Oral Tradition, 28/2 (2013):261-270

The Digital Archiving of Endangered Language Oral Traditions: *Kaipuleohone* at the University of Hawai‘i and *C’ek’aedi Hwnax* in Alaska

Andrea L. Berez

Introduction

In this essay I compare and contrast two small-scale language archives and discuss their relevance for oral tradition research.[1] The first of these is *Kaipuleohone*, the University of Hawai‘i Digital Ethnographic Archive (KUHDEA).[2] KUHDEA is administered by the Department of Linguistics at the University of Hawai‘i at Mānoa (UHM) and curated by the UHM library in an institutional DSpace repository under the purview of the UHM library. The second archive presented here is called *C’ek’aedi Hwnax*[3] (C’H), which serves the Ahtna Alaska Native community in and around the Copper River region of south central Alaska. C’H is fully administered by the Ahtna community itself via a non-profit organization known as the Ahtna Heritage Foundation (AHF).

These two archives have a number of features in common. They could both be called “niche” archives, in that they are both rather small in size, have a well-defined collection scope, and focus on recordings of languages—especially endangered languages—as used within many different genres. Additionally, they are both primarily digital archives, providing for the digitization of older analog materials while also accepting new, born-digital audio and video. They both strive to follow current best-practice recommendations for digital audio formats, storage, and metadata collection. They are both participating members of the Open Language

---

[1] Many thanks to Taña Finnesand, Karen Linnell, Liana Charley John, Kathy McConkey, Beth Tillinghast, Daniel Ishimitsu, and Nick Thieberger for their contributions to the projects described here. C’H is funded in part by the Institute for Museum and Library Services and the Ahtna Heritage Foundation; the Breath of Life workshop was sponsored by the National Science Foundation (award 1207511). KUDHEA receives support from the College of Languages, Linguistics, and Literature and Hamilton Library at UHM. Errors are the responsibility of the author.

[2] *Kaipuleohone* means “gourd of sweet words” in Hawaiian. We are grateful to Laiana Wong for suggesting this name for the archive. Information about KUHDEA can be found at http://www.ling.hawaii.edu/langdoc/archive.html, and access to the archive materials is found at http://www.kaipuleohone.org.

[3] *C’ek’aedi Hwnax* is Ahtna for (roughly) “legacy house.” The name was suggested by Ahtna Elders Jeannie Maxim, Markle Pete, and Virginia Pete, and was confirmed by Chief Ben Neely and Secondchief Fred Ewan in 2010.

Archives Community (OLAC)[4] and they both share metadata publicly via the OLAC search engine.[5] Finally, they were both conceived out of the growing interest over the last two decades in endangered language documentation and preservation, on the one hand,[6] and the increased use of digital infrastructure to serve the needs of social science and the humanities, including linguistics, on the other.[7]

KUHDEA and C'H are also quite different in a number of ways. C'H has a physical facility that is meant to be a gathering place where the Ahtna community can meet to share cultural activities and knowledge; the digital archive is just one part of the larger function of C'H and the AHF. KUHDEA, on the other hand, exists primarily as a virtual entity, with users accessing recordings and interacting with the director online, often from overseas locations. The intended audience for the two archives is different as well: KUHDEA primarily serves an academic audience, while the main audience for C'H is the Ahtna community, though both archives are welcoming to users from academia, the speaker community, and the general public. In addition, the two archives differ in their policies for allowing access to materials: while both archives allow the depositor to stipulate how freely available materials are, KUHDEA allows fairly liberal access online to open materials, while C'H users are intended to visit the physical facility to listen to recordings, and copies are distributed on a much more restricted basis.

History and Purpose

KUHDEA was founded in 2008 as a response to two needs: first, the growing interest in new technologies to assist humanities researchers in discovering and accessing extant research, and second, a need to provide long-term care for language documentation materials collected by researchers at UHM.[8] The first depositors to KUHDEA were long-time UHM linguistics professors and field linguists with a history of fieldwork in the Pacific, including Robert Blust,[9] Derek Bickerton,[10] and Al Schütz.[11] The Department of Linguistics has a five-decade reputation for field-based language work in the Pacific and Asia, and over the last half-century UHM scholars had throughout their careers collected analog recordings on reel-to-reel and cassette. Prior to the digital revolution there was no inexpensive, local, and readily available facility for

[4] http://www.language-archives.org/

[5] http://search.language-archives.org/index.html

[6] See, for example, Hale et al. 1992; Himmelmann 1998; Gippert et al. 2006.

[7] See, for instance, Boynton et al. 2006, 2010, as well as Electronic Metastructure for Endangered Languages Data (E-MELD).

[8] See Albarillo and Thieberger 2009 for a full description of the founding of KUHDEA.

[9] The Blust Collection can be accessed at http://scholarspace.manoa.hawaii.edu/handle/10125/7735.

[10] The Bickerton Collection can be accessed at http://scholarspace.manoa.hawaii.edu/handle/10125/4272.

[11] Other departments at UHM that have deposited materials in KUHDEA include Music, Anthropology, and the Charlene Sato Center for Pidgin, Creole, and Dialect Studies.

digitizing and storing the raw data upon which decades of published language descriptions and linguistic analyses had been based. Thus the founding director, Nick Thieberger, prioritized the digitization of these deteriorating materials and purchased for the archive a suite of digitization equipment including an analog-to-digital converter, a dual cassette deck, a turntable, DAT and minidisk players, a restored reel-to-reel player, monitoring headphones, and a desktop computer with audio-editing software. Thieberger is also a project manager at the Pacific and Regional Archive for Digital Sources in Endangered Cultures (http://paradisec.org.au) in Australia, upon which KUHDEA's digitization workflow, collection and access protocols, and metadata collection procedures are modeled.

Analog materials are digitized at the highest recommended standard for audio digitization (WAV format at 96kHz/24 bit) and stored in a dedicated collection on ScholarSpace,[12] the DSpace repository[13] of the University of Hawai'i Hamilton Library. The dedicated collection allows KUHDEA to collect custom metadata based on the recommendations of OLAC, and it provides an export to a static repository file for harvesting by the OLAC metadata harvester for publication on the web.

In the past few years KUHDEA has actively sought deposits from other researchers and projects associated with UH (for instance, the Hawaiian Sign Language documentation project [Lambrecht et al. 2013]), as well as from students in the graduate program in Language Documentation and Conservation in the Department of Linguistics. In fact, the careful preparation of archival materials from fieldwork and the regular deposit of those materials has become an integral part of the graduate curriculum.

While KUHDEA grew out of a need for an archive to serve a primarily academic community, C'H was built in response to community-based concerns over the protection and sense of ownership of an endangered heritage language. Since the early 1970s, the archiving of recorded and printed materials in and about Alaska's 20 indigenous languages has been primarily the responsibility of the Alaska Native Language Center and, later, the Alaska Native Language Archive (ANLA) at the University of Alaska Fairbanks.[14] However, after a series of village meetings, it became clear that local control of and access to Ahtna language materials was a high priority for members of the Ahtna community.

In 2009 the Ahtna Heritage Foundation stepped forward to assume responsibility for creating a local digital archive to digitize, store, and disseminate Ahtna language recordings. As AHF is a non-profit subsidiary of Ahtna, Incorporated, and is tasked with supporting and promoting Ahtna culture, language, and education, building and maintaining a local digital language archive is clearly in its purview.[15]

---

[12] http://scholarspace.manoa.hawaii.edu/

[13] http://www.dspace.org/

[14] See Berez et al. 2012 for a summary of the history of language archiving in Alaska, as well as a full description of the development of C'H.

[15] Ahtna, Inc., is the Alaska Native Regional Corporation serving the Ahtna region, established in 1971 under the Alaska Native Claims Settlement Act by the United States Congress. See http://www.ahtna-inc.com/ and http://ahtnaheritagefoundation.com/.

AHF applied for, and received, some US $149,746 in grant monies from the Institute for Museum and Library Services, which offers programs for libraries run by federally recognized Native American Tribes.[16] The first step in the building of C'H was to bring together collections of Ahtna recordings from around the country and archive them according to current best practice standards. Tapes were collected from a wide range of sources. The ANLA recording collection of traditional and personal narratives, ecological knowledge, word lists, and research on Ahtna grammar was acquired. Also acquired were collections of recordings from various government and local projects (such as land use interviews in advance of building an Air Force radar facility and historical site interviews conducted under the Alaska Native Claims Settlement Act) and a wealth of tapes in the personal collection of Ahtna individuals (such as potlatch recordings, and so forth).

Once the collections were amassed, AHF purchased digitization equipment based on the suite used at KUHDEA, hired a small staff, and brought in a trainer—incidentally a then-student employee of KUHDEA at the University of Hawai'i—to instruct the staff in digitization procedures, metadata collection, and safe on- and off-site file back up. The catalog is currently a FileMaker Pro relational database that exports its metadata to the OLAC harvester, and files are redundantly backed up both on-site and on servers at the University of Alaska Fairbanks.

Contents and Audience

KUHDEA aims to serve a wide audience ranging from academics to language community members to the general public, although since it is a university-based archive, most of its users are probably academic researchers, including the depositors themselves who wish to access their own collections online. At the time of writing, KUHDEA contains 615 items in 11 collections spanning a geographic region that includes Melanesia, Micronesia, Polynesia, southeast Asia, and China.

The exclusively linguistic content of KUHDEA is of interest to specialists and generalists in oral tradition.[17] Most of the items within KUHDEA are in indigenous languages of Asia and the Pacific, including Marshallese (Marshal Islands), Enga and Kuman (Papua New Guinea), Baba Malay (Singapore), Waima'a (Timor-Leste), Nyarong Minyag (China), and Thao (Taiwan), among others. The items themselves are personal and traditional narratives, conversation, poetry, oratory, and other genres of spontaneous language use. In addition, a great deal of metalinguistic knowledge such as lexical and grammatical information—of interest particularly to descriptive and theoretical linguists—is also recorded in KUHDEA.

Being curated by ScholarSpace has provided some interesting challenges for access. It is generally accepted among endangered language archives that endangered and minority language recordings can be considered to be more sensitive than other types of research data (see, for

[16] http://www.imls.gov/applicants/institution.aspx

[17] However, since the first and only collection of Old Hawai'i Sign documentation is held in KUHDEA, it is perhaps better to say the archive is of interest to specialists in culturally-specific traditions of language use within many genres from many communities.

example, Conathan 2011; Nathan 2011). A recording may contain information that is secret or sacred, it may contain slanderous remarks, or it may contain personal information that is difficult to anonymize. Many dedicated endangered language repositories provide different levels of access to different users (such as the Endangered Language Archive at the School of Oriental and African Studies in London, PARADISEC, and the Archive of the Indigenous Languages of Latin America at the University of Texas), and a user may have to "prove" her affiliation with a tribe or status as a *bona fide* researcher before being allowed access to a recording. Digital libraries, on the other hand, are moving toward providing fully open access, especially for scientific data, and Hamilton Library at UHM has been among the first American university libraries to adopt an open-access policy.[18] This interest in completely open access could potentially pose a problem for a small language archive with special sensitivity needs but without funding to build its own digital infrastructure for custom access.

Fortunately, the administration at ScholarSpace has been extremely supportive of KUHDEA's efforts to make some collections closed to the general public, according to the wishes of the depositor. Currently, users must request access to closed collections from the director via email, who then contacts the depositor. In the future KUHDEA hopes to build in an automated system for handling graded access, but currently DSpace support is not robust. Fortunately, most collections that are not those of current and recent graduate students in the Department of Linguistics allow open online access to the public.

As for C'H, the Ahtna archive of course targets materials specifically relating to the Ahtna community. While a good deal of this material is in the Ahtna language, C'H is not exclusively a language archive, and it contains many English-language items of a historical and anthropological nature. Contributions from linguist Jim Kari, anthropologist Frederica de Laguna, and myself are probably the most linguistically-oriented items in the collection. As the focus of these is Ahtna-language narrative and metalinguistic knowledge, these are the items that are most of interest to oral tradition specialists. However, a great deal of other information (business meetings and interviews, for example) is held in C'H that is of interest to historians, ethnographers, geographers, and anthropologists, as well as to members of the Ahtna community and other Alaska Native or Native American communities.

Audience access is handled differently at C'H from how it is handled at KUHDEA. The majority of items in C'H are potentially sensitive legacy items, most recorded in the past without a clear documentation of informed consent. Access to these items is handled on a case-by-case basis with regard for the sensitivity of the information contained in the recordings: for some of these items, the C'H staff must contact the families of the speakers on the recordings to establish access rights, a process that can sometimes be troublesome (see Berez et al. 2012 for examples of such difficulties). In addition, potential users must currently visit the C'H facility in Alaska in person to listen to recordings, and they must also declare their affiliation as (a) members of the Ahtna community, (b) *bona fide* researchers with an established relationship to AHF, or (c) members of the general public. At the time of writing, no recordings may be removed from the premises, though the C'H staff is currently working to ease these hardships on potential users while at the same time protecting sensitive materials. Possible solutions include the building of

[18] http://library.manoa.hawaii.edu/about/scholcom/oaatuhm.html

workstations in other locations within the Ahtna region to make recordings more accessible in other villages, as well as the establishment of a formal research agreement between AHF and any outside researcher wishing to work in the Ahtna region (Czaykowska-Higgins 2009; Cheecham and Wilhelm 2013). Similar problems face other indigenous language archives with a great deal of legacy material, and progress is necessarily slow.

Value for Oral Tradition Research

Currently, KUHDEA's highest value for oral tradition research is to its student depositors. The Department of Linguistics runs one of only a handful of graduate programs worldwide with a concentration in Language Documentation and Conservation (LD&C). Core courses in the LD&C program advocate the use of digital technologies to create long lasting records of language, and learning how to archive one's materials is a key component of the curriculum. Students are encouraged to deposit materials early in their careers rather than later.

At some point during their time in graduate school, most LD&C students will conduct several months of field-based research, which includes recording, transcribing, and translating oral language. These students usually go on to write descriptive or theoretical theses and dissertations, with examples of linguistic phenomena coming directly from field data they collect themselves. The students are encouraged to deposit recordings and transcriptions in KUHDEA either remotely during fieldwork or immediately upon their return to Honolulu, well before graduation or before the data is thoroughly analyzed. Not only does KUHDEA serve as a backup for the precious data upon which student research is based, but it also allows students to cite examples back to raw data via permanent handles. This ability ostensibly allows readers to check theoretical or structural claims, increasing falsifiability and improving the quality of scholarship in linguistics (see, for example, Himmelmann 2012). Students can of course elect to keep data in the archive private until some predetermined time, for example five years after publication of the dissertation.

C'H serves a rather different kind of research into oral history. Most of the users of C'H are members of the Ahtna community, interested in learning or reinforcing their knowledge of traditional Ahtna language and culture. As an example of how C'H can mobilize their collection to serve the local community, in August 2013 C'H held a weeklong "Breath of Life" workshop. Breath of Life workshops started in 1992 at the University of California, Berkeley, as a way to assist Native Americans with accessing linguistic materials about their languages that are held in archives at the university (see Hinton 2011). Other Breath of Life workshops have since followed (including a national Breath of Life at the National Anthropological Archives in Washington, DC, in 2013) and the model is usually the same: learners team up with linguist-mentors to learn to access published and unpublished materials in the archives.

At the Ahtna Breath of Life, eight learners joined five linguists and four Ahtna Elders to explore the digital collection, learn Ahtna grammar basics (parts of speech, nominal possession, person/number paradigms, and verb structure), and discuss related topics like self-motivation and language identity. Perhaps the most valuable skill gained by the participants was how to use the densely informational Ahtna dictionary (Kari 1990).

Oral literature materials in the archive played a central role in some of the final projects. As an example, one student who is an adult learner of Ahtna wanted to practice developing her transcription skills; that is, she wanted to learn to listen to a recording of spoken Ahtna and be able to at least write down what was being said, so that she could then take the transcription to an Ahtna-speaking elder for help with translating the passage into English. She selected from the archive a recording of a Raven Story told by the late Mildred Buck, a woman who recited the story slowly and clearly, and was known for her skills in Ahtna language teaching. The student, who is familiar with Ahtna orthography (although not an expert) learned to play the recording repeatedly in a piece of linguistic software called TranscriberAG[19] in order to write down what she heard the best she could. After several days and after checking her transcription with linguists and elders, she ended up with a reasonably accurate transcription. Not only could she then work on translating the recording, but she now has much more confidence in her ability to transcribe more of the untranscribed recordings in the archive. A second Breath of Life workshop took place in late 2013.

Summary

KUHDEA and C'H represent two smaller language archives that serve audiences interested in preserving and mobilizing digital records of oral tradition, albeit in two very different realms. While the former is based in a university and serves a mostly academic audience, and the latter is administered by an Alaska Native community primarily to provide access to the Ahtna people, both work within current best practice standards for endangered language digitization, preservation, and access.

University of Hawai'i at Mānoa

References

Albarillo and Thieberger 2009 — Emily E. Albarillo and Nick Thieberger. "Kaipuleohone, University of Hawai'i's Digital Ethnographic Archive." *Language Documentation and Conservation*, 3:1-14.

Berez et al. 2012 — Andrea L. Berez, Taña Finnesand, and Karen Linnell. "*C'ek'aedi Hwnax,* The Ahtna Regional Linguistic and Ethnographic Archive." *Language Documentation and Conservation*, 6:237-52.

Boynton et al. 2006 — Jessica Boynton, Steven Moran, Helen Aristar-Dry, and Anthony Aristar. "E-MELD and the School of Best Practices: An Ongoing Community Effort." In

[19] http://transag.sourceforge.net/

Sustainable Data From Digital Fieldwork. Ed. by Linda Barwick and Nicholas Thieberger. Sydney: University of Sydney Press. pp. 87-97.

Boynton et al. 2010 _____. "Using the E-MELD School of Best Practices to Create Lasting Digital Documentation." In *Language Documentation: Practice and Values.* Ed. by Lenore A. Grenoble and N. Louanna Furbee. Amsterdam and Philadelphia: John Benjamins. pp. 133-46.

Cheecham and Wilhelm 2013 Connie Cheecham and Andrea Wilhelm. "Sharing Worlds of Knowledge: Research Protocols for Communities." Paper presented at the 3rd International Conference on Language Documentation and Conservation, Honolulu. http://scholarspace.manoa.hawaii.edu/handle/10125/26110

Conathan 2011 Lisa Conathan. "Archiving and Language Documentation." In *The Cambridge Handbook of Endangered Languages.* Ed. by Peter K. Austin and Julia Sallabank. Cambridge: Cambridge University Press. pp. 235-54.

Czaykowska-Higgins 2009 Ewa Czaykowska-Higgins. "Research Models, Community Engagement, and Linguistic Fieldwork: Reflections on Working within Canadian Indigenous Communities." *Language Documentation and Conservation*, 3.1:15-50.

Gippert et al. 2006 Jost Gippert, Nikolaus P. Himmelmann, and Ulrike Mosel, eds. *Essentials of Language Documentation.* Trends in Linguistics Studies and Monographs, 178. Berlin: Mouton de Gruyter.

Hale et al. 1992 Ken Hale, Michael Krauss, Lucille J. Watahomigie, Akira Y. Yamamoto, Colette Craig, LaVerne Masayesva Jeanne, and Nora C. England. "Endangered Languages." *Language*, 68.1:1-42.

Himmelmann 1998 Nikolaus P. Himmelmann. "Documentary and Descriptive Linguistics." *Linguistics*, 36:161-95.

Himmelmann 2012 _____. "Linguistic Data Types and the Interface Between Language Documentation and Description." *Language Documentation and Conservation*, 6:187-207.

Hinton 2011 Leanne Hinton. "Revitalization of Endangered Languages." In *The Cambridge Handbook of Endangered Languages.* Ed. by Peter K. Austin and Julia Sallabank. Cambridge: Cambridge University Press. pp. 291-311.

Kari 1990 James Kari. *Ahtna Athabaskan Dictionary.* Fairbanks: Alaska Native Language Center.

Lambrecht et al. 2013 — Linda Lambrecht, Barbara Earth, and James Woodward. “History and Documentation of Hawai‘i Sign Language: First Report.” Paper presented at the 3rd International Conference on Language Documentation and Conservation, Honolulu. http://scholarspace.manoa.hawaii.edu/handle/10125/26133.

Nathan 2011 — David Nathan. “Digital Archiving.” In *The Cambridge Handbook of Endangered Languages.* Ed. by Peter K. Austin and Julia Sallabank. Cambridge: Cambridge University Press. pp. 255-73.

This page is intentionally left blank.