**Data Management Plan**

***1. Types of data and other research products***

This project is itself a kind of data management plan, because its aim is to integrate trait data for broad reuse by the research community. As such, we will produce limited new data *per se*, but we will be managing large quantities of data, as well as new code.

1. *Trait data.* We will compile, integrate, and serve a large number of measurement-based mammalian trait datasets. Many of these data come from previous NSF-funded projects and are already archived in some form. The FuTRES database will provide a central resource for recovering them. In addition to the data sets named in the project description, we expect other trait data compilations to be added by researchers during the course of the project. Our project does not deal with digital images, CT data, or other digital representations of objects.
2. *Code.* Our project will generate computer code for the database interfaces, including R packages.
3. *Ontologies.* We will build application specific ontologies for use within FuTRES by importing needed terms from existing ontologies such as UBERON (animal anatomy), PATO (qualities), and the BioCollections Ontology (biodiversity observations and specimens). We will contribue new ontology terms to external ontologies as needed.

***2. Data standards, formats, and metadata***

1. *Trait data and associated standards.* We will use existing observation and measurement standards (i.e. BCO) and existing ontologies for traits, organisms, and data about them. Where material or digital objects are associated with trait data (e.g., museum specimens or data from *VertNet*), we will associate the object’s URI or other identifier with the *FuTRES* data, so users can recover the object if desired.
2. *Metadata standards*. We still foresee a strong need to utilize both Darwin Core and Dublin Core. Darwin Core is a key standard that is already incorporated into BCO, and there is a need to link trait data back to Darwin Core records in VertNet. Dublin Core is critical for reporting dataset citation metadata, such as creator and license.
3. *Code.* Most of the data handling code will be develop in Python. We can do pre-reasoning and storing using typical data basing tools such as PostgresQL.
4. *Ontologies.* Our ontologies will be developed using OBO Foundry principles. We plan to especially extend BCO to broadly serve the broadest biodiversity user base. Ontologies are developed using the W3C standards Web Ontology Language (OWL) and serialized as RDF-XML or Terse Triple Notation (Turtle).

***3. Management and preservation of source and derived data***

1. A key goal of this project is to deliver a unique and immediately usable new trait data store but also assure that standardize trait data are provided back to aggregators and back to the data publishers from whom it was originally made available.
2. We will leverage resources at CyVerse, which is secure, connected to UAs high-speed network, and backed up at both the UA and the University of Texas’s Texas Advanced Computing Center. Leveraging virtual machines with significant CPU, RAM and storage, we will build the tools for pre-reasoning, data storage and APIs. The data store will be built on top of CyVerse’s iRODS-based Data Store, which permits fine-grained sharing permissions in the form of read/write/own/public access for any object (file or directory), allowing us to grant appropriate permissions to all team members and share with external collaborators as needed. All data products will be backed up regularly using a snapshot process.
3. CyVerse can help assure longer-term sustainability of the data. We also plan periodic large taxon-based slices or "snapshots" in order to assure long-term preservation of data (a similar process is already in place for VertNet). These will be published to CyVerse's Data Commons with DOIs for long term archiving. If FuTRES should be retired, full copies of all data would also be deposited using best practices archiving tools of the time.

***3. Roles and responsibilities***

Primary responsibility for data standards, long-term data archiving, and physical database hosting fall to Walls (PI, University of Arizona), assisted by the postdoctoral researcher. Compliance with metadata standards and ontology development in the OBO Foundry framework are areas of particular strength for Walls. Data management responsibilities secondarily fall to Davis (Lead PI, University of Oregon), and Guralnick (PI, University of Florida), especially regarding code developed under their direction.

Responsibility for processing data and ensuring it is properly entered into the FuTRESdatabase falls to all four PIs, the postdoctoral researcher, and the graduate student researchers. The graduate student researchers have the primary duty of processing data, which is coordinated by PIs and postdoctoral researcher. Should any of the PIs leave the project, the responsibility will fall to their successor on the project.

***4. Data dissemination***

1. *Trait data*. The FuTRESdatabase will be the primary method of dissemination of the trait data processed in this project. It will be an open-access resource from which trait data can be retrieved by members of the research community or other interested parties and to which researchers can contribute. The database is intended to be available indefinitely and all PIs are committed to long-term sustainability.
2. *Code*. All software developed for this project will be free and open source, distributed under either the 3-Clause BSD License or version 3 of the GNU General Public License. Source code will be deposited in publicly accessible repositories on GitHub.

***5. Policies for data sharing, public access, and reuse***

Trait data and code will be freely shared to all interested parties using the least restrictive license possible (e.g., ODC-PDDL). Already published datasets will be reused under the original licenses from data publishers. In VertNet, all datasets and individual records bear formal license information that will be transferred to FuTRES. There will be no additional restrictions to the use of the data. Responsibility for proper citation lies with the end-user, but we will make it as easy as possible by providing downloadable citations. In support of conscientious usage and attribution, we will promote data sharing norms that promote FAIR principles. All data accumulated from literature sources rather than specimens will cite that literature. Factual content derived from published content will note copyright or licensing of the original work.

***6. Plans for archiving data, samples, software, and other research products.***

1. *Code.* We will use GitHub to distribute and version the database code. We will use the Zenodo system to supply DOIs for our GitHubarchive and to make it available long-term. Zenodo is a service provided by the CERN Data Center with funding from the European Union and specifically provides support for archiving code distributed by GitHub*.*
2. *Trait data.* A copy of the full database, in Turtle format, as well as bi-annual data dumps, in the form of the database tables, will be maintained on the CyVerse Data Store
3. *Ontologies.* We will use GitHub to manage ontologies, providing regular, archived stable releases, as per OBO Foundry policies.