Funded Research Records

2017

Uinta-Wasatch-Cache Rare Plants Species Inventory and Distribution Models

Thomas C. Edwards
Utah State University, t.edwards@usu.edu

Recommended Citation

DATA MANAGEMENT PLAN
Uinta-Wasatch Cache Rare Plants Species Inventory and Distribution Models
_____

Project principal	Thomas C Edwards
Title/Name assigned to grant	Uinta-Wasatch Cache Rare Plants Species Inventory and Distribution Models
Primary institution name	Utah State University
Project start and stop dates	05/09/2017-04/30/2020
Granting Agency, grant award number	17-CS-11041914-037

1.   Description of Data

Data formats will vary principally among GIS- and spreadsheet-based formats. All GIS data will be saved in formats amenable to import, manipulation, and transfer in non-proprietary software. These formats are also amenable to proprietary ArcGIS. Basic formats includes ascii grids for raster based data, and generic shapefiles. GIS manipulations will occur in combinations of R, IDL, or GRASS, and will be portable to ArcGIS environments. All code used in GIS manipulations will be archived along with any imported or derived GIS structures. Although much extant spreadsheet data is delivered in proprietary MS-based formats (e.g., EXCEL), all data will be stored and exchanged in flat ascii files as well.

To the extent possible, all statistical and GIS analyses will be performed in R, of which there exist numerous different packages. Given the rapidity with which R packages advance, it is impossible to state explicitly which packages will be used. Some logical R packages, and in which the project principal is familiar, include Raster, rgeos, sp, and others for spatial manipulations and analyses. Some python and java scripting will likely occur. Custom R code for manipulation is likely to be written as well. All written code will be extensively notated to provide sufficient information for others to use post-project.

Data quality is largely a function of securing appropriate metadata for extant datasets used in our research. Such metadata will be archived for all extant data acquired, and developed for primary data. No matter the source, all extant data will be screened using standard summary statistics (central tendency, dispersion, ranges) to determine if data anomalies exist, as well as subjected to QA/QC processes described in Section 1 above. Options for identified data anomalies in extant data are somewhat limited, although the project principal will strive to clear any identified data concerns with the original data source(s).

Estimation of data uncertainties is inherent in our proposed analyses, which rely on SDMs whether predicting a nominal (presence:absence) or count (abundance) response. Base spatial

_____

product errors (e.g., [X,Y], and Z where appropriate) will largely come from the metadata of acquired extant spatial map products. The proposed use of ensembles for spatial analyses will also provide estimates of uncertainty given a set of different models overlapping space. Such errors can be readily mapped as concordance products for categorical map products, or a variety of quartile or other binned groupings for continuous variables. All reports and map products will includes written or mapped estimates of uncertainties.

2. Metadata Content and Format
All spatial products metadata will meet Federal Geographic Data Committee (FGDC) standards (http://www.fgdc.gov/metadata).   Metadata will be generated from one of many possible tools, depending on data characteristics, such as the function spMetadata-methods (R package:plotKML) or the USGS Online Metadata Editor (OME) https://www1.usgs.gov/csas/ome/. As noted above, all nonspatial data (e.g., analytical code) will be notated for access and use by others post-project.

3. Compliance Policies for Data Access, Sharing, and Re-use
No ethical, privacy or copyright issues are foreseen regarding the extant spatial data used in this proposal. Raw, extant, georeferenced location data on plant species is under the direction of the U.S. Forest Service (USFS) (M. Duncan *per* proposal, USFS Project Liaison) and cannot be released by the PI Edwards to third parties except by permission of the USFS. It has been agreed that plants may be assigned to the center [X,Y] of the underlying 1-km grid; this georeferenced point location may be released without permission from the USFS.

No restrictions exist on summary and derived spatial data other than a documented request be made to the PI Edwards. Edwards requests that intellectual publication rights remain with the project for up to a maximum 18 months post-project completion for publication purposes.

4. Long-term Storage and Archival
All data will be backed up and stored on at least two redundant (segregated) systems at Utah State University. Once the project is completed, all data and links will be made freely available, such as through Utah University Digital Commons (http://digitalcommons.usu.edu/) subject only to the intellectual and other caveats described in Section 3, and the data acceptance caveat directly below.

Such release on publically accessible links will occur only after the data have been accepted by the USFS as having met their internal standards. Responsibilities for data preparation and documentation will be the primary responsibility of PI Edwards.