

The Principal Investigators of this proposal have read the NSF Award and Administration Guide on Dissemination and Sharing of Research Results (AAG Chapter VI.D.4), and will manage and share data in accordance with the following Data Management Plan.

1) Kinds of data to be collected: This project will involve cataloging herpetological and ornithological specimens that are backlogged in the MVZ, organizing and cataloging histologic and blood smear slides, cataloging digital images, and rehousing bird skins in new cabinets with updated taxonomic organization. The data will be obtained from original collector field notes and labels, and will include basic information such as: scientific name, collector/preparator name, collector/preparator number, collection method, locality, date/time, latitude, longitude, datum, georeferencing protocol, maximum coordinate uncertainty, parts preserved, tissue barcode, sex, age, stomach contents, reproductive condition, and measurements or other attributes (e.g., weight, snout-vent length, molt, fat, etc.). Data associated with photographs and/or microscope slides will be cross-referenced to specimens.

2) Data standards: MVZ data are managed through Arctos, a collaborative and comprehensive collection management system that currently serves 20 institutions and 65 collections (<http://arctos.database.museum/>). Arctos integrates access to diverse types of collections (e.g., vertebrates, plants, fossils, insects) and data, including specimen records, observations, tissues, parasites, stomach contents, field notes and other documents, and media such as images, audio recordings, and video. It also provides solutions to managing and integrating collections data with object tracking (e.g., barcodes), transactions, geospatial information, people and organizations, and usage (publications, projects, citations, GenBank).

Arctos incorporates a set of controlled vocabularies and authorities for managing data content such as taxonomy, geography, names, parts, attributes (e.g., sex, age), identifier types, media metadata, and data licensing (documentation at <http://arctosdb.org>). Users entering data into Arctos must conform to these standards, and new values are added to the Arctos vocabularies and authorities based on user needs and Advisory Committee approval. For these reasons, data quality in Arctos is relatively high compared to other systems where data entry is more free-form. Arctos also provides views that map Arctos fields to Darwin Core standards. These views have been developed in collaboration with Senior Personnel on the VertNet project, and are being used to share data with VertNet, GBIF, and other data portals via the Integrated Publishing Toolkit (IPT). In addition, the Principal Investigator (Cicero), MVZ archival staff, and Arctos programmer are working together to improve the Arctos schema for media so that it conforms better to Dublin Core and Darwin Core (including Simple Images Extension) standards.

3) Policies for accessing and sharing data: Data are entered into Arctos directly through a web browser, where they are stored in a staging table until curatorial staff with higher-level permissions check and load the data into the production database. Once the data are loaded, they are accessible to anyone with an internet connection. Although some data may be encumbered temporarily upon request by the collector, over 99% of the records in Arctos are available publicly (0.6% encumbered); reasons for encumbrance may include sensitive species

or protection of intellectual property. Darwin Core Archives of Arctos data also accessible to be harvested and shared by other data portals through the VertNet IPT.

Arctos has a policy statement on Ownership and Use of Data that is posted on its website (<http://arctosdb.org/home/data/>) and has the following provisions: Arctos data and media are the property of the originating institution, with all rights reserved; Arctos records are intended for use in education and research; Arctos data may not be repackaged, redistributed, or sold in any form without prior written consent from the collection(s) holding the data; and users wishing to include Arctos data or media in publications, reports, websites, or other means of dissemination must acknowledge the provenance of the original data, cite the institutional catalog number(s), and notify the appropriate curator.

Arctos users can manage their own collection metadata by choosing this general policy statement or a more specific institutional or Creative Commons license. A link to usage rights is downloaded with every data record, and is part of the VertNet metadata for each resource.

4) Re-use and Re-distribution of data: As noted above, Arctos data cannot be repackaged or redistributed without prior written consent of the collection holding the data. However, all collections who use Arctos for their management system agree to share their data with VertNet and GBIF. Recent examples of other data projects that have harvested Arctos data from GBIF include BISON and an effort by the Nature Conservancy to map all known occurrences of freshwater species in major California watersheds.

Arctos has created views of its data and media for harvesting by VertNet, which hosts the different collection resources (Darwin Core Archives, collection metadata) on its IPT instance. Other data portals, including iDigBio, have access to these resources with permission from the collections. In addition, VertNet has an RSS feed that alerts subscribers to publication of a new or updated resource. The resources hosted by VertNet contain globally unique values for Darwin Core terms including occurrenceID that can be used to distinguish records from different publishers when data are harvested by iDigBio or other portals. Arctos already provides data to iDigBio for a Thematic Collection Network project that is digitizing specimens at the University of Alaska Herbarium (ALA). MVZ data will be available to iDigBio through the VertNet IPT.

5) Archiving data and samples for preservation of access: Data and media from MVZ and other Arctos collections are archived securely, with daily backups, at the Texas Advanced Computing Center (University of Texas, Austin). This facility is part of the NSF XSEDE project. TACC has hosted Arctos data for over two years (previously the data were hosted at a commercial facility in Fairbanks, Alaska), and provides infrastructural support including hardware, software, security, and some data processing (e.g., file format conversions for media). MVZ specimens themselves are stored safely in the museum on the UC Berkeley campus, where they are used widely for research, teaching, and public outreach. Staff Curators are members of professional society collection committees and the Society for the Preservation of Natural History Collections. The MVZ and UC Berkeley are committed to preserving access to the museum's collections through staff and budgetary support.