Data Management Plan - DDRIG

1. Types of data to be produced

The types of data to be produced in the course of the project are: (1) audio and video recordings of [Language Name] speech; (2) annotation and translation of recordings through linguistic analysis software (PRAAT, ELAN, and FLEX); and (3) pedagogical materials that I will create in conjunction with the [Language Committee] for community use.

2. Standards for data and metadata

The audio and video recordings will be made to take into account the best practice methods of documentation described in Simons & Bird (2003). The highest standards of language documentation will be used for archival use: .wav format (24-bits at 44.1 KHz) for audio and MPEG2 format for video. Also, copies of audio recordings will be made available in .mp3 format and copies of video recordings in .mp4 format; because these formats are smaller in size, they are more easily accessible by the public for downloading and for presentations. Detailed metadata of the audio and video recordings will reflect the required metadata entry provided by AILLA (Archive for Indigenous Languages of Latin America), which is where the recordings will be archived. AILLA's policy with regard to metadata includes noting information about the speaker, date and place of recording, description of data item, length and quality of recording, contributors to the recording, and other additional notes regarding the recording or other resources that may be linked to each unique file. During the course of the project, a spreadsheet in Excel will be kept with the metadata on each recording or "resource". In the initial stages of the project, each resource will be labeled with a file name according to the following template: [Ethnologue Language Code, Date *(YYYMMDD),* Speaker *(first & last name)*, Genre *(e.g. song, narrative)*, Title], for example LGCODE20120130_ABC_Song_FishingSong. Eventually, once the resources are archived at AILLA, these file names will be converted to a unique AILLA ID.

       Through linguistic analysis programs, the recordings will be manipulated in order to annotate inherent linguistic information. Through PRAAT, the recordings will be analyzed with regard to prosodic patterns such as intensity, pitch, and length of sounds, all which is relevant to the project. PRAAT data is saved under a .Textgrid file format which is open access to the public from the software website: (http://www.fon.hum.uva.nl/praat/). The program ELAN is an aid to transcribe and translate recordings from [Language Name] into Spanish and English, and produce .eaf files, which are also open access from the website (https://tla.mpi.nl/tools/tla-tools/elan/). Finally, the SIL's (Summer Institute of Linguistics) latest program FLEx will be used to index an analysis of the texts and as a database for building a lexicon. The FLEx database is also public and open source (http://fieldworks.sil.org/flex/). All three of these program files are formats which may be included in the AILLA archive.

       Finally, pedagogical materials of [Language Name] language and especially [Language Name] verbal art will be produced by and for the [Language Committee] in order to disseminate among the community and children undergoing pedagogical study. The standards used for the creation of these materials are largely left up to the [Language Committee] and the people from whom they may solicit help (e.g. local school teachers, the co-PI, the [tribal/community council]).

The co-PI will provide the [Language Name] Committee with training in language documentation as well as training of computer use and computer software such as Excel and Word (which will be archived in PDFA format) and well as linguistic analysis software PRAAT and ELAN. All of these programs will help to aid in the production of pedagogical materials. The [Language Name] Committee will also be provided with two years of a printing stipend in order to create copies of papers they produce to circulate among community members. The co-PI will suggest that the [Language Name] Committee scan and backup all [Language Name] language materials on external hard drives which they will be provided with.

3. **Policy for access and sharing**
Research methods used in this study were approved by the IRB at the [University Name] with expedited status from [Date] to [Date], and the study may be reapproved every year after that. The co-PI is responsible for providing consent forms to participants in this project. Because this project aims to aid in transmission of [Language Name] verbal art from the research team to younger [Language Name] speakers and semi-speakers, the goal is to make all of the data open access so that the recordings may be disseminated freely within [Language Communities]. It is the intent to encourage [Language/Culture Name] people to make the data publicly available to non-[Language/Culture Name] communities on a wider scale as a sign of language valorization and pride. However, it is ultimately the decision of the [Language Name / Community] assembly to vote on the public access of recordings, and approval from the [Community Leader Name] and his council. The four levels of data accessibility offered by AILLA can flexibly accommodate the request of [Language / Culture Name] people. In compliance with IRB regulations, in any case, privacy and confidentiality of [Language Name] speakers will be upheld unless it is a consultant's wish to receive recognition for his/her contribution to the project.

4. **Policies and provisions for re-use, re-distribution, and copying of data**
At the level of the [Language Name] Committee and [Language Name] communities, it is encouraged that data be circulated and copied (only at the discretion of the speaker recorded), in a political move to help revitalize [Language Name] verbal art. The data will undergo constant revision by the [Language Name] Committee and co-PI along the course of the research project. However, as mentioned earlier, the ultimate decision as to the restrictions placed on public access to the data is in the hands of the [Language / culture Name] people.

5. **Plans for archiving data and preserving access**
Along the course of the project, data will be regularly backed up on four external hard drives as well as SD cards, which are needed because the size of audio and especially video data is very large. With long-term archiving in mind, this project collection will be a work in progress as it is archived at AILLA and completed upon the project end date. The co-PI will regularly actualize the access level of the project collection based on [Language Name] people's decisions about the extent to which they may be shared.