7-20-2021

Collaborative Research: Ecological legacy effects of megacarcasses in African savanna ecosystems

Johan du Toit
Utah State University, johan.dutoit@usu.edu

Data Management Plan: Collaborative Research: Ecological legacy effects of megacarcasses in African savanna ecosystems

Types of Data

The research described in this proposal will generate multiple data types, including both biological samples and digital data. Biological samples for storage include clipped vegetation (ANPP sampling), soil cores, and leaf samples for nutrient analysis. Digital data include all measurements related to both the field sampling and greenhouse experiment as well as the digital images from game cameras used to survey vertebrate herbivores.

Field data: soil moisture, soil pH, soil nutrients (C, N, P, K, Na, Ca, Mg, Fe), ANPP, soil respiration, litter decomposition, foliar nutrients, plant community composition, woody plant density/identity, invertebrate herbivore abundance, vertebrate herbivore abundance, herbivory, herbivore dung density/nutrient influx,

Greenhouse data: soil moisture, soil pH, soil nutrients (C, N, P, K, Na, Ca, Mg, Fe), soil respiration, plant biomass, soil respiration, foliar nutrients, invertebrate herbivore consumption rates

Standards for Data

Sampling sites: All sites will be in Kruger National Park (KNP), South Africa. Fieldwork will be conducted with permissions and approvals of SANParks and will require that SANParks has full access to all data, metadata, and publications arising from fieldwork in KNP.

Megacarcass locations: Elephant carcasses have been georeferenced in aerial surveys over the past 40 years and the data will be made available to us by Scientific Services, SANParks, Skukuza, South Africa. We will select at least 50 megacarcasses as study sites. Locations (GPS coordinates) will be in spreadsheets as .csv files and only those used in this project will become part of the project data set.

Biological Samples: Biological samples will be stored in several different forms. Clipped vegetation from ANPP sampling will be kept as dried biomass in a climate controlled storage room. Vegetation for nutrient analyses will be ground for analyses and kept in air-tight scintillation vials in a climate controlled storage room. Soil cores will be dried and kept in air tight jars.

Digital data: All digital data will be entered into spreadsheets. For the field surveys, each response variable will have a unique file (e.g., one file each for ANPP, plant diversity, etc.) to be merged during data analysis. For the greenhouse experiment, a single spreadsheet will contain all relevant information for each replicate: time of harvest, plant biomass, soil N, soil N, foliar N, etc. Spreadsheets will be stored as .csv files. Each data file will be accompanied by a .txt document containing metadata (date of study, PI, column descriptions, units, etc.). Standard text formats (.csv, .txt) avoid using proprietary formats (.xls, .doc, etc) that may not be accessible for individuals who lack proprietary software, that may exhibit a lack of compatibility with newer or older versions of software, or that commonly suffer from formatting changes when accessed from different platforms. Raw data and metadata will be accompanied by a R script containing processing and analytical code, as well as a heavily annotated guide to the processing scripts (either a raw Markdown file or a PDF produced from the Markdown file). The annotated guide will contain comments, descriptions, and output (i.e., figures) needed to understand data processing and analysis.

Data Archiving

Biological Samples: All biological samples (e.g., ANPP plant clippings, soil cores, vegetation nutrient samples) will be stored for the duration of the project, and for one year post-publication, in climate controlled facilities at UCSB or at SAEON in South Africa (for those samples that would be onerous to import to the United States).

Digital Data: All data, R/Python scripts, and annotated guides will be stored in several locations. First, all data produced by this project will be stored on the hard-drive of PI Burkepile's main computer and backed up daily to three locations to guard against hard disk failure: a second solid-state hard disk, Burkepile's

could storage (Dropbox), and UCSB's cloud computing server (Box). After completion of the project, all files (data, scripts, and guides) will be archived in multiple online locations: (1) a public GitHub repository that allows anyone access to the data and scripts but requires permission of the repository owner (PI Burkepile) to upload changes, and (2) Dryad, a permanent digital repository. Most scientific journals like Ecology Letters, Functional Ecology, or PLoS One require permanent data archiving, so this step would likely be followed as a matter of publication before the end of the project.

Data Sharing, Public Access, and Dissemination Methods

Dryad and GitHub are open-access digital repositories, and Dryad provides a Digital Object Identifier (DOI) code to all datasets, making each dataset uniquely identifiable and citeable. Furthermore, all data stored on Dryad and GitHub are free to access to any public or private individual (no paid account needed, free to use). Thus, archived data from this project will be publicly available to any interested party and easily shared across institutions regardless of role (individual, corporate, state, and federal). The DOI number enables any person using the generated products to cite this proposal, and GitHub provides strict version control that can alleviate typos, errors, or other mistakes that could render a dataset useless if not backed up properly. In addition, data and/or scripts will be shared freely with any interested individual (after publication) who requests access by either providing that individual with a link/DOI to Dryad and GitHub, or via directly sharing requested items as email attachments. For the numerous camera trap images that we will generate, these will be stored in date-, time-, and plot-specific .jpg files to be made available online in the National Resource for Digitized Collections (iDigBio.org). All data will be made available to the scientific community by the end of the last year of the project, at the latest, and data used in publications before then will be made available upon each respective publication date.

Roles and Responsibilities

PI Burkepile will assume the primary responsibility for implementing the Data Management Plan. That is, Burkepile will store the data files on his institute's data server and cloud-based storage, and will be responsible for sharing cloud storage with collaborators. Burkepile will also be responsible for storing data products on two, solid-state backup hard disks or flash drives as backup copies. Finally, Burkepile will be responsible for digital data archiving on Dryad and GitHub.