rollout in 2010 of the first SNAC Prototype History Research Tool (PHRT) revealed some basic functionalities. The first version of PHRT displayed an authorized archival identity that served as a link to the archival resources regardless of the

¹ An entity is defined as the physical embodiment of the unit being described (the real person), whereas identity is defined as a single representation of that entity. Entities can have multiple real or imagined identities and identities can be shared by multiple entities. Cited from pg. 172: Katherine M. Wisser "Describing Entities and Identities: The Development and Structure of Encoded Archival Context—Corporate Bodies, Persons, and Families" *Journal of Library Metadata*, 11:166–175, 2011. [Also available online.](#)

designation as a creator of the archival resource or as a subject. The PHRT also displayed extracted biographical information associated with the identity which provided essential identity context.²

Following these successes, the Andrew W. Mellon Foundation awarded two separate grants in 2011 and 2012-2013 for further technical research and development. Developments predating the start of the 2012 Mellon grant included , SNAC software creation and refinements in match/merge processing programs. This work resulted in the creation of 128,297 unique EAC-CPF records (authority identities) being extracted from 30,500 finding aids while over 5 million OCLC VIAF clusters were used to normalize these identities. Additionally, regional daylong workshops for archivists and librarians in EAC-CPF began in 2011, funded by IMLS, and conducted by the Society of American Archivists (SAA).³ The goal of this training program was to encourage archivists to adopt this new standard for authority description upon which the future expansion of SNAC depends.

The technical research developments in 2010 led to the writing of separate funding applications focusing on developing an organizational model for the administrative management of these authorities. Recognizing the potential utility of SNAC authorities for cultural institutions, the U.S. Institute for Museum and Library Services (IMLS) awarded the University of Virginia three years of funding, beginning in 2011, for this work; and in 2012 the U.S. National Archives and Records Administration (NARA) joined the SNAC founding institutions in planning for a cooperative. In January 2013, NARA committed to hosting the cooperative and serving as its secretariat, while the technological infrastructure would continue to be developed/maintained by the CDL. Between 2013 and 2014, the possibility of expanding to form an international archive description and access cooperative was discussed, and in August 2014 the Andrew W. Mellon Foundation awarded SNAC a grant to support initial planning that focused on the details of the cooperative's legal, administrative, and technical structures.

On May 21-22, 2012 NARA hosted the first of three meetings, sponsored in part by the IMLS.⁴ Participants were tasked with developing a blueprint for establishing a sustainable National Archival

² The following is taken from [SNAC Webpage: Research and Development section](#) but text has been modified. Processing of source data takes place in three steps. First data is extracted from the source descriptions and assembled into EAC-CPF descriptions. Next the resulting EAC-CPF descriptions are matched and combined with one another, and then matched against Virtual International Authority File (VIAF) records. Data from matching VIAF records supplements the data in the EAC-CPF descriptions. Finally, the resulting EAC-CPF descriptions are added to the foundation of a public research tool (PHRT). Each step is performed by one of the three SNAC R&D collaborators. The first step, extracting/assembling, is done at the Institute for Advanced Technology in the Humanities, University of Virginia. The second step, matching/combining, is performed at the School of Information, University of California, Berkeley. The third and last step, developing the public research tool, is being carried out at the California Digital Library, University of California Office of the President.

³ In September 2013 the IMLS provided a grant extension for training through June 2014, see [Building a National Archival Authority Infrastructure and Scholarship Program for EAC-CPF Workshops](#).

⁴ The NAAI project is funded through a grant from the Institute of Museum and Library Services (IMLS) and support from the Institute for Advanced Technology in the Humanities at the University of Virginia and Simmons College Graduate School of Library and Information Science. See [Building a National Archival Authorities Infrastructure](#).

Authorities Cooperative (NAAC). At the first meeting, the SNAC project served as the model to demonstrate both the feasibility of creating authorized archival identities and the benefits of these resources for research activities. A second meeting in October 2012 resulted in a draft blueprint outlining the essential business, administrative, and technological components of a national cooperative.

Already in March that year, the SNAC Program Working Group charter was signed at NARA and the Work Group announced in August 2015 the beginning of the pilot cooperative program.⁵ A subsequent 2015 funding proposal to the Andrew W. Mellon Foundation designed to build upon on the draft blueprint document mentioned above was written and currently serve as a guide for the continued work both on the cooperative's infrastructure and on the definition of its technical, administrative and governance functions. The seventeen inaugural cooperative members will continue to work on these issues until the grant expires in July 2017.⁶ A future major goal for the fully established cooperative will be to support manual maintenance of the CPF descriptions through an editing interface.

The continued impetus for these cooperative developments was due, in part, to the expansion and refinement of the technical features of SNAC. After August 2014, when the SNAC website underwent a major redesign, more than 2.6 million unique EAC-CPF authority records were contained in the PHRT.⁷ Less than a year later, in July 2015, the number of identities contained in the PHRT had surpassed 3.7 million with links to more than 4000 repositories. As previously mentioned, these linked resource descriptions include identities whether they are described as creators or simply as subjects within those resources. The resultant PHRT identity-display thus serves as a social network of authorized identities linked to the specific research collections where they occur. It adds further researcher value by including descriptive biographical information on the identity harvested from the source EAC-CPF records.

Having reached this stage, a future major technological goal is the transformation of the SNAC R&D platform into a platform that will support ongoing cooperative maintenance of the SNAC description and access data. This operational model would permit the SNAC cooperative authority file to become in effect the local authority file simply through links to local archival record descriptions. Individual institutions' would no longer need to maintain local authority files. To achieve this it is certain that the symbiotic nature of SNAC technical and cooperative development must continue to operate in the future.

⁵ Charter states: The purpose of the Cooperative is to establish an operational international information resource that aggregates reliable descriptive data about people and organizations who created or are documented in archival holdings and for discovering important archival name authorities and their related social networks and context. Cited from [SNAC Webpage Cooperative Program Founding Documents](#).

⁶ See [SNAC Webpage Cooperative Program Member Institutions](#) for full list of pilot cooperative members.

⁷ PHRT functionality was also significantly enhanced at that time through Google analytics demographics feedback from users.

BIBLIOGRAPHY

Consulted the following webpages or web presentations:

<http://socialarchive.iath.virginia.edu/> (viewed many uri's cited under this general path)

<http://socialarchive.iath.virginia.edu/SNACI/>

http://socialarchive.iath.virginia.edu/SNACI/NAAC_index.html

<http://eac.staatsbibliothek-berlin.de>

<https://ecommons.cornell.edu/handle/1813/28718>

<http://www2.archivists.org/news/2011/building-a-national-archival-authority-infrastructure-and-scholarship-program-for-eac-cpf->

<http://www.tandfonline.com/doi/abs/10.1080/19386389.2011.629960>

<https://www.youtube.com/watch?v=iTlodzmEABM&feature=youtu.be>