Data Management Plan

I. Products of Research: Our project will generate ecological data from observations, experiments to be conducted by REU students and other project personnel, code for conducting demographic analyses, and biological collections (voucher specimens). These will all be archived and shared in accordance with NSF policy, US and Brazilian legislation, and our belief that data archiving and open access are essential to advancing our discipline (Bruna 2010).

(a) Demographic data: Demographic data to be collected in 13 permanent plots in forest fragments and continuous forest at the BDFFP. We will collect the following data annually for each plant: if it survived to that year, its height to the tallest leaf, the number vegetative stems it has, and the total number of inflorescences (if any) produced during the flowering season. For a subset of plants we will collect data on the no. of flowers per inflorescence and the number of fruits produced per plant (counted 3-4 times during the reproductive season). Some student projects may require measuring leaf-length (for conversion to leaf area using a previously published regression equation) and seed germination success; in the past we have typically measured these monthly though this may vary based on the question being addressed.

(b) Environmental Data: Once during the growing season we will collect data on canopy cover using photos taken with fisheye lenses and use them to calculate light transmission to the understory. We will also collect data on temperature and relative humidity in all sites throughout the duration of our study using iButton data loggers (15 temp loggers per plot, 5 RH loggers per site); measurements will be made every 5 minutes but only the mean of the previous hour's measurements will be recorded. Finally, we will be collecting data on thermal regimes in the plots using pictures taken with infrared cameras. These photos, 4 of which will be taken at 66 points in the plot, will then be analyzed to pixel level to evaluate the mean, variance, and min-max of temperature recorded in each photo. This sampling will be done each year in each plot (once in the rainy season and once in the dry season).

(b) Curated collections of plant and invertebrate specimens: Any plants or insects collected during the study will be mounted (insects) or pressed (plants) and deposited in INPA herbaria and entomological collections with metadata described below.

II. Data Documentation, Processing, Storage and Preservation: ***(1) Demographic data, experimental ecological data, and data on abiotic conditions:*** Initial data entry and quality control will be done using spreadsheets saved as .csv files to cloud-based servers. Metadata recommended by Michener et al. (1997) will be saved as a .txt file. Collaborative development of R code for data manipulation and analyses will be done using GitHub; the verison of the code used for analyses in papers will be archived and given a DOI at Zenodo but will also remain at GitHub for use and development by others. We will store all data and metadata files and work collaboratively using the Center for Open Science's cloud-based Open Science Framework (link), which allows for version-control of datasets and manuscripts and allows us to immediately share individual datasets with those who request them. Data used in manuscripts will be archived at Dryad; any additional data collected but not used in a manuscript will be archived and made immediately available at Figshare within one year of the project's completion (see #4 below). All original datasheets and notebooks will be stored along with printed and pdf copies of the data and metadata in a secure location at each institution to ensure there is a backup of all data. Finally, As per NSF policy we will also store or mirror software, images, or digitized data in the National Resource for Digitized Collections at the University of Florida.

(2) Curated collections of plant specimens and invertebrates: Any plant specimens with flowers or fruits collected as vouchers during plant surveys will be deposited in the INPA herbarium. Any invertebrates collected will be mounted, labeled, and deposited in the INPA entomological collection. Though we do not anticipate the discovery of new species, holotypes of new species will be deposited in in Brazil as per Brazilian law. The data describing these collections will be organized using the open-source Biota software package (Colwell 2004) and made available to destination museum.

III. Data Formats and Metadata: All data and metadata will be archived in nonproprietary formats (e.g. .csv, .txt) to ensure flexibility and avoid loss due to software obsolescence (Borer et al. 2009). The use of EML to record ecological metadata will integrate our datasets with those from the Knowledge Network for Biocomplexity (KNB) and other DataONE nodes (www.dataone.org).

IV. Data Dissemination & Policies for Data Sharing: All data will be available for use by others, but the manner in which it is made available depends on when they are requested. ***Prior to publication:*** The PI's webpage has a section describing data access policies (link), but all such information will be transferred to the new bilingual Heliconia Project Website (link). We will request that users sign a General Data Use Agreement modeled on the one used by the LTER network (link); it defines conditions of use, citation, and co-authorship (if appropriate). ***Following manuscript acceptance:*** All data used in publications and code used for their analyses will be archived at the Data Dryad (www.datadryad.org) with accession numbers included in the publications *(see Bruna Biosketch)*. Once the embargo period has passed there will be no restrictions on their use by others. ***Access to unpublished datasets:*** Any data/metadata not included in publications will be available on Figshare within 1 year of the project's completion.

V. Roles and Responsibilities: Bruna will have overall responsibility for ensuring that project data are being collected and deposited as described. To ensure that all personnel are aware of the project's data management and archiving objectives, we will teach an annual workshop on data management and archiving at INPA in which all new personnel will participate.

LITERATURE CITED

Borer, E. T., E. W. Seabloom, M. B. Jones, and M. Schildhauer. 2009. Some simple guidelines for effective data management. Bulletin of the Ecological Society of America **90**:205-214.

Bruna, E. M. 2010. Scientific journals can advance tropical biology and conservation by requiring data archiving. Biotropica **42**:399-401.

Colwell, R. K. 2004. Biota 2: The Biodiversity Database Manager. Sinauer, Sunderland.

Michener, W. K., J. W. Brunt, J. J. Helly, T. B. Kirchner, and S. G. Stafford. 1997. Nongeospatial metadata for the ecological sciences. Ecological Applications **7**:330-342.