

# IEP Data Management Plan

## Basic Information

Year: 2020; PEN:339; Date Updated: 2018-09-20; Start Date: 2019-01-01

## Study Title

Landscape-scale analysis of aquatic vegetation response to treatment in terms of growth rates, persistence, community composition, and biodiversity in the Delta

## Principal Investigator

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

Shruti Khanna, California Department of Fish and Wildlife,

2109 Arch Airport Road, Suite 100, Stockton, CA 95206

Email: shruti.khanna@wildlife.ca.gov; Phone: (209) 689-8719

## 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."*

Same as above.

## 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.*

Three datasets will be used in this study: 1) The UCD CSTARS lab has HyMap 128-band hyperspectral imagery acquired at 3x3m pixel resolution (2004-2008; every summer) and AVIRIS-NG 432-band imagery acquired at 2.5x2.5m resolution (2014-present; every fall). Size: >15TB. In addition to this data, classified maps of invasive species will be produced as both rasters and vectors. 2) The species-level field data collected concurrent with imagery acquisitions since 2003. Size: <1GB. All of these datasets are currently located on UCD servers and will be available through UCD CSTARS lab upon request.

## 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.*

The herbicide treatment records (treatment locations, dates, and quantity), will be available from the Division of Boating and Waterways (DBW) for both submerged and floating species dating back to 2003. Size: <1GB. Data located at DBW office in downtown Sacramento.

## 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).*

Two dataset products will be published by June 2020. Since both are geospatial in nature, the FGDC (Federal Geographic Data Committee) will be used.

1) Invasive species classifications (multiple years) will include information about target classes, imagery dates, spatial, spectral and temporal characteristics of imagery, preprocessing steps conducted, projection, and accuracy of classification.

2) Field data polygons (multiple years) will include information on all species observed in a patch, cover of species, % dry/senescent, % water in patch, % algae in patch, sechhi depth, submerged mat depth, date and time, rake composition, density of patch, length and width of patch, flowering or weevil damage, etc. Filename and link made available by June 2020.

## 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 stored on UCD CSTARS servers and on external hard drives in CDFW in the PI's cubicle. UCD servers are backed up every 15 minutes. Final products will be uploaded to an open data repository by June 2020 and will be stored on their servers. The repository is yet to be decided based on space required, metadata standard mandated, etc.

## 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.*

UCD CSTARS archive data on hard drives once a project is complete. CDFW will also maintain the dataset on external hard drives. Additionally, the product datasets, once published online, will be available as long as the data repository is in existence. Extensive documentation and metadata will allow future PIs to easily download and use data. Until the data are published (by June 2020), contact Shruti Khanna for data accessibility.

## 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.*

Access to classification products and field data will be through open data repositories. The original imagery is available publicly through the [Jet Propulsion Laboratory's JPL](#) website or from CSTARS (contact PI) upon request.

There is no embargo period but products will likely only be available at the end of the study period.

Website URLs will be available after datasets are published. Data will be made publicly available to anybody who wants it.

Datasets will be uploaded to repositories by June 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).*

Field data will be captured in Trimble files format (.ssf) and exported out as shapefiles.

Field data will be made available as shapefiles (.shp; .dbf; .shx, etc.) with attribute tables and detailed metadata.

Original image data will be made available in standard ENVI format with a header file and a data file (.hdr and .img).

Classification data will be made available as rasters (standard ENVI format and/or TIFFs) and as shapefiles.

## Quality Assurance

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

The field data are overlaid on the image data (once they are available) and corrected for accuracy. Data points whose attributes do not match with the pictures taken in the field are flagged. For the rest of the data points, information on patch size is used to make polygons around the target patch. Any line data collected in the field is also converted to polygons and checked for accuracy at the same time. There is no checking needed for values input into the GPS unit since all attributes are already coded in the data dictionary. There are lists for all attributes, ranges of values where applicable, hence there is no possibility of spelling errors.

For the classification datasets, accuracy will be assessed using standard methods such as confusion matrices and Kappa co-efficients. These will be published along with the dataset so users know the accuracy of the published maps.

## **Rights and Requirements**

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

The rights and requirements for data use will follow the guidelines of the [CC-BY 2.0 license](#):

This license allows a user to share, copy, adapt, and/or redistribute the data, provided that the user gives an appropriate citation. We request that users cite the DOI as a link to facilitate data use tracking when a DOI is available (by June 2020).