

IEP Data Management Plan

Basic Information

Year: 2020; PEN: 342; Date Updated: 2019-06-03; Start Date: 2019-01-02

Study Title

Spatio-temporal Community Patterns for Early Life Stages of Fishes and their Associations with Zooplankton in the Upper San Francisco Estuary

Principal Investigator

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

Gonzalo Castillo, U.S. Fish and Wildlife Service E-mail: gonzalo_castillo@fws.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."

Gonzalo Castillo, U.S. Fish and Wildlife Service E-mail: gonzalo_castillo@fws.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.

The field data to be used includes species relative abundance (fish, zooplankton phytoplakton), other covariates (chlorophyll a, water quality, flows). Information will be obtained from various IEP agency websites: [20 mm survey & Spring Kodiak Trawl survey](#).

[DayFlow](#)

[Fall Midwater Trawl survey](#)

Environmental Monitoring Program (EMP) at least since 1995 through 2017 will be requested to EMP staff (phytoplankton, chlorophyll a). The entire IEP data set to be gathered is estimated to be 80 to 120 MB.

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 additional data sources are anticipated (already collected or to be collected).

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).

Detailed description of the data collected, compiled, and analyzed will be included in two final manuscripts. Because the period of data analyses for this study will not extend beyond 2017, no separate metadata plan or specific metadata standards are necessary to accomplish all study objectives. We plan to publish two open access peer reviewed articles.

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.

Data will be backed up in the short- and long-term on the PIs networked shared drive at USFWS which has automatic back-up routines. The datasets to be generated will be backed in a folder named "20 mm Community Analyses 1995-2017". The queries used to derive the data used for this study and a backup of the resulting flat file will be stored on the CDFW Stockton office shared drive and backed up on an environmental scientist's hard drive.

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 generated datasets will be archived and preserved on the USFWS and CDFW networked drives. The succession plan for the data is co-equally shared by the two agencies and is not expected to change in the event the principal investigator, Gonzalo Castillo, or another collaborator are not present in their current positions, the remaining collaborators will determine where the files will be stored.

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.

Because this study will utilize data collected by existing IEP long-term projects, the availability of data to the public will depend on each IEP agency and specific program. We anticipate all datasets used in our study will continue to be publicly available either online or by request.

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).

Data will be in a csv (.csv), Excel (.xlsx), or Access (.mdb) format. Data analyses will involve a variety of community and statistical analyses programs. Results will be available from the PI and collaborators upon request.

Quality Assurance

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

We will ensure data has been through QA/QC procedures that are routine for each individual dataset prior to data analyses. For example, in the case of the 20 mm survey, all collected CDFW field data is reviewed by a field crew leader for legibility, completeness, and accuracy. Data entry accuracy is checked by comparing listings of entered data against the data sheet entries twice immediately after their entry into the local database. Data entry is checked again after the end of the field season. The project ES performs reasonableness checks by running computer queries that flag outliers or erroneous entries. After these data checks are complete, a project lead (ES, environmental scientist) creates a flat file for this study to be independently reviewed for accuracy by another ES in the Stockton office. Quality assurance procedures for each dataset used in this study are available upon request from each IEP Program. Preliminary analyses from compiled data sets generated in this study will be reviewed by the PI and collaborators based on Cox et al. 2001 ([Best practices](#) for preparing ecological data sets to share and archive are found online).

Rights and Requirements

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

Data management procedures are informed by:

- 1) CDFW Department Bulletin Number 2017-02 "Department of Fish and Wildlife Scientific Integrity Policy" (CDFW Document Library <https://nrm.dfg.ca.gov/FileHandler.ashx?DocumentID=149016&inline>)
- 2) [Data use rights and requirements for federal government datasets](#) can be found online.