

Finding an Archive for your (Endangered) Language Research Data¹

Susan Smythe Kung

Archive of the Indigenous Languages of Latin America

University of Texas at Austin

Written June 2015

Revised August 2016

The submission deadline for the NSF/NEH Documenting Endangered Languages grant is just around the corner (September 26, 2016). If this is your first time applying for a DEL grant, please be aware that all varieties of the DEL grant (Senior Research Grants, Doctoral Dissertation Research Improvement Grants, and Fellowships) require you to submit a support letter from an archive that has agreed to accept your data for ingestion. So how do you find the archive that will be the future home of your language research data? And once you've identified an archive that is a good fit, what do you do next? The following information and tips will be helpful to researchers who plan to apply for any of the DEL grants, as well as to anyone who plans to submit research data to a(n) (endangered) language archive, regardless of his/her funding source.

DELAMAN Archives

The Digital Endangered Languages and Musics Archives Network (DELAMAN, www.delaman.org) is an international network of archives that preserve materials in or about endangered and/or indigenous languages from all around the world. Links to the individual archives can be found under the Members tab (<http://www.delaman.org/members/>) on the website. Some of the archives specialize in a specific region of the world, and each archive has its own collection policies and fees. Some archives have a self-deposit feature that allows depositors to organize and upload their own born-digital data via a web interface, while other archives do all data ingestion in house. Some archives can digitize analog data while others cannot. The best way to determine if one of the DELAMAN archives is a good fit for your data is to start at the DELAMAN website, then follow the links to the individual archives to read their collection policies and intake procedures.

¹ This blog post was originally posted in June 2015 on the CELP: From the Archivists blog, which was a subsection of the larger blog for the LSA committee on Endangered Languages and their Preservation.

The following DELAMAN member archives provided some specific information for this blog post:

- *Alaska Native Language Archive* (ANLA) at the University of Alaska Fairbanks houses materials relating to Alaska's 20 Native languages, including varieties spoken outside Alaska, and in some cases, languages related to those spoken in Alaska. Information for depositors can be found here: <http://www.uaf.edu/anla/about/deposit/>.
- *American Philosophical Society* (APS) accepts most forms of digital and analog linguistic research data for languages of North and Central America. The APS does not charge fees for depositing data, though depositors needing their data to be fully catalogued by a given date in order to fulfill funding (or other) requirements should consult with the APS ahead of time to ensure this is feasible in relation to the nature and quantity of data. Inquiries on depositing, archiving procedures, and access policies can be directed to the archivist for the APS Center for Native American and Indigenous Research, Brian Carpenter (bcarpenter@amphilsoc.org).
- *Archive of the Indigenous Languages of Latin America* (AILLA) at the University of Texas at Austin accepts any materials relevant to any indigenous language of Latin America and the Caribbean. AILLA is undergoing a major repository and procedural upgrade at the time of writing, so it is not currently accepting deposits until March 2017 at the latest. Nevertheless, *AILLA will continue to provide letters of support for researchers during this temporary suspension of depositor services*. Information about how to make a deposit and associated fees can be found at http://ailla.utexas.org/site/dep_info.html. Please be aware that this information will be updated (and dramatically changed) during the 2016-17 academic year as part of the upgrade. For a support letter, please contact the archive at ailla@ailla.utexas.org.
- *California Language Archive* (CLA) at the University of California Berkeley does not require a Data Management Plan or charge fees, though when depositors have external grants, CLA staff are happy to discuss the latter. Instructions for depositors are found at <http://linguistics.berkeley.edu/~survey/archive/for-depositors.php>.
- *Endangered Languages Archive* (ELAR) at the School of Oriental and African Studies, University of London primarily accepts data that were collected by researchers with support from the Endangered Language Documentation Programme (ELDP); they charge

a fee for all other depositors. Information about depositing materials at ELAR can be found online at <http://www.elar-archive.org/depositing/index.php>; and a PDF can be downloaded from

http://www.eldp.net/pdf/ELAR_Depositing_Instructions_20150610%20%281%29.pdf.

ELAR recommends that new depositors use the ELDP profile for the CMDI Maker (<http://cmdi-maker.uni-koeln.de/>), an offline web app developed by the University of Cologne for adding metadata to linguistic data, to prepare files for deposit. They have two video tutorials on how to use their CMDI maker that can be viewed at

<https://www.youtube.com/watch?v=1YE2RLyTj-Q> and at

<https://www.youtube.com/watch?v=ZKyhlRGfY2M>. More experienced depositors might want to use the Arbil tool (<https://tla.mpi.nl/tools/tla-tools/arbil/>), a more advanced app developed by the Max Planck Institute for Psycholinguistics for organizing files and metadata for deposit in a digital archive.

- [*Kaipuleohone Language Archive*](#) at the University of Hawai'i Manoa accepts materials from University of Hawai'i affiliates, and from anyone else with materials on languages from the Pacific or Asia. They accept born-digital items and can digitize analog materials like reel-to-reel and cassette recordings, images, and fieldnotes. Please read their [Deposit Agreement Form](#) and their [Embargo Policy](#). To inquire about depositing with Kaipuleohone, including getting a Letter of Support for your NSF DEL proposal, please contact them at kaipu@hawaii.edu.
- *Pacific and Regional Archive for Digital Sources in Endangered Cultures*: PARADISEC's digitization services and fees are explained here: http://www.paradisec.org.au/comm_services.html. General instructions for depositors can be found at <http://www.paradisec.org.au/deposit.html>.
- *The Language Archive* (TLA) at the Max Planck Institute for Psycholinguistics currently mainly serves as an archive for its own researchers and associated projects such as DOBES. The archive requires metadata to be provided in [CMDI format](#) (certain profiles) and expects depositors to use their [LAMUS](#) web-based tool for depositing materials. Acceptance of external deposits will be decided upon on a case by case basis, for inquiries contact them at tla@mpi.nl.

Alternatives to DELAMAN Archives

Though the DELAMAN member archives are unique and varied, some researchers have not been able to find one that is a good fit for their research projects. For example, some archives prioritize data that results from research that their sister organizations fund. Other archives accept data only from particular regions of the world. Occasionally archives must temporarily suspend their depositor services in order to catch up on their ingestion backlog or upgrade their repositories and web interfaces. So what do you do when you cannot find a DELAMAN archive that can house your research data?

The first place to look is in your research community, which might already have its own community archive or a relationship with an existing archive or library. These organizations might be willing to add data that is collected as part of your project to their existing collections. If there is a community or affiliated archive, you should consider putting your data here even if you plan to use one of the DELAMAN archives as well.

The next place to look is at your home institution. Almost all US institutions have an Institutional (Data) Repository (IR) for their faculty, students, and researchers. Many universities now require their faculty (and sometimes their graduate students) to put their research results in their local IRs. If your institution does not have its own IR, it might be a member of a larger IR to which you can submit your data. Make inquiries at your institution's main library.

Alternatively, you could consider placing your data in a large, public data repository such as the Dataverse Project at <http://dataverse.org/>. The Tromsø Repository of Language and Linguistics (<http://site.uit.no/trolling/about/>) is a Dataverse instance dedicated specifically to *linguistic* datasets and statistical code (but not audio or video of naturally occurring speech). Please see the website for more information.

A useful tool to search for data repositories throughout the world can be found at the Registry of Research Data Repositories (www.re3data.org).

Contacting an Archive

Once you've identified an archive that seems to be an appropriate repository for your research data, you need to contact the archive (i) to make sure that they will be able to accept your data for deposit; (ii) to find out what requirements they have and what, if any, fees they

charge; and (iii) to request a support letter to be included in your grant application package. The following points are things to keep in mind when contacting the archive:

First contact:

- Contact the archive to which you plan to submit your data well in advance of the grant submission deadline (ideally 2-3 months in advance). If you contact them just a few days before your proposal is due to your Office of Sponsored Projects, they might not have time to review your DMP or write a support letter for you.
- Remember that the people who work at the archive are very busy with the ongoing work of that archive and that institution. Since they have work-related deadlines of their own, they cannot drop everything to rush your support letter because you waited too late to request it. Also remember that you are not the only person that has contacted that archive. Some archives write 10-20 support letters for DEL grants every application round.
- The archive's representative (which might be the manager, the director, an archivist, or some other staff member) will need to know some basic information about your project before s/he can write the support letter for you (see below).

Basic information about your proposed project to give to the archive:

- The title of your proposed project (i.e., the exact title of your grant application).
- The research language(s) and ISO 639 code(s).
- The name of the PI and names of any Co-PIs.
- A summary of the type of data you plan to submit to the archive (e.g., audio, video, translations, transcriptions, annotations, text grids, etc.) and their format (.wav, .mpg, .mov, .eaf, .trs, etc.). Note in particular if you plan to submit any analog data to the archive. If so, the archive might charge a fee to digitize your analog data.
- A timeline indicating when you plan to send your data to the archive (e.g., once a year, after each field trip, at the conclusion of the project). Note that different archives have different preferences for how and when you submit your data. For example, AILLA encourages depositors to submit small batches of data at frequent intervals, but ELAR asks that depositors submit all of their data at once at the conclusion of the project.
- A copy of your Data Management Plan (http://www.nsf.gov/sbe/SBE_DataMgmtPlanPolicy.pdf). The archive representative might want to read your DMP and provide you with any feedback that is relevant to

his/her archive (e.g., you might have included a file format type that the archive is not equipped to handle).

Post-award notification:

- Follow up with the Archive and let them know if you got the award *or not*. The archive needs to know if and when to expect your data.
- If you got the award, send the archive a schedule of when they should expect your materials, a description of what you plan to send, and an estimate of how many files and GB of data you anticipate sending. This way the archive can put your data into its queue, arrange server/shelf space, etc. Note that this tentative schedule might be the same one that was in your DMP, or you might have had to revise it for your grant.

The most important things to keep in mind as you choose an archive to house your research data are these: start your search early and communicate with your chosen archive regularly.