

Data Management Plan

I. Types of Data

This project generates time- and location-stamped image files of natural resources in Delaware County, PA. The images serve as a record of the occurrence of creatures, natural artifacts, and conditions at specific places and times during the period 2003 through 2011. For many of the photos taxonomic information is also available. The occurrence data are observational and qualitative, and in addition to documenting the presence of various taxa they will be used for posters, books, and a web site promoting the appreciation of the natural diversity of the County and the preservation of its remaining natural areas.

II. Data and Metadata Standards

Data will be captured with a digital camera capable of creating images with sufficient taxonomic detail to allow identification to the species level for many taxa. Images are stored as JPG files with embedded EXIF and IPTC information describing the exposure, camera type, lens, and metering mode, along with photographer-generated cataloging tags that may include taxonomic identity of the organism(s) in the photograph. Images will be stored in a date-hierarchical file structure (year, date) on redundant disk drives and managed using the commercial iMatch software. Images will be maintained with unique file names. The dataset contains only photos of reasonably high quality, and quality flags are used to differentiate images of varying quality. J. Smith will be responsible for the data management during the course of the project.

Metadata about timing and exposure of individual images is automatically generated by the camera at the time the photo is taken. GPS locations are subsequently added by post-processing GPS track data based on shared time stamps. GPS data stored in image files depicting rare or locally sensitive species will be obfuscated in the file metadata but can be made available for appropriate, approved uses. Metadata for the image dataset as a whole will be generated by the image management software (iMatch) and will include time ranges, locations, and a taxon list. Those metadata will be translated into Ecological Metadata Language (EML), created using the Morpho software tool, and will include location and taxonomic summaries.

III. Policies for Access and Sharing and Provisions for Appropriate Protection/Privacy

The image collection will be made available beginning in 2015 for sharing with taxonomists, biodiversity study investigators, photographers, children, teachers, and other interested members of the public. They will be available as digital photographs viewable on the web in a restricted form that prevents downloading.

IV. Policies and Provisions for Re-Use, Re-Distribution

Summaries of temporal and spatial distributions of taxonomic groups will be also be made available on Smith's website for use without permission, but users will need permission to access the original high resolution photographs from which these distribution data were derived. Smith will retain copyright and the originals will be licensed using a Creative Commons license (Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License). Data can be cited by referring to the image website, and individual images can be identified by their unique filename

V. Plans for Archiving and Preservation of Access

Long term storage will involve exporting the metadata stored within the JPG files (location, EXIF, and tag information) into text files to facilitate access to that metadata and as a safeguard should the specifics of JPG metadata structure change. The image files will be stored in a single zip file containing the date-hierarchical file structure. As noted above, EML metadata will be created and the resulting dataset will be submitted to the Knowledge Network for Biocomplexity (KNB) archive.