IEP Data Management Plan

Basic Information

Year: 2020; PEN: 337; Date Updated: 2019-04-30; Start Date: 2019-01-01

Study Title

Forecasting Nutria Invasion in the Sacramento-San Joaquin Delta

Principal Investigator

Individual(s) responsible for the project. Include name, agency, e-mail, & phone.

Vanessa Tobias, USFWS, vanessa_tobias@fws.gov, (209)334-2968 ext. 404

Point of Contact

Individuals who data users should contact for access to the data or questions about the data. Include name, agency, e-mail, & phone number or write "same as above."

Vanessa Tobias, USFWS, vanessa_tobias@fws.gov, (209)334-2968 ext. 404

Data Description

A very brief description of the information to be gathered; the nature and scale of the data that will be generated or collected. Include approximate size (in MB) of the resulting data set.

This project will identify impacted and at risk habitats and quantify the rate of advancement in the Delta. To do this, we will identify impacted habitat and forecast distributions of nutria in the Delta before they impact sensitive habitats or water operations. It will use existing data to produce models, maps, and tools for detection. Tasks included in this project include: (1) Identifying potential habitat, (2) summarizing vital rates, and (3) forecasting dispersal through potential habitats. We anticipate that data will be large (approximately 5-10 GB).

Related Data

Optional. Existing datasets that you incorporate into analysis and reporting for this program element, existing data that are relevant to your study, or data that are collected simultaneously.

No new data will be collected, but data from public sightings and trapping efforts will be incorporated. IEP's water quality data, fish distribution data, and products from prior synthesis efforts may also be incorporated. Nutria trapping data (CDFW's Nutria Response Team); vegetation maps (likely derived from CDFW's VegCAMP datasets); water body maps (likely USGS); other data relating to habitat, which we will update as the project gets underway.

Metadata

A description of the metadata to be provided along with the generated data, including the metadata standards used. Provide the file name and information on how users can access the metadata (e.g., a link).

Metadata will be generated for derived datasets as they are compiled, following IEP's DUWG metadata standards. Metadata will be available on the shared data repository while the project is active and the data archiving site when data are published.

Storage and Backup

A description of the short-term storage methods and backup procedures for the data, including the physical and electronic resources to be used for the short-term storage of the data.

All data will be in electronic formats; no physical copies will be part of this project's workflow. Collaborators will back up working copies of data and code on cloud-based drives that are managed by their respective agency IT departments.

Archiving and Preservation

The procedures for long-term archiving and preservation of the data, including succession plans for the data should the expected archiving entity go out of existence.

Derived datasets created for this project will be archived in an open data repository, following IEP's Data Utilization Workgroup's guidelines. We anticipate using EDI (environmentaldatainitiative.org) to publish data and obtain a DOI. It is possible that archiving data products that include specific locations of nutria captures may not be allowed because some individual animals were captured on private land. We are investigating this and will update our DMP as more information becomes available. Derived datasets will be archived no later than the time of report writing and/or manuscript submission. Collaborators will use a shared data repository (GitHub, Bitbucket, or similar) to ensure that all collaborators have access to interim data and code, in case of succession.

Access and Sharing

A description of how data will be shared. Include (1) access procedures, (2) embargo periods, (3) technical mechanisms for dissemination (e.g., website addresses, listserv information), (3) whether access will be open or granted only to specific user groups, and (4) a timeframe for data sharing and publishing.

Compiled datasets will be freely available and published on EDI (environmentaldatainitiative.org). There will be no embargo period. We will announce the publication of data to the IEP and stakeholders via email at the time of publication. Derived datasets will be archived no later than the time of report writing and/or

manuscript submission. We anticipate publication of habitat data at the end of 2019 and forecasting data by the end of 2020.

Format

Formats in which the data will be generated, maintained, and made available. Include BOTH general data type (e.g., spreadsheet, relational database) and file format (extension).

The potentially large size of compiled datasets is a concern so we will choose file types that preserve data quality while minimizing file size. Data will be generated, maintained, and made available in open formats whenever possible. We anticipate that tabular data will be flat files (.csv), code files will be R format (.R; readable in any text editor, executable in program R); raster data will be images (.TIFF) or flat files (.csv); vector data will be shapefiles (.shp, with supporting suite of files) or GeoJSON (.geojson).

Quality Assurance

Brief description of procedures for ensuring data quality. Provide links to Quality Assurance Project Plan and/or QA/QC Standard Operating Procedures.

QA/QC documentation will be obtained for input datasets and data will be inspected for consistency prior to use, using standard data exploration methods. We will prioritize use of datasets that cover the full geographic extent of our project and those with higher georegistration accuracy. We will also prioritize using datasets that have been reviewed and well-documented for QA criteria for geospatial data projects. Derived datasets produced by this project will also be inspected to ensure proper integration among input datasets. Derived datasets will be published with the code that generated them to facilitate review of the quality of the data that is produced.

Rights and Requirements

A link to or instructions to locate the agency's rights and requirements for data use.

Data use rights and requirements for federal government datasets can be found online.