University Library



A gateway to Melbourne's research publications

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Albarillo, EE;THIEBERGER, N

Title:
Kaipuleohone, The University of Hawai'i's Digital Ethnographic Archive

Date:
2009

Citation:
Albarillo, E. E. & THIEBERGER, N. (2009). Kaipuleohone, The University of Hawai'i's Digital Ethnographic Archive. Language Documentation and Conservation, 3 (1), pp.1-14

Persistent Link:
https://hdl.handle.net/11343/282453

License:
CC BY-NC-SA

LANGUAGE DOCUMENTATION & CONSERVATION

Vol. 3, No.1 (June 2009), pp. 1-14
http://nflrc.hawaii.edu/ldc/
http://hdl.handle.net/10125/4422

Kaipuleohone, the University of Hawai'i's Digital Ethnographic Archive

Emily E. Albarillo and Nick Thieberger
University of Hawai'i at Mānoa

The University of Hawai'i's Kaipuleohone Digital Ethnographic Archive was created in 2008 as part of the ongoing language documentation initiative of the Department of Linguistics. The archive is a repository for linguistic and ethnographic data gathered by linguists, anthropologists, ethnomusicologists, and others. Over the past year, the archive has grown from idea to reality, due to the hard work of faculty and students, as well as support from inside and outside the Department. This paper will outline the context for digital archiving and provide an overview of the development of Kaipuleohone, examining both concrete and theoretical issues that have been addressed along the way. The creation of the archive has not been problem-free and the archive itself is an ongoing process rather than a finished product. We hope that this paper will be useful to scholars and language workers in other areas who are considering setting up their own digital archive.

1. INTRODUCTION: THE CONTEXT[1]**.** The use of new technologies in scholarship—known as EScience (UK), EResearch (Australia) or Cyberinfrastructure (USA)—has led to considerable interest in digital tools that can assist scholars in discovering and accessing existing research, as well as converting analog records into digital form, and producing new digital data from present research projects (see Borgman 2007). Linguists need to participate in these larger initiatives in the humanities that are leading to the development of the necessary infrastructure for long-term housing of our research outputs. The UK report titled "Keeping Research Data Safe" (Beagrie, Chruszcz, and Lavoie 2008) recommends that higher education institutions (HEIs) "should consider federated structures for local data storage within their institution comprising data stores at the departmental level and additional storage and services at the institutional level. These should be mixed with external shared services or national provision as required. HEIs should work with and utilize national and international disciplinary data archives where these exist. The hierarchy of data stores should reflect the detailed nature of the content, services required, and the changing nature of its importance over time" (7). Similarly the visionary report "Our Cultural Commonwealth" (American Council of Learned Societies 2006:38) states that "Extensive and reusable digital collections are at the core of the humanities and social science cyberinfrastructure. Scholars must be engaged in the development of these collections.

[1] Kaipuleohone has received the support of a number of people and we are happy to have this opportunity to acknowledge the help of Linda Barwick (PARADISEC), Kurt Brunner, Joseph O'Mealy (Interim Dean of the College of Languages, Linguistics and Literature, University of Hawai'i at Mānoa), Beth Tillinghast, and Daniel Ishimitsu (Hamilton Library).

Licensed under Creative Commons
Attribution Non-Commercial Share Alike License
E-ISSN 1934-5275

[…] The extensive digitization of cultural heritage materials is one of the most exciting developments in the humanities and social sciences in the past century." We agree with these recommendations and have accordingly built the collection using linguistic expertise to identify and target important research data, while relying on existing infrastructure to house and curate the collection over time.

A further and most compelling factor promoting the safe curation of linguistic research data is that, in some cases, it may be the only recording of a language. Speakers of this language and their descendants will look to linguists for access to these recordings, and without a dedicated archive it is unlikely that the records will be available in any accessible form. As pointed out by a student in our department and a speaker of Pollapese (Micronesia), "several questions still play on people's minds such as, 'Whatever happens to the previous research? What benefits were they to our community? Were researchers who seek to get their PhDs merely exploiting us or was it for a greater good? When do we see products and results and not more study?'" (Yourupi 2008:4).

2. BUILDING KAIPULEOHONE[2]. The recent past has seen a dramatic rise in interest in language documentation—that is, in the recording of various aspects of under-studied or endangered languages. This documentation is justified in many ways: languages are disappearing due to globalization; languages provide essential data to linguists; languages are a central part of a community's identity and culture, and so forth. Whatever the reasons or goals, whether community-oriented or linguist-driven, a key product of any language documentation project is raw data: recordings of stories, conversations, rituals, and speeches in the language itself. For any documentation project to be of lasting value, this data—whether analog or digital, and whether text, audio, or video—must be preserved and made accessible, and this is where archives such as Kaipuleohone are extremely valuable (see Johnson 2004 for a general discussion of the role of linguistic archives).

There are a number of digital archives worldwide that are preserving endangered language data: DELAMAN[3] has twenty-one member archives and OLAC[4] lists thirty-six participating archives (but only ten of these have been updated between 2007 and 2009). Kaipuleohone is a necessary addition because it will serve the needs of researchers at the University of Hawai'i (UH) who have already amassed an impressive collection of analog material and who are in the process of creating large amounts of digital data as well. Kaipuleohone is in a unique position to serve these researchers by providing for the twin needs of digitization and preservation of their data. Additionally, the archive will enable people outside Hawai'i to access valuable language data from the region via the Internet, subject to access conditions that accompany each item in the collection.

---

[2] *Kaipuleohone* (Ka-ipu-leo-hone) means a 'gourd of sweet words' in Hawaiian. We are very grateful to Laiana Wong for suggesting this name, and for allowing us to use it as the name of this archive. The archive's website is: http://www.ling.hawaii.edu/langdoc/archive.html.

[3] Digital Endangered Languages and Musics Archives Network, http://delaman.org

[4] Open Language Archives Community, http://www.language-archives.org

Kaipuleohone is modeled on PARADISEC[5]—building on an existing archive means that we don't need to reinvent the wheel but instead can focus on the unique needs of researchers at UH and the communities they are working with. PARADISEC offers a simple model that, once established, requires minimal maintenance. Files are accessioned with a simple filenaming convention and no attempt is made to provide elaborated interfaces to the data itself in the way that, for example, IMDI does for its archives (e.g., DoBeS[6] and its network of regional archives[7]). This could occur in the future, but given the limited personnel and resources available, we have chosen to start small and build up if and when possible. The current interface to the data is given by the ScholarSpace system and so required no special development for our project.

Like many similar projects, we had little in the way of resources to devote to the archive, so we set immediate goals that were achievable, identifying where we needed to buy equipment and where we could negotiate to use existing systems. As already mentioned, the main goals of Kaipuleohone are to provide for the digitization of older language and ethnographic material as well as to preserve and allow access to both digitized and born-digital material, not just for linguists but also for language communities. Therefore, the discussion that follows will be divided into four major sections: digitization, preservation, access, and collaboration.

3. DIGITIZATION. One of the first priorities as we started work on the archive was to establish a digitization suite: a collection of high-quality equipment that would allow us to digitize older recordings and field notes from faculty and student researchers at the University of Hawai'i. Digitization was given priority because cassette and reel-to-reel recordings were deteriorating and needed to be dealt with as soon as possible. Three processes took place, sometimes simultaneously: finding material that needed to be digitized, ordering and installing equipment, and establishing workflows for processing material with that equipment.

In assessing digitization needs, we started with Department of Linguistics faculty, talking to them about any audio recordings or field notes they would like digitized and archived. We kept records of the type of material, length, language, and content for everything available. Students who were currently working on endangered or understudied languages were also consulted, especially those who were working on their own native languages. Finally, we began to contact other departments, including ethnomusicology, anthropology, and the Sato Center for Pidgin and Creole Studies, to see if they were interested in contributing to the archive. Members of all these groups—linguistics faculty, students, and other departments, indicated interest in having material digitized, and as we obtained equipment we began to borrow and digitize their material. Though this search for material

---

[5] Pacific and Regional Archive for Digital Sources in Endangered Cultures (PARADISEC). http://paradisec.org.au.

[6] http://www.mpi.nl/DOBES/archive_info/archive_tasks

[7] http://www.mpi.nl/DOBES/regional_archives/

was an especially prominent feature of the beginning of our work on the archive, we will continue looking for material to digitize in the future, with the knowledge that much of the valuable linguistic and ethnographic data recorded over the past few decades may be sitting in offices and desks, overlooked and slowly deteriorating. We hope that not only linguists but members of as many other departments as possible can take advantage of our equipment and funding to have their material digitized and preserved.

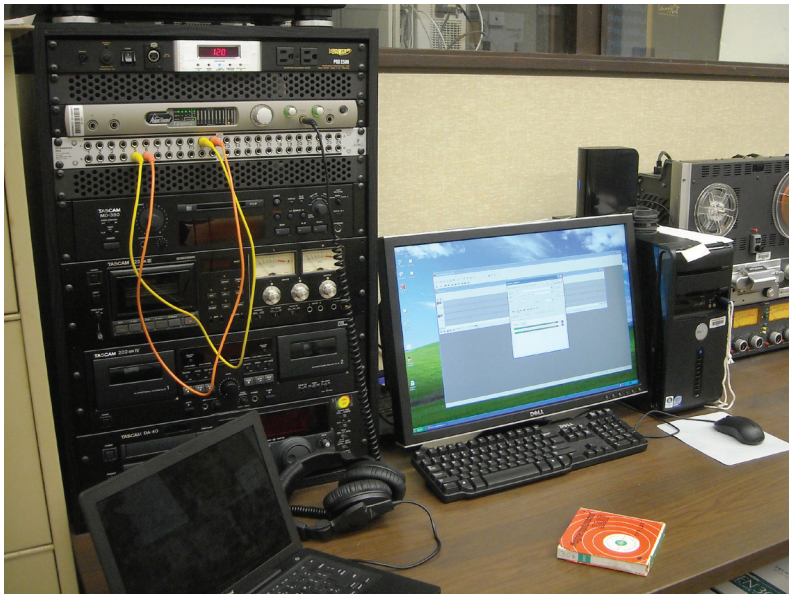


FIGURE 1: The digitization suite in April 2009. In the rack on the left is the power supply (top), then the Prism Orpheus analog/digital converter, a patch bay for patching between inputs (cassette or reel-to-reel), a Minidisc player, a cassette player, a cassette rewinder, and a DAT player. On the right is the Studer reel-to-reel player.

The equipment we bought included a Studer A810 reel-to-reel player, an audio rack with a cassette deck, a turntable, a minidisc player, and a DAT player, along with Prism Sound's Orpheus analog/digital converter, which ran the audio signal to a Dell computer equipped with Sony's Sound Forge 9 software. A pair of monitoring headphones and a 1 TB backup hard drive completed the initial audio setup. With most of the equipment in place, we developed a workflow for digitizing cassette tapes, based on the PARADISEC manual,[8] that could be followed by student workers and would result in quality digital versions of the analog material. The workflow (included as Appendix A) was written up as we began working through our first batch of cassette tapes, and was refined as we developed

[8] http://www.paradisec.org.au/pdsc-manual08.pdf

more experience. We chose to conform to recommended highest standards for audio digitization (based on IASA 2005) that were feasible at the time: WAV format at 96 kHz/24 bit. Files are saved on both the computer and the backup hard drive before being deposited into the longterm storage system (discussed below). We batch process the conversion of WAV files to .MP3, so we store both uncompressed and compressed versions of each file. We have a mass CD burner which can prepare CD copies of the archived media for return to depositors.

With the audio equipment functioning, we moved on to the question of how best to produce digital images of field notes. Background research involved looking at how PARADISEC accomplished this, discussing the issue with technicians in Hamilton Library's preservation department, and reviewing relevant information on the Internet. Scanning papers was one option, but we were concerned about the fragility of the papers and books we wanted to scan, so we ruled out a sheet-fed scanner, and the amount of handling that a flat-bed scanner would require made it more attractive to opt instead for a digital camera. In the end we ordered an Adorama RS-CS1070 copystand with two light arms to support our Canon Powershot G9 12.1 megapixel digital camera. We chose the copystand over other options, namely an invertible tripod, as it would allow for better lighting and a wider range of document sizes. This equipment was installed on a small wheeled table that provided room for a laptop to be attached to the camera via USB. The Canon Powershot G9 came packaged with software (RemoteCapture DC) that allowed for remote control of the camera via a computer. We realized later that there were some limitations to this situation: the camera couldn't shoot RAW (uncompressed) files when connected to a computer, so we had to make do with high-quality JPEGs, suitable for representations of files that would also be housed in an archive but not ideal as preservation copies. Despite these drawbacks, we feel the image digitization equipment is sufficient for our current goals, which are to make previously inaccessible material available in a readable digital form.

Developing a workflow for the imaging equipment was relatively straightforward (see Appendix B). Software settings that would result in images that were easily readable and reasonably sized (1.5 MB per page) were chosen through experimentation and discussion with PARADISEC and Hamilton Library employees. Since the material we will be digitizing is in many different forms—bound volumes, notebooks, loose paper—we left the workflow description as flexible as possible.

In terms of both audio and paper material, the Kaipuleohone archive has made significant progress. We have established a workflow for cassette processing, and student workers are in the process of digitizing various important collections of cassette tapes. We have taken images of some 2,600 items totaling 3.6 GB of digital data. In both cases, there is still work to be done and processes that can be improved, but we feel we are moving toward Kaipuleohone's goal of digitizing various formats of older language and ethnographic material (see Appendix C for list of digitized collections).

4. PRESERVATION. Even the best digitization program is of little value if the digital objects being produced are not properly preserved. One major argument for digitization is that the original analog materials are fragile and deteriorating, but digital objects can also deteriorate and become obsolete with time—often more quickly than we might expect. As Bird and Simons (2003:557) note, "In the very generation when the rate of language death

is at its peak, we have chosen to use moribund technologies, and to create endangered data." Unfortunately, the Kaipuleohone archive does not have guaranteed long-term funding, nor the resources required to build and maintain a digital repository, so we decided to look outside the Linguistics Department to UH's Hamilton Library to assure that the digital objects we are creating will be preserved into the future. We have negotiated an agreement to deposit our material in ScholarSpace,[9] UH's digital repository, based on DSpace.[10] This has created many challenges, as our media files are both larger and in formats different than those ScholarSpace has previously dealt with. The majority of our meetings with the ScholarSpace project manager have focused on how our metadata describing the digital objects will fit into the ScholarSpace catalog, and how they will be aligned with Open Language Archives Community (OLAC) metadata standards. As the metadata we have adopted conforms to the usual standards (Dublin Core and Open Archives Initiative are both part of the OLAC metadata set), we were confident that ScholarSpace metadata could be exported to an OLAC-compliant static repository.[11] The relationship with ScholarSpace has been critical to the ongoing preservation of the data. The success of small and domain-specific collections like ours can rely on establishing good relationships with mass data storage systems with a commitment to preserving valuable data into the future (as is also the case for PARADISEC's association with the National Computing Infrastructure system in Canberra). This commitment includes preservation and back-up, persistent identification and location of items in the collection, and migration of formats of data over time.

5. ACCESS. A major goal of the Kaipuleohone archive is to increase the accessibility of linguistic and ethnographic material that might otherwise be hidden away in a desk drawer or in a box in the back of an office. Methods of access, therefore, are vital. Many of these issues are just beginning to be addressed, because our initial focus was on preservation of the materials themselves. Now, however, with some 200 GB of digitized material and the agreement with ScholarSpace in place, we need to start exploring various ways the material can be made accessible to different communities.

Bird and Simons (2003) emphasize the importance of both discovery and access of language-related material. Discovery refers to how easily people can find out about the existence of material in a given language; access refers to how easily an interested individual can actually obtain a copy of the digital material. We hope to address these two concerns simultaneously by creating multiple access points that will facilitate discovery and make it easy to obtain digital files while respecting any access limitations requested by the depositor.

The first available access point will be the ScholarSpace catalog. ScholarSpace collections can be browsed or searched in a number of different ways, including via any Open Archives Initiative search engine (and Google). As mentioned earlier, we incorporate our

---

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

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

[11] As occurred in April 2009 when Kaipuleohone was registered as an OLAC repository.

metadata for each item into the ScholarSpace catalog, which will give users even more potential information to search or browse by. Ideally an item discovered in ScholarSpace will then link to the actual digital file that can be downloaded to the user's computer (of course, access to any item in the collection is subject to deposit conditions).

Kaipuleohone will also have its own catalog database, which will send metadata to ScholarSpace. Though much of the information in the two will be the same, the archive catalog will accommodate specific linguistic metadata better than the general structure of the ScholarSpace catalog allows, and will have a data entry screen more suitable for linguistic information, using drop-down menus to enforce consistency of data entry using controlled vocabularies (e.g., ISO-639 language names, names of people who are already in the database, names of roles associated with those people, data type names, and so on). The Kaipuleohone catalog is currently being developed based on PARADISEC's catalog. Before the two catalogs were functioning, metadata associated with digitized items was stored in spreadsheets that were imported into the database version.

Third, the metadata associated with the Kaipuleohone catalog is harvested by the Open Language Archives Community (OLAC) by following the OLAC Protocol for Metadata Harvesting. This enables material deposited with Kaipuleohone to be discovered through searching any OLAC compliant search site, and opens the archive up to a wider audience of linguists, other researchers, and the general public.

At the same time that we ensure that items in the Kaipuleohone archive are discoverable and accessible, we will also need to account for copyright issues and access restrictions on some material. We have created a deposit form that allows depositors to indicate any restrictions on access. When requesting copies of such material, users will be required to submit signed access forms, which will then allow ScholarSpace to provide access to just the items requested. In time we hope that this process can be automated to some extent, with a clickable agreement form providing access to all items whose deposit conditions allow such access.

6. COLLABORATION. Inter-departmental collaboration has played, and will continue to play, a vital role in the development of the Kaipuleohone archive. Though the archive is housed in the Department of Linguistics, it is meant to hold a wide range of ethnographic materials. Since the beginning of the project we have been in contact with various other departments in the humanities, offering to digitize, catalog, and store ethnographic material for them. The response has been positive. We are processing cassettes for the Music Department's ethnomusicology program, and a student worker from the Anthropology Department is learning to use our audio equipment to digitize a large collection of recorded folktales from the Solomon Islands. Additionally, we obtained a box of cassettes containing important recordings of Hawaiian Creole English from the Charlene Sato Center for Pidgin, Creole, and Dialect Studies (see Appendix C for details of collections being digitized). Finally, we were able to come to an agreement with the University of Hawai'i Archives, which has been storing Derek Bickerton's Hawaiian Creole English material, a very significant collection that has never been digitized or even fully cataloged. Through this agreement, we will borrow the tapes in batches, carefully digitize them, and return them to the Archives. All of these partnerships are mutually beneficial arrangements—other departments are helping us populate our archive and demonstrate the relevance of

our project, while we are helping them preserve material that they might not have the time or resources to care for.

The collaboration between Kaipuleohone and ScholarSpace has also been productive. ScholarSpace is looking to increase its holdings and its ability to deal with new file formats, while Kaipuleohone needs a secure place to deposit digitized and born-digital material. An institutional repository, like ScholarSpace, also provides continuity and guarantees preservation of the collection if the archiving project itself ceases to function (for example due to a lack of funding or the retirement of key personnel).

7. CONCLUSION. Building an archive like Kaipuleohone involves a number of disparate activities, but of central importance is the development of working relationships with local agencies (the library, technical support, like-minded projects who can provide expertise), depositors, other archives in the OLAC and DELAMAN networks, regional cultural centers who have an interest in the content of the collection, and students who are eager to learn about new methods that will support their research. Starting small and building the collection together with developing these relationships has been central to the development of the archive and will be critical to its ongoing existence. Selecting, buying, installing, and running equipment require local technical support, as does the choice of a metadata schema and a database in which to store the collection's metadata. This kind of repository will become increasingly common, driven by a local and discipline-specific need (in this case UH linguistics), using broader established resources (UH Library systems), and tied into an international framework (OLAC, OAI, and Dublin Core) to ensure interoperability of the resulting systems. Through all of this, the archive will ensure greater accountability for researchers whose analysis can be based on citable archival data, as well as allowing discovery of and access to heritage recordings for those speakers recorded and their descendants.

Appendix A: Kaipuleohone analog tape digitization workflow

(This document is updated periodically and is available from the Kaipuleohone website)

In general:

- Please turn your cell phone off so it doesn't interfere with the audio signal.
- Never place audio tapes near anything magnetic (this includes computers!).
- Recording should always be at 96 kHz, 24 bit.
- Tapes should be stored upright rather than flat on their side.
- Check the tape for signs of deterioration—any breaks, any visible mold, any vinegar smell? If so, the tape needs special care and should not be played.

Starting the equipment:

- Turn on computer and monitor.
- Select "Minimal Recording Mode" when given the option.
- Turn on main power switch at top of equipment rack, and power switch for the top cassette tape deck.
- Clean the heads on tape deck before each session with isopropyl alcohol and a cotton swab (located behind computer monitor).

Digitizing a cassette tape:

- Open Sony Sound Forge and click the red "bull's-eye" record button—a window will appear showing sounds levels.
- Check for 96 kHz/24 bit setting.
- Be sure to have a proper name worked out for the digitized file (see the "Naming files" section below).
- Insert tape into the top deck, fast forward entire side, rewind entire side, begin playing tape.
- As the tape begins to play, adjust the "output" dial on tape deck until levels are within an acceptable range (peaking between -12 and -18).
- Stop and rewind tape, click the red "bull's-eye" record button from the monitoring window, then wait a few seconds before pressing play on tape deck.
- When the entire side of the tape is recorded, click stop (square) button, and click "close." The monitoring window will disappear and the program will open the waveform from the recording.
- Delete any excessive silence before and after the recorded material but leave a few seconds at each end. Don't make any changes to the sound file beyond this.
- Select "Save as" from the File menu, name the file (see "Naming files" below) and save it as a WAV file in C:/Administrator/Documents and Settings/My Documents/New Wavs.
- Once both sides are digitized and saved, fill out a cassette slip with your initials, the date, and the file names; fold and insert this slip in the case of the completed cassette.

Monitoring the recording:

- Plug headphones into the cassette deck headphone jack (labeled "phones") to monitor sound as it records. Adjust headphone volume with small dial, not output dial.
- Plug headphones into the Orpheus headphones jack (labeled "phones 1") to monitor sound from the computer. Adjust volume with small dial next to jack.
- The recording need not be monitored continuously, but should be checked frequently.

Naming files:

- File names consist of three units separated by hyphens plus the file extension (.wav). The first unit identifies the collection, for example "JB1" for the Joel Bradshaw collection. The second unit is the item number; this will generally correspond to the tape number. The third unit refers to the part of the item, for example, "A" for side A. Therefore, "JB1-002-B.wav" would be the proper file name for side B of the second tape in the Joel Bradshaw collection.
- In the metadata spreadsheet, make a note of the filename(s) assigned to the tape (see "Entering metadata" below).

Entering metadata:

- Metadata should be entered into the appropriate spreadsheet at the same time a tape is digitized, rather than left for later.
- All information from the cassette itself and the case should be recorded onto the metadata spreadsheet.
- If the tape has quality issues (distorted playback or strange sections of silence), if only one side is recorded, or if a side is not fully recorded (i.e. only the first ten minutes of a 45 minute side contain sound), this should be noted in the metadata.

Appendix B: Kaipuleohone fieldnote digitization workflow

(This document is updated periodically and is available from the Kaipuleohone website)

Setting up:

- Open the side shelf on the imaging table (make sure it's secure), set up laptop and plug into power strip below table; ensure the computer won't enter sleep mode while you work.
- Attach the power supply to the camera (attaches inside the battery compartment).
- Attach the camera to the copystand column—secure it but don't over-tighten; center the camera as much as possible over the base. The camera attachment should be about 26cm above the base.
- Run the USB cable from the camera to the laptop; turn the camera on.
- Open RemoteCapture DC program on the laptop.
- Turn on the copy lights (there's a switch below each light).

RemoteCapture DC settings should be as follows:

- Image quality: Medium1/Fine
- Focusing point: Automatic selection
- Macro: Off
- AF-assist light: Off
- AF Operation: AF unlock
- Flash: Off
- White balance: Tungsten
- Exposure compensation: +2/3
- Flash exposure level: 0
- Metering mode: Spot
- ISO speed: Auto
- AE Mode: Aperture-Priority AE; Av: 8.0

Digitizing a document:

- Center the document horizontally on the copystand base.
- Using RemoteCapture DC, zoom in enough to leave an even band of grey around all edges of document.
- In the File menu, select "Preferences" and set the file naming as needed.
- Click the focus button then the release button (or command-R) in RemoteCapture DC to take the picture.
- If writing shows through from the back of a sheet or if the paper is very thin, try correcting this by putting a plain white sheet behind it.
- If the camera has trouble focusing on a nearly blank page, replace it with a different page, select "AF lock" to focus, then without refocusing take an image of the nearly blank page.
- After photographing each document, open the files in an image viewer to ensure they are clear and even.

- When a document is finished, fill out a digitization slip with your initials, the date, and the file names; attach this slip to the completed document.
- At the end of a digitizing session, copy all new files from the laptop onto the hard disk ready for transport to ScholarSpace.

Naming files:

- File names consist of three units separated by hyphens plus the file extension (.jpg). The first unit identifies the collection, for example "JB1" for the Joel Bradshaw collection. The second unit is the item number; this will generally correspond to the a complete document. The third unit refers to the page of the document, for example, "023" for page 23. Therefore, "JB1-002-003.jpg" would be the proper file name for the third page of the second document in the Joel Bradshaw collection.
- In the metadata spreadsheet, make a note of the filename(s) assigned to the document (see "Entering metadata" below).

Entering metadata:

- Metadata should be entered into the appropriate spreadsheet at the same time a document is digitized, rather than left for later.
- Enter as much information about the document as possible into the metadata.
- Make note of any problems encountered during digitization in the metadata.

Appendix C: Kaipuleohone collections

Fully digitized:

- 6 cassettes of Numbami (Papua New Guinea) material from Joel Bradshaw
- 8 cassettes of Borneo and Manus island language survey material from Robert Blust
- Notes and maps from surveys of over 100 dialects of Fijian from Albert Schütz
- Images of a manuscript Pohnpeian dictionary produced in 1950 (230 pages)
- 3 cassettes of older Truku (Taiwan) material from PhD student Apay Tang
- 9 cassettes of thesis/dissertation material from the Ethnomusicology Department
- 12 cassettes in Wuvulu (PNG) from PhD student James Hafford

In process:

- Approximately 300 cassettes of Hawaiian Creole English material from Derek Bickerton
- Approximately 150 cassettes of Hawaiian Creole English material from the Sato Center
- Cassettes of Solomon Islands folktales in Cheke Holo from Geoff White in the Anthropology Department at UH.
- 50 reel-to-reel tapes of Fijian dialect survey material from Albert Schütz

Not yet digitized:

- 54 reel-to-reel tapes of thesis/dissertation material from the Ethnomusicology Department

References

American Council of Learned Societies. 2006. *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on Cyberinfrastructure for the Humanities and Social Sciences*. New York: American Council of Learned Societies.

Beagrie, Neil, Julia Chruszcz, and Brian Lavoie. 2008. *Keeping research data safe: A cost model and guidance for UK universities*. London: Higher Education Funding Council for England.

Bird, Steven, and Gary Simons. 2003. Seven dimensions of portability for language documentation and description. *Language* 79(3):557–582.

Borgman, Christine. 2007. *Scholarship in the digital age*. Cambridge, Massachusetts: MIT Press.

IASA (International Association of Sound and Audiovisual Archives). 2005. Standards, recommended practices and strategies: The safeguarding of the audio heritage: Ethics, principles and preservation strategy (IASA-TC 03). South Africa: IASA Technical Committee. http://www.iasa-web.org/special_publications.asp

Johnson, Heidi. 2004. Language documentation and archiving, or how to build a better corpus. In *Language documentation and description*, Volume 4, ed. by Peter K. Austin, 140–153. London: Hans Rausing Endangered Languages Project.

Yourupi, Paulina. 2008. Relevance of archiving to a Chuukese audience. Unpublished manuscript, University of Hawai'i at Mānoa Department of Linguistics.