

## **Text Analysis of Taldash (GAL) in Support of Nuu-wee-ya' Language Revitalization: Indigenous-based linguistic analysis and methodological reflections**

---

### **Roles and responsibilities**

The Primary Investigator is responsible for the digitization and management of the data.

### **Expected data**

Four types of data will come out of this proposed research (shown in table 1 below). The first type of data is the digitized form of the archival data. The data has already been partially digitized through previous NSF funded research. Each page of the archival pdfs that remain to be digitized are typed into Excel. This includes everything on the page, the form in Nuu-wee-ya', its gloss, and any comments by the researcher, as well as the location of each entry in the archival data. This is done by assigning a page number, line number and phrase number to each entry. These codes allow the .csv to be uploaded to the online Nuu-da' Mv-ne' archival database hosted on the Indigenous Languages Digital Archive (ILDA) platform and also allow the researcher keep track of the original data through analysis done outside or inside of ILDA.<sup>1</sup>

The second type of data is the glossed form of the archival data. This data is created by importing the .csv files into Language Explorer (FLEx) and then parsing and glossing the data in the text portion of FLEx.

The third type of data is the lists of the types of morphemes found in the data set and how they compare with data from other Nuu-wee- ya' dialects which were analyzed during my dissertation. This data will be created by extracting a list of all the types of morphemes in the data and putting them in csv files. One file will house the entire list, like a lexicon. Others will contain the semantic groups of morphemes and their cognates in other dialects/languages.

This data complements both research done for an NSF DEL grant and the subsequent analysis of that data in dissertation research. This data is done with the same formats as previous research and will be housed in the same locations.

The materials used in this research project contains sensitive cultural material. While all data will be analyzed, portions of the digitization will be redacted. Access to these materials will be considered by Nuu-da' Mv-ne', the non-profit whose mission is to protect and tend the materials of Nuu-wee-ya' and provide appropriate access to community members and researchers. This project is following the same protocol developed for the ILDA database through NSF DEL grant project # BCS-1562859 to protect yet still utilize the materials for linguistic insight. Analysis and descriptions that contain sensitive material will be done and housed with Nuu-da' Mv-ne' to supervise access.

All four types of data should be preserved and made sharable but only the digitization efforts need to be actively shared. The other types can instead be stored and made available to other researchers. This .csv and .xml files could be used in future linguistic research in sounds, structure and comparative elements of the language as well as to support cultural or pedagogical analysis or materials development.

Materials come from the Jacobs archive at Washington State University. These are in the public domain. All analysis will cite and retain meta-data on the location of the material in the archive. Publications will present the PIs work on these original materials, that includes glossing, description, and comparison.

### **Period of data retention**

All data that is not culturally sensitive will be deposited in two repositories and shared with the tribal communities connected to this research (Confederated tribes of Siletz, Tolowa Dee-ni Nation, Coquille Indian Tribe) within 3 months of the end of the project. A portion of the data types will also be made available on the already developed online digital database ILDA. The data will be deposited at the University of Oregon's Scholar bank along with the other data created in the prior NSF DEL grant and dissertation work. The data will also be included in Humboldt State University Library Special Collection's Digital Indigenous Archive as appropriate. The csv files containing typed forms from the handwritten notes as well as the

---

<sup>1</sup> <https://mc.miamioh.edu/ilda-nuuweeya/>

transliteration in the modern alphabet will also be uploaded to the Nuu-da' Mv-ne' ILDA instance. The ILDA software platform and instance data are maintained by Miami University, Ohio.

The definition and handling of culturally sensitive materials will follow guidelines defined by the non-profit Nuu-da' Mv-ne', which was established by members of the Nuu-wee-ya' speech community to manage the Nuu-da' Mv-ne' ILDA archival database. These guidelines are designed to ensure the needs of the cultural communities are prioritized, while still providing access to sharable data. Information to be redacted will likely include personal and biographical information, locations of historical sites, sensitive historical information, sensitive ethno-cultural information. The purpose of this non-profit is to tend to the digital archive and access to the digital archive in a community-based way.

### Data format and dissemination

The data will either be stored as comma-separated values (.csv) or extendable markup language (.xml) depending on their nature. Meta-data will be in plain-text (.txt). The digitization of the Galice words and the original translation as well as the transliteration of the words into the modern alphabet will be in comma-separated values (.csv) format because this makes it interoperable with ILDA and other databases. The glossed data will be stored in extendable markup language (.xml) to make into interchangeable with SIL language database: Language explorer as well as other databases interoperable with this format.

Metadata will be provided on the researcher and language authority for each document. This will be in a plain-text (.txt) file and will include pertinent background information. Metadata will also be kept on each line of text as it will be assigned with a numeric code that lists its page number, line number and phrase number in the document that it is from.

### Data storage and preservation of access

During the project, data will be stored on the PI's computer and automatically synced to OneDrive. Once a week, the PI will back up the data to a hard drive.

Portions of this data will be stored in three repositories, all the non-sensitive data will be stored in two of those depositories. All data will be shared with cultural communities. All files that are uploaded to the Nuu-da' Mv-ne' ILDA instance are preserved at Miami University, Ohio. All non-sensitive files will be stored in University of Oregon's scholar bank. The PI is working collaboratively with depository managers to determine where materials will be stored and where metadata links will be. Location and type of stored data is shown in table 1.

Table 1. Types and locations of stored data from this project.

<b>Data</b>	<b>Description</b>	<b>Shared with tribes</b>	<b>Open Repository</b>	<b>Restricted Repository</b>
csv	Digitized and searchable version of handwritten archival language data	X	ILDA, UO Scholars bank	Nuu-da' Mv-ne' (around 10%)
xml	Transcribed archival data with glossed morphemes	X	UO Scholars bank	Nuu-da' Mv-ne' (around 10%)
csv	Descriptive lists of the morpheme types and tokens as found in the archival data and as compared to the morpheme types and tokens found in PI's previous (dissertation) research	X	UO Scholars bank	—
txt	Metadata files	X	UO Scholars bank	—