

DMP: Linguistics

Roles and responsibilities

The PI Becks is responsible for the oversight and success of the DMP. There are three types of responsibilities to be carried out in the field: data gathering, data storage, and backing up electronic files. Our research team consists of four researchers. All researchers are required to store any field notes taken in a lockbox owned by the PI. Surveys are permitted to be administered by any researcher, but must be delegated by the PI for field storage on a daily basis. TASCAM and camcorder interviews and recordings may also be administered by any researcher. One researcher will be responsible for ensuring that A) all equipment is functional and batteries are charged prior to daily field work, B) there is ample room on memory devices prior to daily field work, and C) electronic data is backed up to an external hard drive on a daily basis.

After the period of data retention, the PI will be responsible for data sharing and managing accounts with the institutions listed under 'Long Term Archiving and Data Sharing and Access'. The PI is also responsible for alerting the participant community that data and results are available to access, as well as ensuring that up-to-date contact information is provided to participants and archiving and sharing institutions.

Ensuring that the data is properly archived in private and institutional domains, responsibilities related to sharing/publishing, and any other decisions regarding the handling of data post-archiving and post-sharing rest with the PI. Should the PI become unavailable for any reason, these responsibilities will be passed on to Becks.

Period of data retention

All data will be archived with restricted access for a period of 6 months or until results have been published.

Data format and dissemination

Data Formats:

Surveys and field notes will be generated to .docx files and will be stored on an external hard drive. They will then be converted to a single PDF file each. PDF files are easily readable, are not easily retroactively edited, and consolidating all files will streamline the archiving process. TASCAM audio is recorded in uncompressed b-WAV format. WAV is the highest quality, least accessible and largest format, thus the audio will be analyzed in WAV, and all publicly accessible recordings will be archived in 356kbs MP3 format. The video/audio camcorder records are encoded as .AVI files which are compatible with the software FinalCut-Pro.

Metadata Tools:

We will use EAD to generate metadata for domain-local archiving. After the results of our research are published, we will use a TEI (Text Encoding Initiative) tool to generate metadata for a public access archiving. This tool is optimal for creating machine-readable text for stories, oral history and other narratives which we are collecting from the interview participants.

Sensitive Information:

A large amount of the data collected, including surveys and interviews, may contain personal information about research participants. In accordance to the obligations laid out by IRB protocol, we will obtain

written and/or verbal consent prior to administering surveys or conducting interviews, which grant us permission to use and publish any information provided for educational and academic purposes. The IRB board has tentatively classified this research as minimal risk research. Its status is under review with anticipated approval within 30 days.

Data storage and preservation of access

Short Term Archiving:

During the retention period all data will be archived privately through the University's online Box account. Only the PI and team researchers will be given access to the files during this time. Backups of data will be maintained by the PI on an external hard drive, which will be stored with completed surveys and field notes at the PI's office at the University. Data that will be submitted to scientific literature will be open access and available at no or minimal cost, including: A comparative study on the phonology of the target language, a synchronic study of the status of the target language, and an article proposing new language revitalization methodology.

Long Term Archiving and Data Sharing and Access:

After the period of retention our data will either be restricted or open access depending upon its content. Restricted access data, including raw audio and video recordings and physical and digitalized field notes and surveys will remain archived on the University's Box account. This content will be openly accessible only to the research team and those who obtain special permission from the PI. A second complete set of all data will be archived and partially shared with the University Archives. Content stored as 'open access' (not restricted access) will be available to anyone at no cost upon request by contacting either the PI or Becks. This publicly available content includes details regarding research methodology and any additional files or media that should be publicly available but have not been archived or shared at another institution. All content shared with the foregoing institutions will also be available to anyone at no cost through University Archives.

Select data will be archived through The Archive of Indigenous Languages of Latin America (AILLA). This content will be freely and publicly accessible and includes: edited video and audio files and transcripts containing content that is relevant to the language, lore, history and culture of the participant community. Select data will be archived through Ethnologue and SIL International. This content will be freely and publicly accessible and includes: data that pertains to the vitality of the target language and areas where the language is spoken.