IEP Data Management Plan

Basic Information

Year: 2020; PEN: 340; Date Updated: 2019-05-30; Start Date: 2020-01-01

Study Title

Understanding Climate Change Tools for San Francisco Estuary Analyses and Investigation of Thermal Refugia in Warming Waters

Principal Investigator

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

Larry Brown (USGS, lrbrown@usgs.gov, 916-278-3098; Bruce Herbold (Consultant, bherbold@gmail.com); Ted Sommer (DWR, Ted.Sommer@water.ca.gov, 916-376-2792); Catarina Pien (DWR, Catarina.Pien@water.ca.gov)

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."

Catarina Pien (DWR, Catarina.Pien@water.ca.gov)

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 is a synthesis project, leveraging existing water temperature data from the network of high-frequency monitoring stations in the Delta, Suisun region, and San Francisco Bay. All data will be gathered directly from the California Data Exchange Center (CDEC), and the National Weather Information System (NWIS). The data from these sources will be integrated into a single dataset, and QA/QCed. The estimated size of the final, integrated dataset is 3-4GB.

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 integrated water temperature dataset is relevant to discrete water temperature data collected by IEP routine surveys. However, there are currently no plans to incorporate data from discretely collected data. Data from NOAA/SFNERR and local universities

may also be added to the integrated water temperature dataset to increase geographic representation of water temperature data.

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 include a detailed description of the exact data sources (all stations included in the integrated water temperature dataset, with their associated periods of record). Metadata will also include a detailed description of how the QA/QC protocol used to arrive at the final dataset used for data visualization and analysis. All of the data will be retrieved, cleaned, and subjected to QA/QC using a reproducible script in R. This R script will be publicly available at the IEP Open Synthesis GitHub site (https://github.com/IEP-Open-Synthesis). In addition to detailed metadata on data sources and temperature sensors (excel table), the IEP-EDI metadata template (word document on IEP GitHub) will be completed for the study in preparation for publication of the dataset to EDI. The IEP-EDI metadata template is standardized to Ecological Metadata Language (EML).

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.

The water temperature dataset, and associated electronic files (e.g. R scripts) will be kept at a Networked drive at DWR and on a cloud drive. The DWR Networked Drive: S:\M & A BRANCH\[final file name TBD]. (automatically backed-up). Information managers for each individual water quality monitoring station can also be contacted for data.

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.

The data, metadata, and key products will be placed in a shared, networked drive at DWR-DES.

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.

The integrated water temperature dataset, as well as the R script for producing the dataset will be publicly available at the IEP Open Synthesis GitHub site (https://github.com/IEP-Open-Synthesis). Raw data from individual stations is available on CDEC or other publicly accessible sites. Original data sources will be listed in the metadata file.

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 integrated dataset will be in a spreadsheet format (.csv).

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 QA/QC protocol is still under discussion with the DWR QA/QC group, but will follow protocols of information managers for high-frequency water quality monitoring stations in combination with cited literature for appropriate filtration methods. The protocol will be detailed in the metadata document.

Rights and Requirements

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

All data will be drawn from publicly accessible sources. DWR does not currently have a description for QA/QC or data requirements.