

Accessing Anthropology: Digital Collections at the National Anthropological Archives

Stephanie Ogeneski Christensen, National Anthropological Archives, Smithsonian Institution; Washington, D.C., USA

Abstract

Since the start of the National Anthropological Archives digital imaging program in 1999, the archive has produced over 100,000 digital assets from archival material to include sound recordings. Several of the projects have promoted outreach and access to the collections through avenues of collaboration. Working with both source communities and researchers, our collections are continually reaching new audiences and are being used in creative ways. In a web 2.0 and social networking world, the NAA is looking ahead towards new ways to promote access. This paper highlights considerations, challenges and successes in achieving these projects.

History and Background

The National Anthropological Archives (NAA) is part of the National Museum of Natural History, Smithsonian Institution and is located at the Museum Support Center (MSC) in Suitland, Maryland. The Museum Support Center contains over half a million square feet of space dedicated to the storage and safeguarding of museum collections. Contained within the pods of MSC, the National Anthropological Archives (NAA) accounts for approximately 16,700 cubic feet of storage space for its collections. There is also off-site nitrate negative storage. The archived collections contain historical and contemporary materials that document the world's cultures and the history of anthropology. The NAA collections represent the four fields of anthropology – ethnology, linguistics, archaeology, and physical anthropology. The collections consist primarily of textual and photographic materials, but also include graphic works of art and sound recordings created by Smithsonian anthropologists and other preeminent scholars. The NAA has in its collections: 9,700 linear feet of manuscripts, 402,000 ethnological and archaeological photographs, 21,000 works of art, 3,800 sound recordings, and approximately 40 born-digital collections.

The collections include the Smithsonian's earliest attempts to document North American Indian cultures (began in 1846 under Secretary Joseph Henry) and the research reports and records of the Bureau of American Ethnology (1879-1964), the U.S. National Museum's Divisions of Ethnology and Physical Anthropology, and the River Basin Surveys. The NAA also maintains the records of the Smithsonian's Department of Anthropology and of 25 professional organizations, including the American Anthropological Association, the American Association of Physical Anthropologists, the American Ethnological Society, and the Society for American Archaeology. Among the earliest ethnographic collections are the diaries of John Wesley Powell, which recount his exploration of the Colorado and study of the region's Indians, and the pictographic histories of Plains Indians collected by U.S. military officers and BAE ethnographers. Other

significant manuscript collections include the ethnographic and linguistic research of Franz Boas, Frances Densmore, Alice Cunningham Fletcher, Albert S. Gatschet, John Peabody Harrington (including sound recordings), and J.N.B. Hewitt, as well as the expedition logs, photographs, and film record produced on Matthew Stirling's explorations in New Guinea (1926-29). [1]

The audience for the archives typically includes linguists, historians, artists, publishers and people researching their cultural heritage. In 2009, the NAA had approximately 700 on-site research visits, 650,000 web visitors, and approximately 9 million images viewed online through its collection information system (CIS). Our collection information system holds a little over 100,000 catalog records that include both collection and item level records. In 2007 the NAA had 77,500 high-resolution digital images, of which, 42,900 of these digital images were online allowing for 55% of its digitized collections accessible in digital format. In 2009, the NAA had approximately 110,000 digital images and approximately 76,000 of them online, changing the archive's online presence of digitized collections to 70%. Included in material not accessible online is a percentage of culturally sensitive images, images with rights management restrictions, and backlogged images that are awaiting processing. While the archive continues to achieve a greater online presence each year, we constantly face issues that require balancing accessibility, resources, and limitations.

The Value of Collaboration

The NAA's interest is to expose the collections to new audiences through innovative outreach that makes the archived digital assets available to the public without compromising the historical content and value contained within them. The NAA is interested in promoting knowledge of anthropological research, repatriating knowledge to source communities, and cultivating relationships with communities of interest. The selection criteria for the analog collections that appear in digital form was established to address common issues of preservation and access: to preserve the originals and reduce their handling, increase availability to the collections (i.e. glass plate negatives), and increase access to items with high cultural value. The vast holdings of the NAA have such a rich amount of information that digitization of these collections has allowed that information to extend beyond the reference room, which in turn has encouraged visitors to come out to the facility and conduct more extensive research. Over the years, many of the digitization projects have been accomplished through collaboration. These collaborations have included ArtStor (www.artstor.org), the Chaco Digital Initiative (www.chacoarchive.org), partnering with native tribes, such as the Cherokee and Coushatta, and global collaborations such as the Rosetta Project (www.rosettaproject.org).

One of the early collaborative projects the archive embarked on was with the nonprofit initiative of the Andrew W. Mellon

Foundation called ARTstor. Artstor's original intention was to aggregate and provide online access to canonical works of art, but the Mellon Foundation was receptive to the NAA's suggestion to include a collection of Plains Indian Drawings and Glass Plate Negatives of the American Indians created by the Smithsonian's Bureau of American Ethnology. Researchers using Artstor receive authoritative data from catalog records from the Smithsonian's own collections information system (CIS), SIRIS (Smithsonian Institution Research Information System), alongside a set of sophisticated Java-based tools that allow them to zoom, pan, and rotate images. This collaboration has greatly increased exposure of one of our most popular and requested collections to educators and scholars from a variety of non-profit institutions that may not have been aware of our online catalog previously.

Following the ArtStor project, the archive partnered with the Chaco Digital Initiative. This collaboration was sponsored by the Mellon Foundation, the Institute for Advanced Technology in the Humanities at the University of Virginia, and the National Science Foundation. Its goal was to create an electronic research archive that would integrate the widely dispersed archaeological data collected from Chaco Canyon in the late 1890's and the first half of the 20th century. The integrated archive would allow scholars to more effectively and efficiently address unresolved issues regarding culture change and organization in the canyon and surrounding region. For this project, the NAA contributed research data and photographic documentation from the papers of two Smithsonian archaeologists: Neil Judd and Frank Roberts. Our collaboration with the Chaco Digital Initiative has been rewarding in unexpected ways: Chaco researchers have helped to identify some of the Smithsonian's earliest archaeological photos of Chaco Canyon, including structures that have never been recorded and have collapsed since. The expertise of the Chaco researchers is steadily improving the Smithsonian's collection cataloging data, while making that information available online to an ever-larger audience. The archive still faces challenges on the backlog of information generated from this collaboration and implementation of it into the archive's cataloging system.

Source communities, such as the Cherokee and Coshatta native tribes also contributed to our digitization efforts. The Cherokee collaboration involved working with the Museum of the Cherokee Indian to digitize in two phases over 9,000 pages of the Smithsonian's Cherokee language manuscripts for the Cherokee Virtual Museum. The NAA is one of the world's most important repositories of material related to the Cherokee culture, language and history. Our collection includes manuscripts written in the Cherokee syllabary as well as ethnobotanical accounts of the Cherokee culture written by Smithsonian anthropologist James Mooney, whose association with the Cherokee began in 1887. The collection includes songs and musical transcriptions; lists of Cherokee personal names and place names; early maps and censuses; copies of Cherokee treaties; and a wealth of ethnobotanical and medical information. Cherokee museum staff also have carried out fieldwork with elders and native speakers in order to assess these materials, translate them, and develop ways to use them in Cherokee language preservation programs. The digital Cherokee collections have also been made available through our online collection information system, SIRIS, with the exception for the manuscripts too culturally sensitive to display in an online

environment. The Cherokee project has been one of our largest digitization projects involving text. Issues of handling and conservation concerns of certain manuscripts needed to be addressed before digitization occurred. Due to the fragility of many of the manuscripts, decisions to capture text were defined giving consideration to areas such as color and fading. Manuscripts were digitized in their entirety as the archives felt that preserving the entire manuscript true to its original form was also of significant value to our collections.

Continuing our collaboration with native tribes, the archive worked with the Coshatta Tribe of Louisiana. By digitizing over 6,000 pages of endangered language manuscript material for the Coshatta Tribe, the tribe has utilized the digital assets and their contents to integrate into interactive displays, learning websites, and video games to help teach younger generations about their language and heritage. [2]

A final language collaboration was with the Rosetta Project. The NAA has contributed to this project digitized copies of some of the first sound recordings from Native American Indian languages. These recordings are from the J.P. Harrington Collection, a collection originally containing wax cylinder and aluminum disks. The recordings document some of the last remaining speakers of native languages. These audio files are highlighted on the Rosetta site (www.rosettaproject.org/projects/google-earth/) through a set of global linguistic visualizations for Google Earth. The Google Earth layers allow for a more intuitive way to browse, locate, and discover the special collections that Rosetta holds. These capabilities to the site, allows administrative users the ability to reorganize linguistic trees to explore alternative visions of historical language development. [3]

While I have highlighted some of the archive's partnerships over the years, it should be noted that without working in a collaborative nature with our audience and researchers, many of our early digitization projects would not have seen fruition. These projects have helped provide funding to our digital imaging program and have also allowed for discoveries of knowledge, which has benefited the archives as well as its audiences. As the material begins to be disseminated beyond the collection catalog and into public websites and social networking sites, the re-use and re-purposing of these images starts to allow for more unexpected discoveries of the materials.

Towards a Greater Access: Modes of Outreach for Collection Material

Part of the Smithsonian's mission is the "increase and diffusion of knowledge." As digitization increases across the institution, new ways of learning about and modes of discovery of the collections and vast resources that the Smithsonian holds will begin to take place. Users in today's world will expect to have greater access and search functions on the institution's digital assets. They may want to relate one collection in the National Anthropological Archives with holdings in the Smithsonian Institution Libraries. A photograph in the NAA may be helpful in describing an artifact held in the Department of Anthropology's collections artifact database. The list of possibilities and roadmaps is endless. As the archive works with researchers and source communities to create and disseminate digital assets to outreach

environments, the archive is also looking at ways internally and through Smithsonian collection information systems, websites, and social networking channels to increase the display of their digital collections to the public.

In the past few years, the archive was able to get a backlog of early digitization projects online. One of the issues standing in the way was the fact that item level catalog records of these collections lagged behind digitization efforts. The other issue that the archive faced was that many of these images required multiple image display capability, which our collection information system was limited in at the time. By utilizing a flash xml image viewer, this project was to develop the capability to display multiple images in large quantity. Working with the SIRIS office, the NAA identified existing collections and records to achieve getting approximately 22,000 backlogged images online in one year. The level of complexity of the collections contained in this backlog ranged from a simple group of 20 drawings to a manuscript that contained 1587 digital files with a complex hierarchical structure. Within this particular manuscript, there was the collection level record, 15 series records, and within some of those series records even more division. For example, one of the series records is divided into 16 folders, containing 429 digital objects. Overall, this project has allowed greater access to digitized collections that would otherwise not be available outside of the NAA reading room. It has continued to promote knowledge of anthropological research through our digital assets. For example, a researcher saw a particular image from a manuscript online, it helped in facilitating his research regarding the stone from which the particular tablet was made, he visited the archives and also the department of anthropology's collections area where the three dimensional object resided. In the end, it provided new evidence to his research. [4] The ability of the multiple image display capability facilitated this manuscript going online.

While the image viewer capability proved to enhance accessibility, it did have its limitations. For instance, when the images are contained within the image viewer, they are not individually searchable by the Collection Cross Search Center (CCS). As digitization across the institution has increased, SI has created the consequent need for centralized searching of collections and the inherent need for long-term storage of digital materials. In order to meet these ever increasing needs, a Collections Search Center (CCS) and a Digital Access Management System (DAMS) were created. The CCS (<http://collections.si.edu>) is one stop searching based on a central index, with data pulled from and linked to a number of collections, which employs an index metadata model based on standards and internal data needs, while the DAMS is intended for the long-term storage of digital assets and uses a core set of metadata created through a pan-institutional committee process. The searchable metadata is dependent on the comprehensiveness of the description of the original analog material. The metadata quality begins to work in conjunction with the quality of the metadata contained within the images. While the archive is continually working on enhancing its records, budget cuts and staffing over the years has presented challenges.

To overcome these challenges, the archive has worked on collaborative pan-institutional efforts towards a DAMS core metadata and core embedded metadata to help define core

metadata that will assist in the ability to readily describe digital assets, determine which digital assets are available for use, which have restrictions, or which assets belong to a particular unit. The metadata model of the DAMS as well as core embedded metadata will become essential to the distribution of the archived digital assets not only internally but also externally. While this has proven to be a productive collaboration, we have recognized that there are challenges within our own institution. Defining a core set of fields that can accommodate a broad range of materials has not been an easy undertaking. For instance, our efforts towards a core embedded model suggests elements that are *consistent* across SI, while our suggested set of elements have been established to accommodate the needs that one unit may have over another. As these efforts begin to move forward, we are testing the capabilities of the CCS and DAMS systems with new ways to extract and deliver metadata. For example, the archive will be working with our collection information office to extract embedded metadata from images within the DAMS as well as from core DAMS metadata to help assist in creating MARC records for use in the CCS system. This pilot project will involve a collection of 5,000 images that have been digitized but for which no item level records exist. The goal is to create minimal MARC item level records based on the embedded metadata and DAMS core metadata that can then be enhanced by staff. Should this project prove feasible, this would help in automating some of the preliminary steps of the process, free up staff time, and provide a solution that balances a multiple image display capability and the ability to enhance search functions.

Where Do We Go From Here?

As our digital assets move from creation to dissemination, the archive is looking internally to enhance our online presence of our digital assets through our public websites. With the implementation of the Collections Cross Search (CCS) a metadata search and access layer, known as EDAN (Enterprise Digital Asset Network) was implemented. This layer offers on-line discovery tools such as an Image Delivery Service (IDS). Utilizing this image delivery service, the capability to serve up to public websites digital assets from the original source where the content is maintained now exists. While this IDS is being utilized through the archive's collection information system to access the images, one of the website projects in development, *Accessing Anthropology: The Collections and Archives Program at the Department of Anthropology*, is to utilize this feature through the Department of Anthropology's public website to create a dynamic revolving portal that pulls from the collections at the Anthropological Archives, the Department of Anthropology's artifact database, and the collections of the Human Studies Film Archives. This portal, while drawing on current technologies developed SI wide, will access the anthropological collections and also highlight some of the department's online exhibitions, such as the John Marshall Film Collection. [5] It will serve as a collections commons for the department, and an additional online point of access to provide greater exploration of the anthropology collections contained in the Collection Cross Search (CCS). In terms of developing web 2.0 capabilities to the site, the collaborative SIRIS blog (<http://si-siris.blogspot.com/>) from our collections catalog is providing guidance. The SIRIS blog is a pan-

institutional collaboration in which units that utilize the SIRIS collections information system, have the ability to post blogs that range from collection highlights, interesting finds within the collection, or current events as they relate to particular collections. This blog has provided additional visibility and points of access to the NAA's collections. In our commitment to provide a greater digital access to our collections the archive will continue to strengthen the foundation from which we build and disseminate our digital assets. Continually looking at the life cycle of the asset, the archive will need to keep in mind that these digital assets will assist in the "increase and diffusion of knowledge" from the creation of the digital asset to the metadata quality associated with that asset. This framework will help guide us in administrative and management decisions that balance preservation and access while maintaining the goals of the National Anthropological Archive's digital collections.

Acknowledgements

I would like to thank all current and previous staff, volunteers, and interns of the National Anthropological Archive and Human Studies Film Archive, and all those who have who have assisted in the realization and continuation of the digital imaging program at the archive. I would also like to thank Robert Leopold, Director, National Anthropological Archives, John Homiak, Director, Anthropology Collections and Archives Program, and the staff of the SIRIS office.

References

- [1] Glenn, James R., Guide to the National Anthropological Archives, Smithsonian Institution. (Smithsonian Institution, 1996). Leopold, Robert, "About the Archives," <http://www.nmnh.si.edu/naa/about.htm>.
- [2] Coshatta Tribe of Louisiana, "Kosati (Coshatta) Language Project," (2008). <http://web.wm.edu/linguistics/coshatta/>. Capriccioso, Rob, "An Indian Nation Preserves Its Heritage with a Magic Joystick" New York Sun (September 7, 2007). <http://www.nysun.com/national/indian-nation-preserves-its-heritage-with-a-magic/61862/>.
- [3] The Rosetta Project: A Long Now Foundation Library of Human Language, <http://www.rosetta-project.org/projects/google-earth/>.
- [4] Steelhammer, Rick, "Grave Creek Stone hoax linked to Wheeling doctor" Charlotte Gazette (October 13, 2008). Lepper Bradley T., "Great Finds on West Virginia nothing more than a fraud" The Columbus Dispatch (November 11, 2008).
- [5] Smithsonian Institution, "A Million Feet of Film/ A Lifetime of Friendship: the John Marshall Ju/'hoan Film and Video Collection, 1950-2000" (2009) <http://anthropology.si.edu/johnmarshall/index.html>.

Author Biography

Stephanie Ogeneski Christensen is the Digital Imaging Manager at the National Anthropological Archives, Smithsonian Institution. Prior, she served as the manager of the digital imaging facility at the Chicago Albumen Works. She is a member of the Federal Agencies Digitization Guidelines Initiative Still Image Working Group. She holds a Certification in Photographic Preservation and Archival Practice from the George Eastman House (1998) and an MFA from Indiana University (1997).