

## Data Management Plan

Lists the the type of data to be collected.

### Data management during fieldwork

Raw data collected during our fieldwork will consist of (i) audio in .wav format, recorded in mono at a 48 kHz sampling rate and 24 bit resolution; and (ii) written notes taken during recording sessions. We will have portable external hard drives on-site during each fieldwork trip, and will back up all audio data on the day that it is collected. Whenever reliable internet access is available, we will also back up our data to off-site servers using utilities like Dropbox and/or external server space at our home institutions.

Explains how data will be stored in the field.

The PIs will store digital data indefinitely on the portable hard drives mentioned above. The data will also be stored on separate servers at the PIs' home institutions. Finally, digital data and associated annotations will be publicly archived at the *Archive of the Indigenous Languages of Latin America* (AILLA; [ailla.utexas.org](http://ailla.utexas.org)). AILLA has confirmed their willingness to archive the material output of our proposed project (the agreement is included as a supplementary document). We will deposit digital materials with AILLA after the completion of each fieldwork trip, upon our return to the U.S.

Explains where data will be archived.

Clearly indicates that the archive has been contacted & has agreed to ingest the data.

### Public data sharing

Results of this project will be shared in several ways. First, we will publicly share all audio recordings stemming from our field sessions, provided that our [REDACTED] participants give free, prior, and informed consent for the sharing of such materials. All efforts will be made to anonymize data before it is shared: recordings will be associated with basic information about the speaker (initials, gender, age, town of residence), but no other personal information will be tied to the recordings. Speakers may nonetheless divulge personal information during recording; we will never share such recordings if we think that doing so could put the speakers at any kind of risk.

Second, we will share any and all annotations that we make in the process of analyzing audio data. These annotations will be in the Praat TextGrid format or the ELAN .eaf format. For spontaneous speech data, we will also share transcriptions as simple text and/or .xml files. All of these formats can be accessed and manipulated using widely-available no-cost software. Once completed, these audio annotations will be archived at AILLA alongside the associated audio files.

Third, while we do not plan to take extensive free-form fieldnotes, any such notes (including handwritten notes) will be converted to digital .pdf files upon return from [REDACTED]. The resulting PDFs will always be stored with the accompanying audio to preserve their linkage.

We will place absolutely no restrictions on the non-commercial use of our recordings, annotations, and transcriptions, other than those restrictions which are expressly noted in the AILLA use conditions, such as proper citation practices. Commercial use of our research materials will be absolutely prohibited, consistent with pre-existing AILLA policies. We allow for one exception to this restriction: if members of the [REDACTED] community, such as the [REDACTED], request the right to use our recordings and/or annotations in the production of derivative, for-cost materials (e.g. a for-sale print version of our collected narratives), we will grant such rights provided that the resulting products fully respect the prior consent agreements established between us (the researchers) and our consultants.

We intend to place a temporary embargo on public access to our research materials. AILLA has several options for controlling access to archival materials. We plan on depositing our materials

Explains how data will be shared.

Explains how data may be used by others.

Explains temporary access restrictions to be placed on the data.

under “Level 3: Time limit”. Our fieldwork materials will be uploaded to AILLA as soon as possible (see above), but will not be publicly accessible until 24 months after the original date of depositing. The rationale for such an embargo is that it allows us time to analyze these materials and publish research results based on our work before those materials are made available to other researchers.

↗ Lastly, the analytical results of this project will be shared through major journal publications. We have budgeted funds for the publication of papers under an open access license. This will ensure that the publications resulting from this project will be maximally accessible to researchers across the globe, most importantly Latin American scholars who may not have the financial means to access articles published in major for-profit journals.

The storage and data sharing plans discussed above are in full compliance with IRB requirements at our home institutions.

### Timeline for archiving activities

- Year 1: Raw audio data will be deposited with AILLA shortly after completion of the first fieldwork trip (within two weeks). Annotations, transcriptions, and digitized fieldnotes will be shared with AILLA as they are completed. These materials will be under an 24-month access embargo, to expire in Year 3.
- Year 2: Raw audio data will be deposited with AILLA shortly after completion of the second fieldwork trip (within two weeks). Annotations, transcriptions, and digitized fieldnotes will be shared with AILLA as they are completed. These materials will be under an 24-month access embargo, to expire roughly 9 months after completion of the grant.
- Year 3: Raw audio data will be deposited with AILLA shortly after completion of the third fieldwork trip (within two weeks). Annotations, transcriptions, and digitized fieldnotes will be shared with AILLA as they are completed. These materials will be under an 24-month access embargo, to expire roughly 21 months after completion of the grant.

Data collected in Year 3 will be posted on the second archival website as it becomes publicly available on AILLA.

Will publish under open access for greater accessibility by target audience (people in Latin America).

Mentions IRB compliancy of DMP.

↑ Provides a specific time line for ingesting data into the archive and for making data publicly accessible.