Expected Data Types

I. Products of Research: Our project will generate ecological datasets (e.g., demographic data, data on temperature and rainfall), computer code used to carry out statistical and demographic analyses with these data, and the results of these analyses. 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 has been collected in 13 permanent plots in the BDFFP's forest fragments and continuous forest reserves. During annual surveys, we collect the following data on individual plants: if they survived from one survey to the next, the height to the tallest leaf, the number vegetative stems a plant has, and the total number of inflorescences (if any) produced during the flowering season. For a subset of plants we 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). We also have data on seed germination rates in different habitat types, disperser feeding behavior and movement, pollinator visitation rates, etc.

(b) Environmental Data: Our analyses will require compiling and organizing daily rainfall data collected at each of the BDFFP field camps and data on temperature and humidity from the weather station located at the Manaus airport. We will also be organizing and archiving previously collected data on environmental conditions in plots that will be used in analyses (e.g., canopy cover and the canopy photos used to estimate it. soil properties in our plots, etc.).

Data Formatting Standards & 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).

Policies for dissemination and data-sharing

The PI's webpage and the Heliconia Project Website (http://brunalab.org/heliconia) describe the data access policies (http://brunalab.org/datasets). 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: If the paper including the dataset has yet to be published, we will request that users sign a General Data Use Agreement modeled on the one used by the LTER network. This agreement defines conditions of use, citation, and co-authorship (if appropriate). Following publication: As is our current practice (see Bruna Biosketch), all data used in publications will be archived at the Dryad Digital Repository (www.datadryad.org) and accession numbers will be included in the publications [Disclosure: Bruna is a member of the Dryad Board of Directors]. Once the embargo period has passed there will be no restrictions on their use by others. Unpublished datasets: Any data not included in publications (and their associated metadata) will be available on within 1 year of the project's completion (see II above).

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. We will store all data and metadata files 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 easily share individual datasets with those who request them. Datasets used in manuscripts will be archived at Dryad; any additional data collected but not used in a manuscript will be archived and made available at the UF Library Institutional Repository 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. The version of the code used for the analyses presented in papers will be archived and given a DOI at Zenodo, but will also remain at GitHub to allow for improvement by other researchers. 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.

Roles and Responsibilities

Bruna will have overall responsibility for ensuring that all personnel are aware of the project's data management and archiving objectives and that project data are being collected, curated, and archived as described.

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.

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.