**Data Management Plan.** This project will generate three kinds of data: 1) assessment data on the students and the training, 2) materials generated for teaching in workshops, and 3) data generated in the course of research by grant participants.

**Assessment data**: The types of data generated for the assessment portion of the project will include: instruments (surveys, interviews, observation field notes), data generated from the instruments, and reports (annual, summative).

Formats. Participant and other stakeholder data will be collected for the duration of the project. Qualitative data will be generated through open-ended survey items and noted during interviews and observations. It will be collected and transcribed by qualified staff and stored electronically in Word or PDF format. Quantitative data will be generated via surveys, meeting minutes, and project documents. It will be stored as data files in Excel format. Both formats will ensure ease of use among project personnel. All data collection instruments will be saved in commonly used formats.

Archiving. Data will be saved on secure file storage within University of Tennessee.

Sharing. Participant and stakeholder data will be shared with personnel without personally identifying information through a shared file server approved for FERPA-eligible data. Assessment data will be stored and backed up at both EMEC and the University of Tennessee. EMEC will collect, organize, de-identify data into raw files that will be provided to the Project Coordinator for archival storage. Annual and summative reports and manuscripts generated from the assessment data will protect the confidentiality of participants who provided the data. Sharing, including manuscripts for publication, will adhere to the IRB rules instituted by the University of Tennessee, Knoxville.

Confidentiality and Security. All participant data will remain confidential. The data will be collected according to rules of the Institutional Review Board (IRB) at the University of Tennessee, Knoxville. The Project Coordinator if not already, will be trained regarding the rules of research data. Any data collected via pen and paper and consent forms will be stored in a locked cabinet of the Project Coordinator. All data (transcribed or electronic) will be stored on a password protected computer in secure location under the control of the lead PI.

**Teaching materials** will be distributed under an open access license allowing for educational reuse. Materials will be available via dedicated training websites but also deposited on GitHub, with citable DOIs provided by Zenodo. As best practices change, we will adapt to ensure the materials are reusable and available without any registration or login barriers. Teaching materials will include slides, exercises, and notes. Staton and O’Meara have a significant record of doing this already.

**Research data** will be placed in appropriate repositories for each kind of data at the time of publication. Possible research products:

* Genetic data will be deposited in NCBI’s GenBank, which maintains long term storage and free reuse.
* Phylogenetic trees generated will be placed in both the TreeBase and Open Tree of Life repositories under a Creative Commons Zero license allowing free reuse.
* GIS data will be pushed to the Global Biodiversity Information Facility (GBIF), also under a CC0 license. The only exception would be for species of conservation concern where disclosure of locations could lead to their targeting by collectors or other bad actors.
* Research software created will be published via GitHub and language-specific repositories (such as CRAN for the R language) under an open source license.
* Specimen vouchers will be deposited in appropriate museums.

In addition to the specialized types of data above, data for all all publications resulting from this grant will be available via Dryad, a journal and society-sponsored repository that maintains data long term and allows its citation and reuse. All data deposited on Dryad are CC0 licensed.