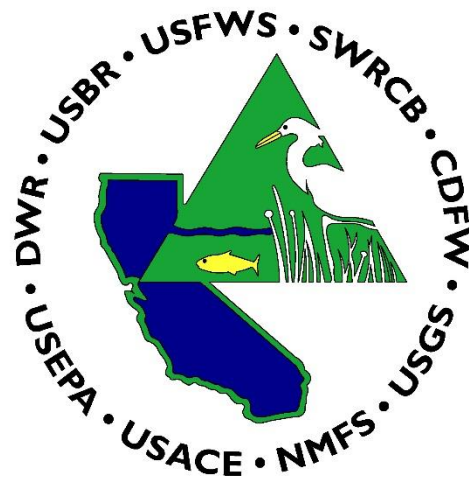


# Historical Metadata Template

*Version 1.0*



## Interagency Ecological Program

COOPERATIVE ECOLOGICAL  
INVESTIGATIONS SINCE 1970

*Last Updated: 10/16/2023*

## ***Suggested citation***

Data Publication Working Group. "Historical Metadata Template". Version 1.0. Interagency Ecological Program, 17 Aug. 2023, <https://iep.ca.gov/Data/Data-Utilization-Working-Group>. Date of Access.

Version Number	Date Created	Description of changes	Justification for changes	Version editor(s)	Contact information
1.0	10/16/2023	Document created		DPWG	trinh.nguyen@wildlife.ca.gov

# Table of Contents

Introduction .....	4
Conceptual Model.....	5
The four components to describe historical metadata .....	5
Fish data .....	6
Phytoplankton and Zooplankton data.....	7
Environmental and water quality data.....	9

# Introduction

**Vision:** This template provides guidance to improve documentation of historical metadata of IEP surveys. Complete, transparent, and standardized metadata is essential for data to be accurately analyzed and interpreted over time. Our vision is an easy-to-follow document that facilitates comprehensive metadata capture to enable thorough analysis and interpretation of IEP data.

**Definition:** Historical metadata describes the intentional and unintentional changes to established methodologies throughout the history of a survey.

**Scope:** This template is deliberately limited to changes that can directly affect data analysis and interpretation. While broader procedural changes may occur, the document focuses on metadata elements that provides crucial context needed to effectively evaluate data trends and quality over time.

**Disclaimer:** Some surveys may already document historical metadata within their metadata documentation. This template is meant to guide surveys that do not currently do this. However, we still advocate for all surveys to adopt the framework provided here, as it is helpful to end users to navigate a standardized approach across surveys.

**Usage:** The template identifies elements for surveys to consider when documenting their historical metadata. Guidance is provided specifically for fish, phytoplankton and zooplankton, and environmental and water quality data. Elements are repeated across these data types. We intend that surveys only refer to the data section(s) that is/are relevant to their data. Surveys with both biological and water quality data should consider including both sections.

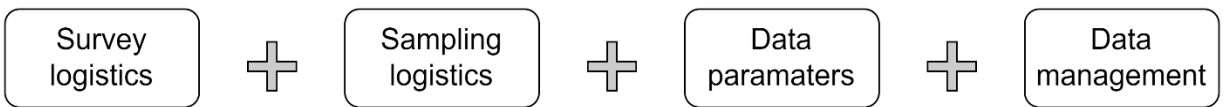
**Supplementary documents:** Surveys should reference or link to external documents if additional context is relevant for understanding the project's evolution. Examples include:

1. the project description of the original study design
2. special studies documentations
3. current metadata documentation

# Conceptual Model

This template outlines components that the Data Publication Working Group (DPWG) subgroup of the Data Utilization Work Group (DUWG) recommends IEP surveys consider reporting to adequately document the historical metadata of their datasets. Specifically, historical metadata is defined as the intentional and unintentional changes to established methodologies throughout the history of a survey. Emphasis is placed on changes or deviations that can influence the analysis and interpretation of the data. This template is intended to produce metadata that is in addition to any existing metadata documentation of current methods.

## The four components to describe historical metadata



1. Survey logistics: changes in the spatial and temporal extent or study design of the survey
2. Sampling logistics: changes in how the samples are collected
3. Data parameters: changes in what variables are measured to describe the samples
4. Data management: changes in how survey information is organized and published

All components outlined below should be viewed as changes to these components over time. All examples are for demonstration purposes only.

# Fish data

1. Survey logistics:
  - a. Sampling region (spatial)
    - i. Addition, deletion, or shifting of station(s)
    - ii. Changes to how stations are selected (e.g., loss of access to site, fixed versus random stations, changes to methods for choosing random stations)
  - b. Sampling season (temporal)
    - i. Extension or reduction of the season
    - ii. Changes in general start and end times of a sampling day (e.g., beginning before sunrise)
  - c. Sampling frequency (effort)
    - i. Changes in sampling effort throughout the season (e.g., biweekly to daily sampling in December, number of tows per station, or sample processed per station)
2. Sampling logistics
  - a. Trawl changes: e.g.,
    - i. Changes in targeted tow speed
    - ii. Changes in targeted tow duration
    - iii. Changes in the number of tow replicates per station
    - iv. Others
  - b. Gear changes: e.g.,
    - i. Changes in net size, mesh, composition, cod ends
    - ii. Changes in sled configuration
    - iii. Changes in rope composition
    - iv. Changes in type of boat used (e.g., “BigBoat” decommissioned in 1975, changes in size, horsepower, hull type, etc.)
    - v. Changes in data instruments and their calibration (e.g., a Van Dorn was used until 1989)
    - vi. Additions or discontinuations of gear types (e.g., a plankton net was deployed until 1989)
    - vii. Others
3. Data parameters
  - a. General
    - i. Start or end dates of recording a variable (e.g., started recording distance towed on 01/01/2020)
  - b. Environmental
    - i. See [Section Environmental and Water Quality](#) for further details
  - c. Taxa
    - i. Start or end dates of records of a taxon (e.g., shrimp only counted after 1983)
    - ii. Exclusion of taxa from data collection (e.g., jellyfish not enumerated from 1990-1991)

- iii. Changes in measurements protocols (e.g., transition from measuring fork length to standard length after 1995)
- iv. Changes in specimen preservation protocols (e.g., transition from formalin to dry ice after 1998)
- v. Changes in subsampling protocol
  - 1. Number of individuals measured
  - 2. Treating plus counts (e.g., all enumerated or a volumetric approximation)
- vi. Changes in QAQC protocols of taxon identification
- 4. Data management
  - a. Data collection and entry protocols (e.g., written to electronic datasheets)
  - b. Database software
    - i. Changes in software use (e.g., Microsoft Access to SQL Server)
    - ii. Changes in calculation protocols (e.g., removing special studies prior to calculations)
    - iii. Changes in data structure (e.g., relational to nonrelational database)
  - c. Changes in QAQC protocols of data entered (e.g., addition of an outlier detection system)
  - d. Changes to methods for any calculated parameter (e.g., CPUE or abundance indices)
  - e. Changes in data publication protocols, e.g.,
    - i. How often is data published?
    - ii. Where is data published?

## Phytoplankton and Zooplankton data

- 1. Survey logistic
  - a. Sampling region (spatial)
    - i. Addition, deletion, or shifting of station(s)
    - ii. Changes to how stations are selected (e.g., loss of access to site, fixed versus random stations, changes to methods for choosing random stations)
  - b. Sampling season (temporal)
    - i. Extension or reduction of season
    - ii. Changes in general start and end times of a sampling day (e.g., beginning before sunrise)
  - c. Sampling frequency (effort)
    - i. Changes in sampling effort throughout the season (e.g., biweekly to daily sampling in December, number of tows per station, or sample processed per station)
- 2. Sampling logistics
  - a. Trawl changes, e.g.,
    - i. Changes in targeted tow speed change
    - ii. Changes in targeted tow duration change
    - iii. Changes in the number of tow replicates per station

- iv. Others
- b. Gear changes, e.g.,
  - i. Changes in net size, mesh, composition, cod ends
  - ii. Changes in sled configuration
  - iii. Changes in rope composition
  - iv. Changes in type of boat used (e.g., “BigBoat” decommissioned in 1975, changes in size, horsepower, hull type, etc.)
  - v. Changes in data instruments and their calibration
  - vi. Additions or discontinuations of gear types (e.g., a Clarke-Bumpus net was deployed after 1989)
  - vii. Others
- 3. Data parameters
  - a. General
    - i. Start or end dates of recording a variable (e.g., started recording distance towed on 01/01/2020)
  - b. Environmental
    - i. See [Section Environmental and Water Quality](#) for further details
  - c. Taxa
    - i. Start or end dates of identification of a taxon (e.g., identification of *Pseudodiptomous forbesi* to the species level after 1994)
    - ii. Exclusion of taxa from data collection (e.g., all non-clams not enumerated from 1990-1991)
    - iii. Specimen collection protocols (e.g., sweep nets as a sampling gear)
    - iv. Taxa identification protocols, e.g.,
      - 1. Magnification of microscope used (e.g., 10x, 100x)
      - 2. Subsampling protocol (e.g., use of a sorting tray, individuals randomly chosen)
      - 3. Number of individuals measured (e.g., minimum of 20 individuals must be measured)
      - 4. Treating plus count
      - 5. Counting of linked/chain-producing phytoplankton (natural units) (e.g., component units counted rather than chains as whole beginning 2021)
      - 6. Counting of invertebrate parts (e.g., heads are counted while other body parts are ignored)
    - v. QAQC protocols of specimen ID
- 4. Data management
  - a. Data collection and entry protocols (e.g., written to electronic datasheets)
  - b. Database software
    - i. Changes in software use (e.g., Microsoft Access to SQL Server)
    - ii. Changes in calculation protocols (e.g., removing special studies prior to calculations)
    - iii. Changes in data structure (e.g., relational to nonrelational database)



- c. Changes in QAQC protocols of data entered (e.g., addition of an outlier detection system)
- d. Changes to methods for any calculated parameter (e.g., CPUE or abundance indices)
- e. Changes in Data publication protocols, e.g.,
  - i. How often is data published?
  - ii. Where is data published?

## Environmental and water quality data

This section describes components to report for both environmental data collected by surveys and by continuous data gages. Data relevant to this section can be biotic or abiotic data that has been collected to describe the samples taken. Studies collecting both biological and environmental and water quality data should consider including both sections and repeating information in both sections.

1. Survey logistics
  - a. Sampling region (spatial)
    - i. Addition, deletion, or shifting of station(s)/sensor(s)
    - ii. Changes to how stations are selected (e.g., loss of access to site, fixed versus random stations, changes to methods for choosing random stations)
  - b. Sampling season (temporal)
    - i. Changes in general start and end times of a sampling day
  - c. Sampling frequency (effort)
    - i. Changes in sampling effort throughout the season (e.g., biweekly to daily sampling in December, number of tows per station, or sample processed per station)
2. Sampling logistics
  - a. Changes in sampling depth (e.g., top/bottom of the water column)
  - b. Changes in instruments
    - i. Sampling method (e.g., sensor changed)
    - ii. Model number
    - iii. Reporting limits (e.g., 10-30°C to 0-60°C)
    - iv. Estimates of precision
  - c. Changes in instrument calibration methods
  - d. Changes in sample processing
    - i. Water grabs (e.g., how grabs are taken)
    - ii. Processing logistics (e.g., water grabs vs continuous sampling)
    - iii. Laboratory analytical methods
3. Data parameters
  - a. Start or end dates of collection per parameter
  - b. Changes in data resolution (e.g., daily to hourly data)
  - c. Changes in data units (e.g., NTU to FNU)
4. Data management

- a. Data collection and entry protocols (e.g., written to electronic datasheets)
- b. Database software
  - i. Changes in software use (e.g., Microsoft Access to SQL Server)
  - ii. Changes in data structure (e.g., relational to nonrelational database)
- c. Changes in QAQC protocols of entered data e (e.g., addition of an outlier detection system)
- d. Changes in Data publication protocols, e.g.,
  - i. How often is data published?
  - ii. Where is data published?