

## Data Management Plan

"Documenting the Ethnobiology of Mexico and Central America (DEMCA): A Digital Portal for Collaborative Research in the Humanities, Social Sciences, and Natural Science" will develop a web portal of ethnobotanical and botanical materials pertinent to the aforementioned disciplines. The substantive content of the portal—botanical collections and in-situ photographs; ethnobotanical data on plant nomenclature, classification, and use as well as digital recordings, transcriptions, and translations—will be developed by Amith and Beck from projects they currently direct in Nahuat, Totonac, and Mixtec communities, projects supported by the NSF (Documenting Endangered Languages), the Endangered Language Documentation Programme (SOAS, London) and the NEH (Preservation and Access). Seven additional linguists (see support letters, pp. 18–26, section 9) will provide smaller amounts of similar data to fully test DEMCA functionalities and the viability of the standards for ethnobiological metadata developed for the Symbiota/DEMCA portal.

This data management plan includes management of all materials that will be used in the DEMCA portal even though these materials were produced with funds from other grants. This Level II project will, however, produce new material: an integrated database of botanical and ethnobotanical materials. Brandt will ensure that a system is implemented to produce backup files of the entire DEMCA database and contents on a nightly basis. Each nightly backup file will be preserved for the duration of a one week on a machine separate from the server to ensure several options for recovery if data restoration is ever required. At project end the last backup will be deposited at the Archive of Indigenous Languages of Latin America, University of Texas (AILLA) and at the American Philosophical Society (APS) for long-term archiving (see Letters, pp. 36, 37)

### 1. Type of materials    + indicates material produced by other grants but which will be used as content for DEMCA    \* indicates material produced by this NEH project

- +a. A database of Indigenous nomenclature, classification, and use of plants based on information derived from various sources: (1) Amith's and Beck's work on Sierra Norte/Nororiental de Puebla Nahuat and Totonac; (2) Amith's work on Pacific Coast of Guerrero Mixtec; and (3) contributions by the seven collaborating linguists and anthropologists (see Letters, pp. 18–26).
- +b. An extensive set of archival quality (48KHz, 16-bit) digital Indigenous language recordings, the majority from sources (1) and (2) in (pt. a) above. Towards project completion, seven linguists; see (3) above, will provide DEMCA test data from other languages and hopefully include recordings. Amith and Beck will develop some video (mp4) to test DEMCA's video presentation.
- \*c. Metadata of all ethnobotanical and botanical information built into DEMCA as documented in the Appendix. This metadatabase will provide the informational foundation for the DEMCA web portal.
- +d. Voucher specimens of fertile (flowering or fruiting) plants from areas mentioned in 1.a.. These materials will be collected according to herbarium specimen standards with full collection data (e.g., GPS coordinates, habitat, plant description) and ethnographic information (nomenclature, classification, and use in Indigenous languages and cultures). All collected plants will be extensively photographed in situ (flower, leaves, stem, fruit, entire plant) and the mounted voucher specimens will be digitally photographed. Approximately 10,000 collections and 2,500 species will be made.

### 2. Standards for data and metadata

There are three basic types of data, each of which will be managed with distinct metadata and standards.

- a. Ethnographic material includes (1) database of plant nomenclature, classification, and use, linked through metadata to a physical botanical specimen and its occurrence data; (2) digital recordings, transcriptions, and translations of Indigenous experts talking about a particular plant or group of plants. The recordings will be archived in uncompressed .wav format (at 48KHz, 16-bit). The metadata for the recordings will be archived in IMDI format (<https://tla.mpi.nl/imdi-metadata/>), a shared standard used by many archives for endangered language material. The transcriptions and translations will be in ELAN format (<https://tla.mpi.nl/tools/tla-tools/elan/>), again a standard XML scheme used by most linguists and endangered language archives. Data on plant nomenclature, classification, and use will be archived in an XML schema developed in this present project.
- b. Plant specimens and photographs: The collections will have unique numerical identifiers; the database will record standard collection data: coordinates, biological form, habitat. Specimen labels

for herbarium sheets will include this information as well as ethnographic data: indigenous name, classification, and use. The Darwin Core standard will be used for all metadata pertinent to botanical collections. The physical specimens will be accessioned at major herbaria (e.g., Mexican and US national herbaria, Missouri Botanical Garden). The photographs will be incorporated into the Tropicos system at the Missouri Botanical Garden (see Letters, pp. 34, 35, 17).

- c. **DEMCA database:** DEMCA will develop an integrated botanical and ethnobotanical database according to the work plan and appendix. This database will contain the metadata and content from the botanical and ethnobotanical material. It will be backed up daily and periodically deposited at project conclusion at AILLA and at the APS.

### 3. Policies for access, privacy, confidentiality and IPR

Ethnographic material (e.g., recordings) will be gathered with strict adherence to recognized IPR protocols of informed consent and privacy when the collaborator so desires. Amith has secured IRB approval for this project with Gettysburg College and follows a protocol that he has been using for 13 years. Communities in whose lands collections are undertaken will be approached and collection will only proceed if a signed agreement is reached (see example, Letters, p. 27). When pertinent, additional agreements will be reached with community authorities and schools (Letters, pp. 28, 29). All those who record will be given a mutually signed "memorandum of understanding" about privacy issues, potential dissemination, and rights as authors (Letters, p. 30). Both orally and in this agreement it is explained that the material will be used only for education and research and any "commercial" publication would require new written consent with a clear statement that any economic benefit would be for the author. All botanical collections will be made with the proper permits and export from Mexico will be through the national herbarium (MEXU).

### 4. Policies and provisions for re-use, re-distribution and the production of derivatives

All ethnographic material (digital recordings, transcriptions, translations) may be used for academic, non-commercial purposes as per provisions that will be posted on the open access website (e.g., AILLA, APS, DEMCA). Distribution will follow the Creative Commons License most appropriate to this material: Attribution-NonCommercial-ShareAlike (CC BY-NC-SA) although it is recognized there are some grey areas.

The voucher specimens and collection data will also be freely available through the herbaria in which they are deposited and collection data may be cited. Photographs (both of plants in situ and voucher specimens) may be used under the same CC BY-NC-SA license.

The Symbiota software developed for this project will be open source and available on GitHub.

### 5. Plans for archiving data, samples and other research products

**Ethnographic material:** Arrangements have been made to archive all Indigenous language digital recordings, time-coded transcriptions, and translations at AILLA and the APS digital archive. Proper metadata records in the IMDI format (<https://tla.mpi.nl/imdi-metadata/>) will facilitate discoverability and provide researchers with the necessary information about the recordings (e.g., name, origin, age at recording, of narrators) and about the plants and voucher specimens to which they refer. AILLA and APS will also archive the database of all collections (which will include data on the collection and ethnographic information about plant nomenclature, classification, and use). **Support letters: AILLA and APS (Letters, pp. 36, 37)**

**Plant specimens:** Of three plant vouchers, one voucher of all collections will be deposited at MEXU (Mexican National herbarium), one at US (US National Herbarium), and one at the Missouri Botanical Garden. The Darwin Core compliant metadata will be provided to each herbarium for porting to their database system. In-situ photos will be accessioned at Missouri and the metadata for the photos, incorporated into the Tropicos database. **Support letters: MEXU, US, MO (Letters, pp. 34, 35, 27)**

**DEMCA database and content:** At project end, all metadata and content will be exported from DEMCA and deposited at AILLA and APS. **Support letters: AILLA and APS (Letters, pp. 36, 37)**

**Symbiota software:** This software is open source and will be posted on GitHub. The posted version will include the data structure for ethnobiological information developed by this project.