Data Management Plan: California Cooperative Oceanic Fisheries Investigations (CalCOFI)

NOAA Data Sharing Template

I. Type of data and information created

What data will you collect or create in the research?

Contextual statement describing what data are collected and relevant URL (IOOS Certification, f 1. ii)

CalCOFI Data Reports are a compilation of the physical, chemical, and biological data that is collected on quarterly cruises. These reports include materials & methods, reference citations, plots of cruise track, personnel, physical & biological parameters, tabular physical and productivity data, zooplankton biomass, and avifauna (bird) observations. Supplementing these reports, electronic hydrographic bottle data, CTD cast data and surface underway data are publically available from each cruise.

What data types will you be creating or capturing?

CalCOFI imports all of the final hydrographic data into a Microsoft Access Database. Then, csv's, XML and SQL files are generated from this database. Column Definitions for Cast and Bottle Tables: **http://www.calcofi.org/new.data/index.php/reporteddata#database**

CalCOFI CTD Data are available in three formats:

1. CTD cast files
2. CTD preliminary files
3. CTD final files.

How will you capture or create the data?

Describe how the data are ingested (IOOS Certification, f 2.)

Bottle samples are collected and analyzed at sea. Conductivity, temperature and depth (CTD) data are also collected at every station. Surface underway data is collected for the duration of the cruise.

Describe how data are managed (IOOS Certification, f 2.)

Each seawater property analysis produces data files which are then ingested into a single station csv file. CTD data are processed using SEASERV software. Underway surface measurements combined with fluorometry, pH, ship instrumentation (such as heading & speed) & meteorological data are combined into 30 second records and collected into daily files.

Describe the data quality control procedures that have been applied to the data. (IOOS Certification, f 3.)

For each seawater property analysis, certified standards are used to assure data quality. Once the hydrographic bottle data are combined into a single station csv, the data are then plotted and compared to the CTD data for further quality control. Once bottle data are finalized, the CTD data can be re-processed using the bottle data to make regressions. Underway data are converted into a universal ascii data format. Final underway data files contain both the processed ship specific RAW data and calculated values.

If you will be using existing data, state that fact and include where you got it.
What is the relationship between the data you are collecting and the existing data?

N/A

II. Expected schedule for data sharing

Adheres to the NOAA Data Sharing Procedural Directive. The System is an operational system; therefore the RICE should strive to provide as much data as possible, in real-time or near real-time, to support the operation of the System. (IOOS Certification, f. 4.)

How long will the original data collector/creator/principal investigator retain the right to use the data before opening it up to wider use?
If requested, raw data are available in real-time at sea.

How long do you expect to keep the data private before making it available? Explain if different data products will become available on different schedules (Ex: raw data vs processed data, observations vs models, etc.)
All raw data are available upon request. Finalized date are available within six months of each quarterly cruise.

Explain details of any embargo periods for political/commercial/patent reasons? When will you make the data available?
N/A

III. Standards for format and content

Which file formats will you use for your data, and why?

How can the information be accessed? (IOOS Certification, f 1. ii)

Data can be accessed through the CalCOFI website in an access database and also through NOAA's ERDDAP data portal.

What file formats will be used for data sharing?

1. ACCESS DATABASE

Microsoft Access format; version 1July2015 (mirror link). Includes final hydrographic data from March 1949 -July 2014 (IEH data + published Data Report data).

a. CSV

- CSV format; version 1July2015 (mirror link). Includes final hydrographic data from March 1949 -July 2014 taken from the Cast, Bottle and DIC Tables of the Access Database.

b. XML

- XML format; version 1July2015 (mirror link). Includes final hydrographic data from March 1949 -July 2014 taken from the Cast, Bottle and DIC Tables of the Access Database.as separate XML-XSD files.

c. SQL

- SQL format; version 1July2015 (mirror link). Includes final hydrographic data from March 1949 -July 2014 as Access to MySQL 'dump' SQL file. Complete web version, includes Cast, Bottle, Standard Station, Standard Depth, and metadata tables, uncompresses to over 870mb

2. ERDDAP

ERDDAP is a data server that gives you a simple, consistent way to download subsets of gridded and tabular scientific datasets in common file formats and make graphs and maps. This particular ERDDAP installation has oceanographic data (for example, data from satellites and buoys).

a. Tabledap

- Tabledap lets you use the OPeNDAP constraint/selection protocol to request data subsets,
  graphs, and maps from tabular datasets (for example, buoy data).
  For details, see ERDDAP's tabledap Documentation.
  (http://coastwatch.pfeg.noaa.gov/erddap/tabledap/index.html?page=1&itemsPerPage=1000)

b. Universal Ascii data files

- Underway data is available in raw, 30sec MET ascii format
  and 10 minute binavg, corrected data.

What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?

1. ACCESS DATABASE METADATA

a. Table 1.: http://www.calcofi.org/new.data/index.php/reporteddata/2013-09-30-23-23-27/cast-table-column-descriptions

b. Table 2.: http://www.calcofi.org/new.data/index.php/reporteddata/2013-09-30-23-23-27/database-tables-description

2. ERDDAP DATABASE METADATA

a. FGDC

b. ISO 19115-2/19139

c. .xml

What contextual details (metadata) are needed to make the data you capture or collect meaningful?

Contextual details describing attributes within the downloadable table are essential to understanding these data in detail.

How will you create or capture these details?

These details are capture within the metadata documentation.

What form will the metadata describing/documenting your data take?

ERDDAP has Web Accessible Folders (WAF) with FGDC and ISO 19115-2/19139 .xml metadata files for all of the geospatial datasets.

Which metadata standards will you use and why have you chosen them? (e.g. accepted domain-local standards, widespread usage)

FGDC and ISO 19115 metadata are both accepted standards and mandated by the US Federal Government.

2. **Polices for stewardship and preservation**

http://www.calcofi.org/data/underway/151-underway-data-format/261-underway-data-format.html

What is the long-term strategy for maintaining, curating and archiving the data?

Points of contact- Individuals responsible for the data management and coordination across the region (CV's attached); (IOOS Certification f 1. i)

James Wilkinson- Employee, Programmer Analyst
(858) 822-0674
jawilkinson@ucsd.edu

Jennifer Rodgers-Wolgast- Employee 12 Years, Staff Research Associate
(858) 534-2888
jrwolgast@ucsd.edu

Identify the procedures used to evaluate the capability of the individual (s) identified in subsection 997.23(f)(1) to conduct the assigned duties responsibly. (IOOS Certification, f 1. iii)

The University of California has a process in place for personnel evaluation. These evaluations are on file with UC San Diego Human Resources. All personnel listed have received excellent evaluations.

Which archive/repository/database have you identified as a place to deposit data?

Documents of the RICE's data archiving process or describes how the RICE intends to archive data at the national archive center (e.g., NODC, NGDC, NCDC) in a manner that follows guidelines outlined by that center. Documentation shall be in the form of a Submission Agreement, Submission Information Form (SIF) or other, similar data producer-archive agreement (IOOS Certification, f 6.).

Data are archived and stored through a local hard drive, backup hard drive, webserver and mirror web serve. Also, data are sent to NOAA for their ERDDAP Serve:
http://coastwatch.pfeg.noaa.gov/erddap/tabledap/index.html?page=1&itemsPerPage=1000

What procedures does your intended long-term data storage facility have in place for preservation and backup?

Addressing long-term data storage, CalCOFI plans on maintaining current infrastructure and continuing to work with national data archives such as ERDDAP.

How long will/should data be kept beyond the life of the project?

Data will be stored indefinitely.

What data will be preserved for the long-term?

All data will be preserved.

What transformations will be necessary to prepare data for preservation / data sharing?
Raw data are analyzed and quality controlled.

What metadata/ documentation will be submitted alongside the data or created on deposit/ transformation in order to make the data reusable?
Descriptive tables are available online with the intention of being used as metadata, also, ERDDAP metadata documentation is available.

What related information will be deposited?
Hydrographic Data Reports
We report the data that we collect on our cruises in a Hydrographic Data Report. These reports consist of:

INTRODUCTIONS
Contains the standard procedures, tabulated data descriptions, associated ancillary projects, footnote definitions, and literature cited.
CRUISE DATA REPORTS
Quarterly tabulated data reports. The 60+ year dataset includes temperature, salinity, oxygen and phosphate observations. In 1961, nutrient analysis expanded to include silicate, nitrate and nitrite; in 1973, chlorophyll was added; in 2008, the nutrient analysis expanded to include ammonium. Data Report Format.
PRIMARY PRODUCTIVITY DATA REPORT
These are a subset of the quarterly tabulated cruise data report containing only stations in which primary productivity experiments were carried out. Includes hydrographic data along with primary productivity data.
PERSONNEL
List of personnel per cruise, cruise dates, ports of call, ship name, and captain's name.
MACROZOOPLANKTON BIOMASS REPORT
Contains the macrozooplankton biomass, as wet displacement volume, after removal of large (>5 ml) organisms from the starboard side of a 71 cm mouth diameter paired net (bongo net) equipped with 0.505mm plankton mesh; towed obliquely from 210 meters to the surface.
SPATIAL PATTERN FIGURES
Contains vertical (line 90 only) and horizontal (10 and 200 meters along the standard CalCOFI grid) contour plotted hydrographic data.
BIRD FIGURES
Quartely Avifauna distribution observations over the standard CalCOFI grid.

3. Procedures for providing access

What are your plans for providing access to your data? (on your website, available via ftp download, via e-mail, or another way)
Describe how data are distributed including a description of the flow of data through the RICE data assembly center from the source to the public dissemination/access mechanism. (IOOS Certification, f. 2.)
Online database files and ERDDAP Server:
http://coastwatch.pfeg.noaa.gov/erddap/info/siocalcofiHydroBottle/index.html

Will any permission restrictions need to be placed on the data?
No, these data are publically available.

With whom will you share the data, and under what conditions?
Data are publicly available.

Will a data sharing agreement be required?
In general, a data sharing agreement will not be required. However, data should be properly acknowledged.

Are there ethical and privacy issues? If so, how will these be resolved?
N/A

Who will hold the intellectual property rights to the data and how might this affect data access?
These reports are not copyrighted, except where otherwise indicated, and may be reproduced in other publications provided credit is given to California Cooperative Oceanic Fisheries Investigations and to the author(s).

4. **Previous published data**

a. http://www.calcofi.org/new.data/index.php/publications/2014-02-18-22-26-15/2014-2015-publications

d. http://www.calcofi.org/new.data/index.php/publications/2014-02-18-22-26-15/2013-2014-publications

e. http://www.calcofi.org/new.data/index.php/publications/2014-02-18-22-26-15/2012-2013-publications

f. http://www.calcofi.org/new.data/index.php/publications/2014-02-18-22-26-15/2011-2012-publications