Digitizing the Audio Archive of Linguistic Fieldwork at the Berkeley Language Center

> Mark Kaiser, Associate Director with the help of Marianne Garner, Librarian Susan Stone, Programmer

Historical Background

- Fieldwork began in the 1950's
- Collection at the UCB Language Lab dates from the 1960's
- 1979-1980 UCB Language Lab received NEH funding to organize the archive
 - Archive mastered to reel-to-reel tape
 - Paper catalog created



Examples of the various media on which original materials are deposited (various reel-to-reel formats, magnetic wire, phonographs)



Storage of the original materials (left) and the mastered reel-to-reel tapes (right)

State of the Collection Late 1990's

- Factors influencing decision to digitize
 - Exercising the masters
 - Tape and tape machine manufacturers were leaving the business
 - Limited access to information about the archive
 - Distribution was limited
 - Patrons' needs

Technical Issues

- File format
 - Archive: 24 bit / 96 kHz, uncompressed .wav
 - Production: 16 bit / 44 kHz, uncmp. .wav
 - Web: Windows Media Audio (.wma)
- "Master" and backup
 - Hard drive array
 - Tape and DVD, 2 copies each

Legacy Permission Forms

- Collectors sign off on extent of access, contractual arrangement
- Three interest groups with the potential for disparate opinions
- Permission forms did not anticipate
 Internet
- Permission forms did not anticipate repackaging media
- In many cases original parties now deceased

Resegmentation

- Segmentation of the analog recordings was based on time constraints of media, done by audio engineers
- Segmentation of the digital recordings is based on the content of the media, done by graduate students in linguistics

Item-level Cataloging

Creation of segment titles

- Used collector's data when available
- Graduate students apply a limited uniform set of descriptors
- Searchable in the online catalog
- Modification of content type terms

 Controlled vocabulary, mappable to OLAC
- Lessons learned

Conforming to OLAC Standards

- OLAC as a moving target
- Methodology to conform
 - Database mapping tables where our vocabulary does not conform to OLAC's
 - Personal roles, content type terms, language names

 Java program generates an OLAC static repository, using SQL database tables from OLAC and a Sybase stored procedure in our database

Current Status

- Online catalog is up and running
 - Currently contains a mix of batch-generated data from the old catalog and newly digitized items together with the linked audio files
 - We are 6 months into the grant and on schedule to complete digitization and cataloging on time
 - We will register our collection with OLAC shortly

😥 BLC Audio Archive of Linguistic Fieldwork - Mazilla Firefox	
Bie Edz Bee So Bookmarks Look Help	
🐢 • 🥼 • 🚰 💿 🕎 📾 http://www.mp.berkeley.edu/bk/fa/templauery_demo.html	- 0 % K.
Oustomae Units 🔄 Pree Hotmail 🗋 Google Search 🔄 Google.com 🗋 windows Medie 📑 Windows	
Search the Meatalog Berkeley Linguage Center	
Audio Archive of Linguistic Fieldwork Enter the terms you wish to search for in the form below, then click on "Submit Search."	
Native American Language Native American Language Family Other Language	
Laibeta Makah Makah Makah Masa	
Collector	
Lingustic Consultant	
Title	
Content Type: Stores Linguistic data Ethnographic data Songu/Charlas	
Type of Search	
Submit Search Cear To the BLC Collections Page	
Note: Cataloging is still in progress for this collection	
Done	

http://blc.berkeley.edu

```
UC Berkeley, BLC Audio Archive of Linguistic Fieldwork - Mazilla Firefox
                                                                                                                                                                                                                                                                                                                                                             Elle Edit View Go Bookmarks Tools Help
                                                                                                                                                                                                                                                                                                   - O to K.
  G . D . M
                                                Res and the standard stand
Standard stand
Standard stan
Standar
     Customize Links 🔄 Free Hotmail 🔄 Google Search 🔄 Google.com 📘 Windows Media 🔄 Windows
 Title:
                                  Maidu / collected by William L. Shipley; spoken by Roxie
                                                                                                                                                                                                                                                                                                                                                                           .
                                  Peconom et al.
 Series: Survey of California and Other Indian Languages
 Language: Maidu, Northeast
                                  1955-1957
 Date:
 Fublisher: University of California, Berkeley, Berkeley Language Center
 Format: Sound recording
 Collectors:
                Shipley, William L.
 Linguistic Consultants:
                Peconom, Roxie
                Peconom, George
                Morales, Leone
                Williams, Dan
                Benner, Lena T.
                Gallagher, Maym
                Benner, Marvin
                Gallagher, [Mr.]
                unidentified child
 BLC Moit LA 30
                The Regents of the University of California
 Collection: Audio Archive of Linguistic Fieldwork
 Duration: 5 hrs. 34 min. 14 sec.
 Documentation: Extensive documentation at Linguistics Department.
 Availability:
                Internet access
 Purchase:
                 may be purchased at cost for personal use
Done
```

http://blc.berkeley.edu

BLC Audio	Archive of Linguistic Fieldwork Conte Iowi Go Bookmarks Tools Help	nts Record - Mozilla Firefox		
📮 • 🗊 • 🚰 💿 🐔 🕋 http://www.nip.be/keley.edu/cgi/bir/bik_cont_disp.plPLA_30+10083				- 0 to K.
Customice	Links 📄 Free Hotmail 🔝 Google Search	Google.com 🔛 Windows Media 🔝 Windows		
		Contents for Maidu (LA 30)	-
Segment	Tude	Language Consent/Note	Personal Names	
:	Miscellaneous words especially animalic numbers, and body parts includes simple phrases, i.e., "I drink water."	Maida, Northeast / Inspartic data / Paynes Creek, CA. Formerly numbered 1.1 English glosses	Benner, Letse T. (Inguistic consultant), Shipley, William L. (collector)	
2	Words and phrases especially knobig terms	Madu, Northeast / Inguistic data / Paynes Creek, CA. Formerly numbered 1.2. English glosses.	Benner, Lens T. (Ingustic consultant); Shipley; William L. (collector)	
1	General talk, includes some words and planates pertinent to discussion.	Maidu, Northeast / Paynes Creek, CA Formerly numbered 1/3. English glosses for some	Benner, Lesa T. (Ingustic consultant), Shipley, William L. (collector)	
4	Earthquake bong with discussion	Maida, Northeast / songa/chants / Paynes Creek, CA. Formerly numbered 1:3. Discussion in English English glosopi for some	Benner, Lens T. (Enguintic consultant); Shipley, William L. (collector)	
	Grassgume song with discussion	Made, Northeast / songs/chants / Paynet Greek, CA. Formerly membered 1-3. Discussion in English and Maidu. No English glosses	Benner, Lens T. (Inguistic consultant); Shipley, William L. (collector)	
	Gambling song with deceasion	Maida, Northeast / songs/chants / Paynes Creek, CA. Formerly numbered 1:3 Directorison in English and Maida. No English glosses	Benner, Lens T. (Iniguintic consultant); Shipley, William L. (collector)	
-	Grasseme sone with discussion	Maida, Northeast / songs/chants / Paynes Greek, CA. Formerly numbered 1.4.	Benner, Lena T. (Inguistic consultant): Sheley, William L.	

http://blc.berkeley.edu

Contact Information

- Mark Kaiser, Associate Director, BLC
 <u>mkaiser@socrates.berkeley.edu</u>
- Marianne Garner, Librarian
 <u>II-lib@socrates.berkeley.edu</u>
- Susan Stone, Programmer
 sstone@socrates.berkeley.edu