Research-driven approaches to improving archival discovery

Diana E. Marsh[1]

Abstract

The National Anthropological Archives (NAA), part of the Department of Anthropology at the Smithsonian's National Museum of Natural History, holds some 18,000 cubic feet of materials of relevance to qualitative researchers. These archival collections—manuscripts, fieldnotes, audio recordings, drawings, maps, and still and moving images—are used by not only anthropologists, but increasingly scholars from a range of qualitative research fields. In 2016, the NAA received a grant to support a 3-year post-doctoral fellow to conduct research that would lead to the improved discovery and use of archival resources. This article discusses some of the practical ways the fellowship was designed to ask interdisciplinary research questions, and describes how that premise, as well as findings from a pilot study run in the first year, are helping to improve the research experience for our increasingly interdisciplinary users. Both the project's preliminary findings and its overall design may provide valuable insights to qualitative researchers and their institutions.

Keywords

access, anthropology, archives, collections, users

Introduction

The National Anthropological Archives (NAA), part of the Department of Anthropology at the Smithsonian's National Museum of Natural History, holds some 18,000 cubic feet of materials of relevance to qualitative researchers. These archival collections—manuscripts, fieldnotes, audio recordings, drawings, maps, and still and moving images—are used by not only anthropologists, but increasingly scholars from fields such as history, art history, and social studies of science. In 2016, The NAA received a grant to support a 3-year post-doctoral fellow to join the staff of the NAA and the broader scientific staff of the department. The goal of the fellowship was to conduct research that would explore archival access, and ideally lead to the improved discovery and use of archival resources of value to researchers. This article discusses some of the practical ways the fellowship was designed to ask interdisciplinary research questions, and describes how that premise, as well as findings from a pilot study run in the first year, are helping to improve the research experience for our increasingly interdisciplinary users. Both the project's preliminary findings and its overall design may provide valuable insights to qualitative researchers and their institutions.

The NAA Collections

The National Anthropological Archives is the United States' largest archival repository dedicated to the history of anthropology and the world's cultures, with over 18,000 cubic feet of historical documents, photographs, audio recordings, and film. It holds one of the world's largest archival collections of both American Indigenous languages and ethnographic film.

Such collections are crucial to understanding how published research was generated. Archival materials include a range of unpublished work by scientists and fieldworkers and their colleagues, such as drafts of manuscripts with annotations and corrections, loose sheets or cards containing

raw linguistic documentation, notes containing free-form thoughts and observations, or letters illustrating scholarly relationships and networks. Such archival documents provide the full context in which knowledge was obtained, synthesized, and produced.

As I have written about elsewhere, it was clear through the range of publications generated from NAA's archival materials that these collections are increasingly being used by a diverse range of qualitative researchers (Marsh 2018). Raw research materials, such as fieldnotes, personal diaries, correspondence, annotated maps, photographs, sound recording, and video, are now being used not only by anthropologists, but linguists (Davis 2010), environmentalists (Anderson 2005) and ecological historians (Loring & Spiess 2007), immigration scholars (Schmidt, Seguchi, & Thompson 2011), apparel scholars (Marks 2014), the historians of science (Hinsley 1994; Rich 2012), musicologists (Troutman 2013), and ethnomusicologists (Moon 2010), English literature scholars (Applegarth 2014), and art historians (Naeem 2018).

Non-academic researchers, such as artists, documentary filmmakers, exhibit designers, journalists, and even children's book authors are also researching these collections for a range of uses with much wider public exposure. Native and Indigenous community members are also researching their own histories, languages, and cultures in these collections, especially in service of language revitalization programs.[2] Yet, little had been done to analyze this apparent shift.

Furthermore, it was known anecdotally that these collections were not reaching the broadest possible range of researchers because of barriers to their access both in-person and online.

Research Questions

This postdoctoral NSF project was therefore driven by three premises: 1) that despite the importance of NAA archival collections and their increased digital presence, usage remains below the immense potential that the collections hold; 2) a general institutional desire to see NAA collections have more scholarly centrality and citation, as well as overall circulation and secondary use; 3) the hypothesis that collections discovery and access are hindered by current descriptive practices, discoverability tools, and interfaces.

Preliminary research questions included:

1. How can the NAA make its collections more discoverable, accessible, usable to researchers?
2. What difficulties are encountered by anthropological researchers and source communities in seeking information in the archives?
3. What attitudes or understandings about archival research are held by anthropologists and other researchers?
4. How can archivists more effectively involve anthropologists and source communities in the archival processes of collection representation?
5. How can archival descriptive practices better represent elements of the collection to increase discoverability by anthropological researchers?
6. How can "traditional" archives such as the NAA better engage with emerging digital data repositories such as the Digital Archaeological Record (tDAR), the Archive of Indigenous Languages of Latin America (AILLA), the Open Language Archives Community (OLAC),

and the Digital Endangered Languages and Musics Archives Netork (DELAMAN), and how can we best develop shared understandings of "archives," and "digital data"?

My research sought to establish a better understanding of both archival repository and user needs to improve researcher success in the discovery of archival sources. In addition, a core goal of the fellowship was to fill a gap in professional training between archival studies and anthropology. Organizationally, the NAA (including the Human Studies Film Archives) sits within the Collections program of the Department of Anthropology, within the National Museum of Natural History at the Smithsonian Institution. Being embedded within an anthropology department (staffed with experts in anthropological research) as well as within an archive (staffed with experts in archival science and practice) was key to bridging this gap. The fellowship was designed to meld hands-on work in the NAA with research by joining the NAA staff, and work on an ongoing collections assessment to learn the collections.

Because my background is primarily in anthropology, I spent a good deal of time in the first year of the fellowship reading archival science basics and learning how to help with the NAA's daily work, especially in reference. I also assisted our contract archivist, Gabriela Sanchez, with the NAA's current comprehensive collections assessment. My task was to assess the "intellectual value" of collections based on documentation quality (based on the types of materials and their uniqueness), researcher interest (based on topical focus and past use), and local importance (based on relationships to other collections or inherent institutional value to the NAA). In total, we assessed over 300 collections in the first year.

Concurrently, I began an environmental scan that included a) informal interviews with NAA staff about current users, uses, discovery tools, and access issues to glean NAA staff understandings of the grant's research questions; b) compiling a project bibliography on archival users, access, discoverability and other relevant readings, and c) reviewing previously produced institutional reports and studies relevant to the current project.

Environmental Scan & Assessment Findings

My literature review made clear that this project has novelty due to its disciplinary emphasis on anthropological and Indigenous collections and its implementation component.[3] In particular, few user studies have the benefit of being undertaken at a repository, rather than by university-based researchers, or of a three-year timeline in which findings can be implemented and reflected upon.[4] It is not so much that, as Elizabeth Yakel (2004, p. 65) noted over a decade ago, "user evaluation has rarely been mentioned as an integral aspect of implementation," but rather that most studies lack the longitudinal timeline and internal institutional support to directly apply evaluation findings to practical change.

Findings from year one of our assessment revealed that NAA collections have major intellectual access barriers. Of 314 assessed collections, 253 (81%) do have at least a collection-level catalog (MARC) record (e.g. the Beatrice Medicine papers). However, only 25% have a more detailed finding aid online. Many of these are pdf documents. At the time of writing this piece, due to the recent redesign of the NMNH's website, many of those pdf finding aids are no longer findable online, and it will take staff time to get them back up. Only 15% of the 314 assessed have a fully keyword searchable, EAD finding aid (in ArchivesSpace) that comes up in all Smithsonian search

platforms (and therefore has been guaranteed to survive the NMNH website redesign). Therefore, on a scale from 1 to 5, where 5 is highest, only 15% of our year one assessed collections are rated a 5 and considered highly accessible (See Figure 1).
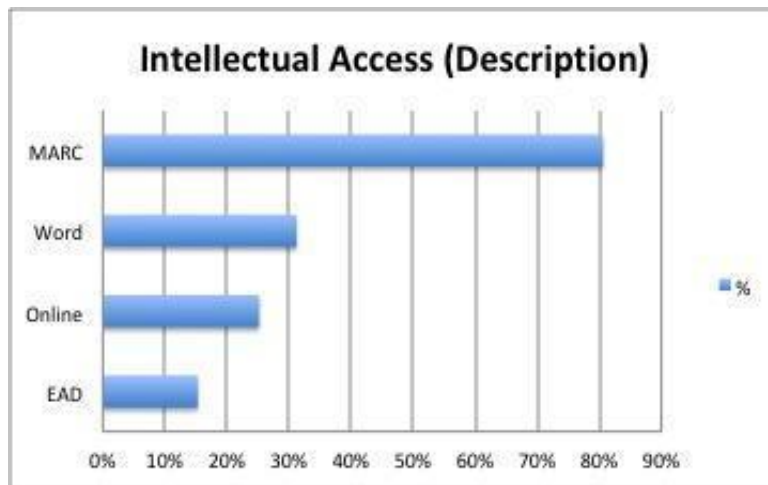

Figure 1. Percentage of Collections Accessible by Description Type of 314 Assessed

Of course, the NAA is not alone in this. According to a study of their backlog, the National Archives and Records Administration reports only having 26% of its textual collections processed sufficiently to allow "researchers to easily identify records of interest." Thirty-three percent of its collections records lacked basic elements of intellectual control such as titles or dates (Bucciferro 2008). According to a 1998 study, the mean of special collections repositories' backlog is 33% (Panitch 2000).

I found that the NAA's collections also had additional major barriers to discovery due to the design of its website. Working with two Spring Break interns from the University of Michigan's iSchool, Thanhthu Nguyen and Wendi Ding, we also found that the NAA website lacks discovery functionality. From Google Analytics, NAA sites have a 50-60% exit rate (percentage of users that leave the site from a page) and a 75% drop off rate (percentage who don't click through to a next page). In other words, researchers were not aided in finding materials of relevance on our institutional website.

For qualitative researchers, this means that both at the NAA and elsewhere, many materials of relevance are almost impossible to find, or may not be available for use at all. Typically, unprocessed archival collections are not made available to researchers because they have not yet been vetted by archivists for potentially sensitive, personally identifiable, or legally complex materials. Many archival repositories have websites and interfaces that are not intuitive to researchers and require institutional knowledge to navigate. Moreover, these systems often change, so that researchers may need to learn and re-learn interfaces at institutions with relevant collections over the course of their project.

Pilot Study Findings

The primary research conducted in 2017-2018 was a pilot study to better understand these NAA users and their information-seeking behaviors. The pilot study included: a) exploration,

preliminary coding, and analysis of available NAA FY2016 users from the NAA's remote reference log, visitor appointment database, and permissions database; b) scheduling and completion of 22 targeted 1-hour interviews and three focus group discussions with user communities identified during the existing user data analysis; and c) transcription of all recorded interviews and focus groups for coding and analysis.

An analysis of our FY2016 databases confirmed that the NAA has a highly diverse set of users (See Table 1). Native community-based researchers are now the NAA's second largest user group, and we have almost an equal number of academic (47%) and non-academic (46%) users. In addition, from central FY2017 Smithsonian data, the NAA serves users from 49 US states and territories and 33 countries around the globe. Thus, we are serving a range of qualitative researchers from across different disciplinary and professional backgrounds.

Total Frequency of Users in Top 7 Groups FY2016	
User Group	**Frequency (n=1004)**
Anthropologist	170
Community Researcher	91
Historian	85
Heritage Professional	82
Art Historian	41
Filmmaker	40
Other Social Scientist	32

Table 1. Frequency of Users in Top 7 Researcher Groups, FY2016

Twenty-two participants were recruited from these top user communities. I loosely correlated the number of participants from each designated community to the number of total users from that group in 2016. In total, I interviewed: 6 anthropologists from different subdisciplines, 5 community-based researchers, 4 heritage professionals, 3 historians, 2 filmmakers, and 2 humanities scholars (one an art historian and community member). In total, these researcher interviews include 20 hours of audio .wav recording and three written responses. Sixty-four participants were invited to participate in three Smithsonian focus group discussions, with a total of 14 Smithsonian staff participants and 4 hours of .wav audio recording. All interview and FGD transcripts are transcribed and were coded using TAMSanalyzer.

Key interview findings thus far include that:
1. **Search tendencies make collections harder to find for community and non-academic users.** Pathways to NAA collections differ by user community. Academics and community researchers tended to find out about the NAA through word of mouth, either in a fellowship or directly from colleagues. All heritage professionals and filmmakers, and all who identified as photo researchers found out about the NAA through online searches. Only academics found out about the NAA through bibliographic sources. Only academic users mentioned using a Finding Aid to identify relevant collections. Many academic users search by specific anthropologists' (record

creator) names or by collection; all non-academic and community-based users tend to search by cultural group name or subject.

2. **Researchers lack training in archives.** Very few researchers receive any training in archival research (the logistics of conducting archival research or how archives are organized), and describe learning "as they go," even if they attended graduate programs in anthropology or history.

3. **Current Smithsonian search platforms are not intuitive for users**, even if they know what they are looking for. Multiple entry points at Smithsonian and nested nature of NAA exacerbates this problem.

4. **Outdated information and thin description is problematic.** Users also noted the existence of problematic and incorrect catalogue information, where community members specifically mentioned outdated, problematic, or racist terminology (and collections' description non-Native perspective) as an issue. Desire for more depth of collections description was identified by 4 academic participants and was thus a lesser factor than expected.

5. **User expectations are shifting**. Users expect more collections easily accessible digitally, especially through the presence of more digital surrogates online. The availability of digital surrogates and the lack of on-demand or on-site digitization were listed as *the* top barrier to collections access.

6. **Gaining on-site access is confusing, difficult, and cost-prohibitive**. Many interviewees mentioned difficulty in finding out how to contact the NAA, our appointment process, our security process, and the prohibitive cost and distance to visit. One community user noted that the security process evokes historical trauma. The confusing nature of the Smithsonian's organizational structure (and what collections are where) adds to the feeling of institutional impenetrability.

Research during year two of this project will include a survey with a number of professional organizations such as the American Anthropological Association, the Association of Tribal Archives, Libraries, and Museums, Native American, the National Association of Tribal Preservation Officers, and the Native American and Indigenous Studies Association.

Implementation of Improvements

Instigated by this research, we have begun to make a number of practical improvements, however small. As part of the NMNH's [website redesign](), ongoing during the first year of my fellowship, we worked with other anthropology department staff on the development of a [new appointment form]() that will allow on-site researchers to self-identify their research interests and disciplinary background or subject expertise. This will allow us to more accurately track our users and their backgrounds, and how their interests are changing through time.

We are also partnering on broader initiatives. In May, I assisted with a joint University of Maryland-NAA workshop aiming to understand the impacts of digitized ethnographic archives. The workshop, run by UMD Assistant Professor Ricardo Punzalan, brought together for 30 community-based users of the John Peabody Harrington papers. We are also collaborating with the Washington State University Center for Digital Humanities' Mukurtu Shared project (a web-based version of the [Mukurtu CMS]()). This project brought two researchers (Mukurtu fellows) to the NAA for four months to identify collections of interest to partner Native community partners, which will be brought into the Mukurtu Shared system for community comment and the

development of 'copyright'/access protocols. Additionally, Ricardo Punzalan and I are working together to revitalize the Council on the Preservation of Anthropological Records (formerly at http://copar.org/) of which the NAA has historically been an active part, which will provide researchers with a resource for connecting to anthropological records and resources across the globe.

Thus far, we have applied for two grants to process and selectively digitize collections of high potential interest to users. We are working to initiate more cross-Smithsonian collaborations to showcase NAA collections through initiatives such as Smithsonian's Transcription Center, which will allow online volunteers to transcribe, and thereby make keyword searchable, our digitized collections.

Most recently, to address the lack of training in archives for many graduate-educated researchers, Gina Rappaport, NAA Photo Archivist, and I collaborated with Alessandro Pezzati at the Penn Museum Archives and Guha Shankar at the Library of Congress American Folklife Center to run an Archives 101 Workshop at the November 2018 American Anthropological Association Meetings. The workshop leaders were all from repositories with significant anthropological holdings. The workshop aimed to "demystify" archival repositories and jargon, and to provide an orientation to conducting research in archives. We introduced the general principles that govern archival organization and descriptive practices, described the types of records that are found in archival repositories and how they can be used, and helped our participants determine strategies for locating materials of interest to them in archival repositories, especially by searching online catalogs and finding aids. We gave participants time to search for archival materials relevant to their own research interests. We had 25 participants attend and received highly positive feedback. We think this is a model that can be replicated at a wide range of conferences for qualitative researchers. It might also be possible to create online videos, modules, or tutorials to train researchers in archival principles to help a wide range of researchers better intuit how to search for potential collections of relevance.

Conclusion

Archival collections are a primary research site for many qualitative researchers, and that interest is growing. Yet, it is clear that archival collections like those at the NAA are not easily accessible for remote research. As a result, many researchers may not be aware of collections of direct relevance to their projects. It is often through professional networks, fellowships, and in-person research that users may find out about collections without full descriptions online.

This project's interviews illustrated that many qualitative researchers would benefit from training in archival principles and practices. Most researchers, without knowing how archives are processed, described, and made accessible by archivists, lack the knowledge of collections' provenance to intuit where collections of relevance may be held in an archives.

Other repositories may find it useful to carry out similar projects, interviews, or interdisciplinary fellowships at their institutions. Our initial NSF grant proposal aimed to develop new anthropological leadership attuned to the needs and practices of both anthropologists and archivists. Applied fellowships of this kind help to bridge disciplinary gaps and to realign archival and disciplinary expectations and can encourage the broader use of archival collections as sites for knowledge production and to demonstrate the continued value of archival data for us all.

References

Anderson, K. (2005), *Tending the Wild: Native American Knowledge and the Management of California's Natural Resources*, University of California Press, Berkeley, CA.

Applegarth, R. (2014), *Rhetoric in American Anthropology: Gender, Genre, and Science*, University of Pittsburgh Press, Pittsburgh, PA.

Bucciferro, A. (2008), "Attacking the Backlog: NARA Archivists Mobilize to Make Unprocessed Records Available to the Public", *Prologue Magazine,* vol. 40 no. 2, pp. 46-51.

Davis, J.E. (2010), *Hand talk: Sign language among American Indian nations,* Cambridge University Press, Cambridge, UK.

Hinsley, C. (1994), *The Smithsonian and the American Indian: Making a Moral Anthropology in Victorian America,* Smithsonian Institution Press, Washington D.C.

Loring, S. & Spiess, A. (2007), "Further documentation supporting the former existence of grizzly bears (Ursus arctos) in northern Quebec-Labrador", *Arctic*, vol. 60 no. 1, pp. 7-16.

Marks, D. (2014), "The Kuna Mola", *Dress*, vol. 40 no. 1, pp. 17-30.

Marsh, D.E. (2018), "Toward Inclusive Museum Archives: User Research at the Smithsonian's National Anthropological Archives", In Yun Shun Susie Chung, Anna Leshchenko, & Bruno Brulon Soares (eds.), *Defining the Museum of the 21st Century: Evolving Multiculturalism in Museums in the United States* (pp. 129-142), ICOM/ICOFOM, Paris.

Moon, K. R. (2010), "The Quest for Music's Origin at the St. Louis World's Fair: Frances Densmore and the Racialization of Music", *American Music*, vol. 28 no. 2, pp. 191-210, doi:10.5406/americanmusic.28.2.0191.

Naeem, A., Knipe, P., Nemerov, A., Shaw, G. D., & Verplanck, A. (2018), *Black Out: Silhouettes Then and Now*, Princeton University Press, Princeton, NJ.

Panitch, J. M. (2000), *Special Collections in ARL Libraries: Results of the 1998 survey sponsored by the ARL Research Collections Committee*, Association of Research Libraries, Washington, DC.

Rich, J. (2012), *Missing Links: The African and American Worlds of R.L. Garner, Primate Collector*, University of Georgia Press, Athens, GA.

Schmidt, R.W., Seguchi, N., & Thompson, J.L. (2011), "Chinese immigrant population history in North America based on craniometric diversity", *Anthropological Science,* vol. 119 no. 1, pp. 9-19.

Troutman, J.W. (2013), *Indian Blues: American Indians and the politics of music, 1879–1934*, University of Oklahoma Press, Norman, OK.

Yakel, E. (2004), "Encoded Archival Description: Are Finding Aids Boundary Spanners or Barriers for Users?", *Journal of Archival Organization*, vol. 2 nos. 1/2, pp. 63-77.

Endnotes

[1] Diana E. Marsh is postdoctoral Fellow, National Anthropological Archives, Department of Anthropology, National Museum of Natural History, diana.e.marsh@gmail.com.

[2] See, for instance, Baldwin, D. (2017), *Language Reconstruction and Strengthening Community : The Role of Archival Resources*. Paper presented at the Irving K. Barber Learning Centre Events; Fitzgerald, C.M., & Linn, M.S. (2013), "Training Communities, Training Graduate Students: The 2012 Oklahoma Breath of Life Workshop", *Language Documentation & Conservation,* vol. 7, 185-206; Hinton, L. (2013), "The Use of Linguistic Archives in Language Revitalization: The Native California Language Restoration Workshop", In Leanne Hinton & Ken Hale (eds.), *The Green Book of Language Revitalization in Practice* (pp. 419-423), Brill, Boston, MA; Roy, L., Bhasin, A., & Arriaga, S.K. (2011), *Tribal Libraries, Archives, and Museums: Preserving our language, memory, and lifeways*, Scarecrow Press, Lanham, MD.

[3] There are related studies focused on historians and historical collections, such as: Anderson, I.G. (2004), "Are You Being Served? Historians and the Search for Primary Sources", *Archivaria* vol. 58; Torou, E., Akrivi K., Costas V., George L., and Constantin H. (2010), "Historical Research in Archives: User Methodology and Supporting Tools", *International Journal on Digital Libraries,* vol 11, no. 1, pp. 25-36; Toms, E.G. & Wendy D. (2002), "'I Spent 1 1/2 Hours Sifting through One Large Box...': Diaries as Information Behavior of the Archives User: Lessons Learned", *Journal of the American Society for Information Science and Technology*, vol. 53 no. 14, pp. 1232-38; For a broad study of Native user needs, see Association of Tribal Archives, Libraries, Museums, and Miriam Jorgensen. Sustaining Indigenous Culture: The Structure, Activities, and Needs of Tribal Archives, Libraries, and Museums. Association of Tribal Archives, Libraries, and Museums, 2012.

[4] For repository-based exceptions, see for instance, Altman, B. & Nemmers, J. (2001), "The Usability of on-Line Archival Resources: The Polaris Project Finding Aid", *The American Archivist*, vol. 64 no. 1, pp. 121-31; Conway, P. (1994), *Partners in Research: Improving Access to the Nation's Archive*, Archives & Museum Informatics, Pittsburgh, PA; Goggin, J. (1986), "The Indirect Approach: A Study of Scholarly Users of Black and Women's Organizational Records in the Library of Congress Manuscript Division", *The Midwestern Archivist*, vol. 11, no. 1, pp. 57-67; On applied implementation approaches based on user studies, see Brancolini, K.R. (2000), "Selecting Research Collections for Digitization: Applying the Harvard Model", *Library Trends*, vol. 18, no. 4, pp. 783-98; Daines, J.G. & Nimer, C. L. (2011), "Re-Imagining Archival Display: Creating User-Friendly Finding Aids", *Journal of Archival Organization*, vol. 9 no. 1, pp. 4-31; Nimer, C. & Daines, J.G. (2008), "What Do You Mean It Doesn't Make Sense? Redesigning Finding Aids from the User's Perspective", *Journal of Archival Organization*, vol. 6 no. 4, pp. 216-32.