

Data Management Plan for National Breath of Life 2.0: Creating a 'Second Breath' for Indigenous Language Revitalization (BoL 2.0)

This project results in two major data deliverables: (1) the *Indigenous Languages Digital Archive* (ILDA) software and (2) the language research artifacts produced by the user communities. The following is the data management plan for each deliverable.

ILDA Software

Intellectual Property Agreements

The relationship between Miami University, the Miami Center, and the Miami Tribe of Oklahoma is governed by a perpetual Memorandum of Agreement, which outlines proprietary interests over products produced through the Myaamia Center, such as Miami-Illinois Digital Archive (MIDA) and ILDA. Article Four - Intellectual Property, states:

WHEREAS, It is understood by the terms of this agreement that the Miami Tribe, by right of self-determination, has control over its cultural and intellectual property on behalf of the citizens of its nation. Neither the Myaamia Center nor Miami University may copyright materials produced through and/or by the Myaamia Center. No reprinting or distribution of materials produced by the Myaamia Center may occur without the express written consent of the Miami Tribe of Oklahoma. (Memorandum of Agreement signed March 31, 2016)

The ability of ILDA to meet the needs of other communities who are revitalizing their languages from documentation was tested during the 2017 National Breath of Life Archival Institute for Indigenous Languages (BoL). Given the success of the pilot as outlined in Appendix 2, the proposed BoL 2.0 will offer ILDA through future workshops. Permission is granted by the Miami Tribe of Oklahoma to continue developing ILDA for the purpose of sharing it with BoL alumni (see Miami Tribe letter of commitment). Based on the outcomes of BoL 2.0, additional long-term permission will be required and will be handled through appropriate legal advisement.

ILDA Software Access

ILDA will be made available through a publicly accessible website which will provide read-only access to language research artifacts as the default. Accounts with secure credentials will be provided to BoL workshop trained researchers that enables them to maintain their language artifacts. Each account allows users to maintain the data for their respective research projects only.

ILDA Software Architecture

ILDA (and MIDA) are implemented using the industry-standard web service solution commonly referred to by the acronym LAMP. The 'L' in LAMP refers to the operating system, Linux. The web server is Apache, the 'A' in LAMP. The underlying database is a relational database management system using MySQL (the 'M' in LAMP). The web pages, search logic, and database management code are programmed using a combination of the HTML, PHP (the 'P' in LAMP) and JavaScript programming languages. The use of these industry-standard languages,

operating systems, and applications assures ongoing software support, maintainability, and technical documentation.

ILDA Hosting Environment

ILDA is hosted by Miami University's Information Technology Services, the provider of the University's Internet and Web infrastructure. Miami's IT Services infrastructure includes a fully redundant data center located in Oxford, Ohio. Internet and Internet2 access is provided to the data center by OARnet on redundant connections. Computer servers and associated storage are provided by a VMware vSAN cluster. The ILDA software and language research data is backed up to another physical location on campus using IBM Spectrum Protect. Miami's IT Services provides continuous operating system and select application security updates to the entire infrastructure. Included are Linux operating systems, Apache web servers, and supporting middleware to assure future compatibility and security.

ILDA Software Development Procedures

The ILDA software development environment utilizes industry-standard best practices and tools. GITlab is used for the source code repository and version control. The development process follows the DevOps industry standards. DevOps requires a development site for the programmers, test site for pre-deployment, and a production site for the release version. The test site is used for user testing before updates are released to production. This process is consistent with Miami's IT Services development and operational policies. It assures that ILDA is developed following robust programming, testing, and release procedures.

ILDA Project Management

To mitigate risk and be responsive to user needs, the ILDA developers use an Agile project management approach. Agile practices include close communication with users, iterative development, frequent releases, and continuous improvement of work practices. Dr. Troy is an accredited Agile instructor by the International Consortium for Agile (ICAgile).

Language Research Artifacts

Three main types of data constitute the language artifacts for each community: (1) digital image surrogates of archival language documents, (2) alphanumeric data such as transcriptions, translations, and linguistic analyses, and (3) digital audio recordings. The management of each is described below.

Intellectual Property Agreements

The management of intellectual property rights for language research artifacts, including obtaining permissions from archives, rests with and is the responsibility of each of the participating user communities and will be defined in an end user license to be developed to support this project.

Digital Image Surrogates

Digital images are high resolution JPEG images of archival source documents. Each user community will be trained in the process of preparing and uploading the images corresponding

to each document to be analyzed. Images are indexed by source document, page, line, and phrase. Once uploaded the images are stored in the ILDA Linux server directory structure, which is organized by document within each participating community section of the public website.

Alphanumeric Data

This data consists of transcriptions of source documents, translations, and linguistic analyses. This data is stored in the ILDA MySQL database. ILDA provides for bulk upload and download capability, allowing users to export this data from the database at anytime. This capability can be used to migrate information out of ILDA, if needed. It can also be used to perform bulk edits, with many changes being subsequently uploaded in bulk.

Audio Recordings

Audio recordings are stored as MP3 (MPEG) files. Audio recordings are stored in the ILDA Linux server directory structure similar to the way that digital images are stored.

Language Artifact Storage and Backup Procedures

As described above, the The Miami University hosting environment provides secure data storage in its fully redundant data center. The Linux file system and the MySQL database are backed up daily.

Conclusion

The ILDA software will be available to the research community through a public website. The ILDA software and database is implemented, hosted, and backed up using industry-standard practices with support from Miami University's Information Technology Services. Data produced by the community researchers, stored in the ILDA database and Linux file system, is secured through password-protected accounts and backed up daily. Bulk download capability allows community researchers to migrate data out of ILDA if necessary.