Data and Metadata Standards

About this Document

This Data and Metadata Standards document was prepared by the Ocean Science Trust, building on region-specific standards developed previously as part of the regional marine protected area (MPA) baseline monitoring programs. The North Central Coast MPA Baseline Program Data and Metadata Standards were prepared by Rani Gaddam (University of California, Santa Cruz), in consultation with the academic, agency, and independent scientists that conducted MPA monitoring as part of the North Central Coast MPA Baseline Program. The South Coast MPA Baseline Program Data and Metadata Standards built on the North Central Coast standards and were prepared by the Ocean Science Trust in collaboration with Carol Blanchette (University of California, Santa Barbara) and Jennifer Casselle (University of California, Santa Barbara). The Indigenous Traditional Knowledge data and metadata standards were prepared by the North Coast MPA Baseline Program Indigenous Traditional Knowledge project collaborators, coordinated by Megan Rocha (Tolowa Dee-ni' Nation).

Data files and associated documentation, including metadata and technical reports are submitted to the California Department of Fish and Wildlife, California Ocean Protection Council, and California Ocean Science Trust at or before completion of state-funded MPA monitoring projects. To advance data coordination, integration, and utility, this document provides data and metadata standards designed to guide data submission by program participants. The recommendations and standards included in this document will continue to be refined and updated, as needed.

About the California Ocean Science Trust

Ocean Science Trust is an independent, non-profit organization that brings together governments, scientists, and citizens to build trust and understanding in ocean and coastal science and management. We empower participation in the decisions that are shaping the future of California's ocean. We were established by the California Ocean Resources Stewardship Act (2000) to support managers and policymakers with sound science. With our state partners, we develop and implement marine protected area (MPA) monitoring to assess the effectiveness of California's statewide network of MPAs. For more information, visit our website at www.oceansciencetrust.org.

SUMMARY

Approach

California's statewide network of marine protected areas (MPAs) was completed on December 19, 2012. Supported by an investment of state funds from the California Ocean Protection Council, regional baseline monitoring was conducted to assess ecological and socioeconomic conditions at or near the time of MPA implementation, and to detect initial changes occurring following MPA implementation. Data and results from baseline monitoring are shared publicly through OceanSpaces.org.

Data and metadata standards have been developed for California's Statewide MPA Monitoring Program, including the regional MPA baseline program, to supplement data coordination and integration, facilitate data submission to the Ocean Science Trust, and increase the consistency and utility of MPA monitoring data. These standards build on data and metadata standards developed for the North Central Coast MPA Baseline Program in 2011, the South Coast MPA Baseline Program in 2015, and for Indigenous Traditional Knowledge in 2016.

Data & Metadata Standards

This document includes required and recommended data and metadata standards. *Required* Data/Metadata Standards must be adhered to by all monitoring program participants when submitting data and metadata, while *Recommended* Data/Metadata Standards should be adhered to when possible, but are not required.

Ecological Metadata Language (EML, developed by the Knowledge Network for Biocomplexity, KNB) is adopted as the metadata standard for data collected as part of the Statewide MPA Monitoring Program. EML offers a number of benefits for use as the standard for MPA monitoring data: it is flexible, extensible, and machine readable, it has been the standard for the regional MPA baseline programs, and it has already been adopted by a number of organizations and institutions throughout California. Required metadata standards focus on project- and file-level metadata fields and descriptions. These metadata files may be augmented by more detailed metadata adhering to EML, or International Organization for Standardization (ISO) standard 191**, which is support by the Federal Geographic Data Committee (FGDC) as the geospatial metadata standards.

Next Steps

These data and metadata standards should be considered a "living document" that will be updated and refined as data collection and submission continues. Together with partners in the region, Ocean Science Trust will review and adapt these standards as needed to apply to the breadth of data collection and analysis. Beyond facilitating integrated scientific analyses, these standards serve as the foundation for instructions to guide data and metadata submission, or upload, via a web interface.

TABLE OF CONTENTS

INTRODUCTION **5**

DEVELOPMENT OF THESE DATA AND METADATA STANDARDS 5

USING THESE DATA AND METADATA STANDARDS 6

DATA PACKAGES **7**

DATA STANDARDS **10**

INDIGENOUS TRADITIONAL KNOWLEDGE INFORMATION 10

STANDARD UNITS AND FORMATS 11

TABULAR DATA 11

Organism-Specific Observation Data Tables *12*

Physical Observation Data Tables *12*

Derived Data *12*

LOOKUP TABLES 13

MPA Boundary Information Lookup Table *13*

MPA Baseline Program Information Lookup Table *13*

Site/Location Information Lookup Table *13*

Organism Information Lookup Table *14*

Sampler/Interviewer Information Lookup Table *15*

NON-TABULAR DATA 16

GeoSPATIAL Data *16*

Interview Transcripts *16*

Survey data *16*

Databases *16*

Products from archival research *17*

MEDIA ASSETS 17

Imagery Files *17*

Videography Files *18*

Audio Files *18*

METADATA STANDARDS **18**

BASIC PROJECT INFORMATION METADATA 18

METHODS AND SAMPLING METADATA 19

DATA AND LOOKUP TABLE METADATA 21

General Data/Lookup Table Information *21*

Data/Lookup Table Attribute Information *21*

NON-TABULAR DATA 23

GeoSPATIAL data *23*

Interview Transcripts *23*

Survey data *24*

Databases *24*

Products from archival research *25*

MEDIA ASSETS METADATA 25

Underwater Imagery metadata *26*

Above-water imagery *27*

INTRODUCTION

In December 2012, a network of 124 marine protected areas (MPAs) and 15 special closures was completed along the California coast under the California Marine Life Protection Act (1999). MPA monitoring, which is a key component of California's MPA Management Program, is guided by a monitoring framework, which was adopted by the Fish and Game Commission in 2010. Monitoring in California is implemented in two phases: (1) regional, baseline monitoring, and (2) statewide long-term monitoring.

Key ecosystems and human uses were monitored as part of regional baseline monitoring, which had two key goals:

1. Baseline Characterization: A summary description, assessment and understanding of ecological and socioeconomic conditions in the South Coast region, inside and outside MPAs established under the MLPA, at or near the time of their implementation. Baseline characterization provides a frame of reference to support subsequent assessment of MPA network performance against MLPA goals and facilitate future adaptive management.
2. Assessment of Initial Ecological and Socioeconomic Changes - Measurement of initial ecological changes and the short-run net benefits or costs to consumptive and non-consumptive user groups following MPA implementation.

MPAs were implemented regionally, across four coastal regions in California. Baseline monitoring followed this regional approach, and was launched regionally, following implementation of each of the four regionally MPA networks. Data collected and generated as part of MPA baseline monitoring is submitted to the Ocean Science Trust and made publicly-available on OceanSpaces.org. The data and metadata standards presented in this document provide guidance to MPA monitoring project leaders and collaborators as they prepare their data, and will help to ensure that those data can be used by others in the future. These Standards are designed to encompass the wide variety of data that will be submitted by MPA monitoring projects in California.

DEVELOPMENT OF THESE DATA AND METADATA STANDARDS

The data and metadata standards in this document were prepared by the Ocean Science Trust, building on region-specific standards developed previously as part of the regional MPA baseline monitoring programs. The North Central Coast MPA Baseline Program Data and Metadata Standards were prepared by Rani Gaddam (University of California, Santa Cruz), in consultation with the academic, agency, and independent scientists that conducted MPA monitoring as part of the North Central Coast MPA Baseline Program. The South Coast MPA Baseline Program Data and Metadata Standards built on the North Central Coast standards and were prepared by the Ocean Science Trust in collaboration with Carol Blanchette (University of California, Santa Barbara) and Jennifer Casselle (University of California, Santa Barbara). The Indigenous Traditional Knowledge data and metadata standards were prepared by the

North Coast MPA Baseline Program Indigenous Traditional Knowledge project collaborators, coordinated by Megan Van Pelt (Tolowa Dee-ni' Nation).

Because of the variety of data types and formats to be submitted, the Ecological Metadata Language (EML) metadata specification, developed by The Knowledge Network for Biocomplexity (KNB), was identified as offering the most flexibility. EML offers compatibility with the International Organization for Standardization (ISO) standard 191**, which is support by the Federal Geographic Data Committee (FGDC) as the geospatial metadata standards. Some other key benefits of using EML are as follows:

- EML allows for both basic and highly detailed metadata. Some standards only accept high detail; others do not provide the level of detail possible with EML.
- The highly structured format of EML allows for metadata to be machine processed.
- Free software tools such as Morpho (https://knb.ecoinformatics.org/ - tools/morpho) are available that allow users to create, edit, and store EML.
- EML has already been adopted, or is in the process of being implemented, by Statewide MPA Monitoring Program researchers across the state.

USING THESE DATA AND METADATA STANDARDS

This document is intended to define the data and metadata standards to which MPA monitoring project leaders and collaborators will adhere when submitting data to the California Ocean Science Trust. This document is not intended to provide a complete definition and description of metadata, instructions for creating and editing EML, or specific instructions on data submission. Please see the OceanSpaces Data Upload Tool (DUT) User's Guide for a detailed description of the step-by-step process for uploading data; a link to this document is provided within the DUT.

Required data and metadata standards should be adhered to by all Statewide MPA Monitoring Program participants, while *recommended* data and metadata standards should be adhered to when possible, but are not required. Data packages that contain sensitive data[1], including endangered species data and Indigenous Traditional Knowledge (ITK) information and assets, shall conform to these data and metadata standards, and shall refer to the Data Upload Tool User's Guide for information about how sensitive data will be handled. Note that in some situations, a required data and metadata standard may not apply to a specific data package (e.g., organism information for non-biological data). In some instances, only recommended data and metadata standards are listed.

[1] Some data packages may contain information on the location of Endangered Species Act-listed species (e.g., black abalone). Data marked as containing endangered species data in the data upload process will not be made public on OceanSpaces. Any project with endangered species data must also upload a separate data package that is identical to each sensitive data package, but with the sensitive data removed from all files.

DATA PACKAGES

Data and/or derived data should be submitted, along with corresponding metadata, in the form of one or more data packages. A *data package* represents a collection of metadata, data files, and lookup tables that together form a complete description of a particular survey type (i.e., data that can be described by a single project information metadata file and a single methods and sampling metadata file). For example, data from surveys of species density and percent cover collected at the same site could represent two different data packages. Organism information (i.e., taxonomy), site/location information, and sampler information should be included as lookup tables in each data package.

Below is an outline of the contents that should be included in each data package, all of which are described in this document. Within each data package, each data file and lookup table should have a corresponding metadata file. Each data package will include a methods and sampling metadata file that encompass all of the data in the package.

Example data package files for ecological monitoring projects:

1. Tabular data and associated metadata
    a. Organism-specific observations
    b. Physical observations
    c. Derived data
2. Lookup tables and associated metadata
    a. MPA boundary information (*automatically appended*)
    b. MPA baseline program information (*automatically appended, if applicable*)
    c. Site/Location information
    d. Organism information
    e. Sampler information
3. Non-tabular data and associated metadata
    a. Geospatial data
    b. Databases
4. Media assets (e.g., imagery, videography, audio)
    a. Imagery and associated metadata
    b. Videography metadata[2]

[2] Videography produced through a state-funded monitoring projects must be made available to the California Department of Fish and Wildlife, California Ocean Science Trust, and California Ocean Protection Council. If videography is governed by a data use agreement (e.g., non-disclosure agreement, memorandum of understanding, human subjects research protocol, guidelines for use of Indigenous Traditional Knowledge), then access may not be required. Contact Ocean Science Trust to discuss access to and/or delivery of videography assets.

    c. Audio metadata[3]

5. Other information (*automatically appended, if applicable*)
    a. Data Use and Attribution statement

Example data package files for human dimensions projects

1. Tabular Data
    a. Derived data
2. Lookup Tables
    a. MPA boundary information (*automatically appended*)
    b. MPA baseline program information (*automatically appended, if applicable*)
    c. Site/Location information
    d. Organism information (*if applicable*)
    e. Sampler/Interviewer information
3. Non-tabular Data
    a. Geospatial data
    b. Interview transcripts
    c. Survey data
    d. Databases
4. Media assets (e.g., imagery, videography, audio)
    a. Imagery and associated metadata
    b. Videography metadata[1]
    c. Audio metadata[2]
5. Other Information (automatically appended, if applicable)
    a. Data Use and Attribution statement

[3] Audio produced through a state-funded monitoring project must be made available to the California Department of Fish and Wildlife, California Ocean Science Trust, and California Ocean Protection Council. If audio is governed by a data use agreement (e.g., non-disclosure agreement, memorandum of understanding, human subjects research protocol, guidelines for use of Indigenous Traditional Knowledge), then access may not be required. Contact Ocean Science Trust to discuss access to and/or delivery of audio.

Example data package files for ITK projects

1. Tabular Data
    a. Derived data
2. Lookup Tables
    a. MPA boundary information (*automatically appended*)
    b. MPA baseline program information (*automatically appended, if applicable*)
    c. Site/Location information
    d. Organism information (*if applicable*)
    e. Sampler/Interviewer information
3. Non-tabular Data
    a. Geospatial data
    b. Interview transcripts
    c. Survey data
    d. Databases
    e. Products from archival research
4. Media assets (e.g., imagery, videography, audio)
    a. Imagery and associated metadata
    b. Videography metadata[1]
    c. Audio metadata[2]
5. Other Information (automatically appended, if applicable)
    a. Data Use and Attribution statement
    b. Considerations and Guidelines for Accessing and Applying Indigenous Traditional Knowledge under the North Coast MPA Baseline Monitoring.[4]

[4] While data packages containing ITK will contain the same basic components as other examples provided here, they will also include a document that provides additional information and guidelines on accessing and using ITK data, while protecting and appropriately contextualizing ITK and its confidentiality. This document will be automatically appended to appropriate data packages.

DATA STANDARDS

The following Data Standards have been developed for all data files created and/or compiles as part of the MPA monitoring. Standards are presented in separate subsections below for tabular data, lookup tables, non-tabular data, media assets, and other information. Required Data Standards should be adhered to by all projects, and Recommended Data Standards should be adhered to whenever possible.

Please note that each data file should be accompanied by a metadata file that provides in-depth information about its contents. Metadata Standards for metadata files are presented in the next section of this document. Note that some data standards in this section briefly mention items that should be included in the associated metadata files; please refer to the metadata standards section below for a full description of these items.

INDIGENOUS TRADITIONAL KNOWLEDGE INFORMATION

The State of California, in its work to understand current and historical marine conditions, has made important strides in recognizing the role of Indigenous Traditional Knowledge (ITK) with regard to the function and stewardship of marine ecosystems. The North Coast regional MPA monitoring program was the first in the State to incorporate ITK. Because ITK encompasses complex knowledge and belief systems, it is important to understand that ITK cannot be distilled down to discrete pieces of “data”. Rather, ITK is an entire worldview that incorporates knowledge, teachings, beliefs, and practices that operate in iterative and holistic ways of life that have emerged over time and across generations since time immemorial. This knowledge is not solely comprised of ecological observations, but also incorporates elements of spirituality, ceremony and belief that play an integral role to understanding and interacting with the natural environment, as well as provide for the laws and protocols for stewarding these resources and places. Additionally, understanding, meaning and value are rooted in place and cannot be extrapolated from Tribal peoples’ relationship to, and connection with, those places. ITK can provide a contextualized and qualitative understanding of place and interaction between human and other parts of the environment over time, and often cannot be quantified.

Tribes have a responsibility to protect individual and collective cultural heritage and ITK from appropriation and misuse. Tribal research includes information about specific places, cultural beliefs and practices, and other information that is sensitive and holds high cultural value to the Tribes. Often these practices and beliefs convey important information to Tribes that is culturally sensitive, and thus are not made available to the public. Additionally, locations of certain Native American cultural places, as well as sensitive information about their nature and uses, are confidential and legally protected from public disclosure under various State and Federal laws, including the Freedom of Information Act (5 U.S.C. § 552). Protecting confidential information is an issue of utmost importance to the Tribes involved in this project, and to Indigenous Peoples worldwide, and is widely recognized in government-to-government consultation protocols and guidelines.

These standards are provided throughout this document. They have been developed to support the utilization of ITK data[5], while protecting and appropriately contextualizing the data. See Appendix A for additional information about Indigenous Traditional Knowledge information and assets.

STANDARD UNITS AND FORMATS

Required Data Standards

- All units used should adhere to the standard units accepted by EML, including the use of metric units where possible. See https://knb.ecoinformatics.org/ - external//emlparser/docs/eml-2.1.1/./eml-unitTypeDefinitions.html for more information.
- All date fields, (i.e., year, month, and date), should be in separate columns, in the following format: year (YYYY, e.g., 2011), month (MM, e.g., 06, 10) and day (DD, e.g., 02, 28).
- All time fields (i.e., hours, minutes, and seconds) should be in a single column in HH:MM:SS format.
- Temperature data should be in degrees Celsius.
- For all time fields, there should be a column indicating the time zone used (e.g., Pacific Standard Time, Greenwich Mean Time, etc.); this should also be included in the metadata.

Data packages containing ITK data will be required to adhere to the same basic standard units and formats, with additional standards that support the access and use of ITK data, while protecting and appropriately contextualizing the data and confidentiality.

TABULAR DATA

Tabular data generated as part of the Statewide MPA Monitoring Program are stored in data tables, each with a corresponding metadata file. These data tables can fall into the following categories:

1. Organism-specific observations
2. Physical observations
3. Survey data
4. Derived data

Required and recommended data standards are presented below for each category of data table.

[5] "Data" refers to knowledge, information, and assets gathered.

ORGANISM-SPECIFIC OBSERVATION DATA TABLES

Required Data Standards

When organism-specific observation data are included in a data package, those data should be submitted in tabular format, with the project short code provided in a separate column. If data are collected in a plot or along a transect, etc., a plot identifier (e.g., plot_id) and information about the precision of the location data should be included in the data table.

All data tables that present measurement data (i.e., non-count data) should indicate the unit of measurement. Count data should be defined as per a given area, time, etc., and the standard units should be provided. For organism/species-specific data tables that are derived from ITK data collection, other qualitative codes/categories may be used to characterize species abundance/count.

The corresponding metadata file should include a description of the unit of measurement, clear definitions of the type of data collected (e.g., counts, percent cover, sizes, etc.), and sample size. If qualitative codes are used in the data table (e.g., rare/ present/ common/ abundant, small/ medium/ large), they should also be clearly defined in the metadata file.

PHYSICAL OBSERVATION DATA TABLES

Required Data Standards

When physical data measurements are included in a data package, those data should be submitted in tabular format, with the project short code provided in a separate column. Data tables that present measurement data (e.g., temperature, wind speed, swell, visibility, depth, etc.) should indicate the unit of measurement. If data are collected in a plot or along a transect, etc., a plot identifier (e.g., plot_id) and information about the precision of the location data should be included in the data table.

The corresponding metadata file should include a description of the unit of measurement, and clear definitions of any qualitative codes used, (e.g., low/ medium/ high).

DERIVED DATA

Required Data Standards

The tables used to generate derived data (e.g., graphs, analysis, etc.) should be submitted in tabular format, following the data standards for the type of data being analyzed.

Each corresponding metadata file should include a description of the unit(s) of measurement, a description of the methods used to generate the derived data (e.g., how averages/sums/means were calculated, including script files if applicable; how data were aggregated to protect anonymity; etc.), and

information regarding any reference standards used (e.g., chlorophyll standards, etc.) for any calculations.

For physical data tables that contain ITK data, other qualitative codes/categories may be used to characterize physical attributes.

LOOKUP TABLES

MPA BOUNDARY INFORMATION LOOKUP TABLE

Required Data Standards (*automatically appended*)

Information about all corresponding MPA boundaries (and associated metadata) will be provided by the Ocean Science Trust, and these files will be automatically appended to each data package by the Data Upload Tool. The information file will include, in tabular format, the name, designation, region, and boundary global positioning system (GPS) coordinates for each MPA. The headers and values in this table will be consistent across all projects within a given regional baseline program.

MPA BASELINE PROGRAM INFORMATION LOOKUP TABLE

Required Data Standards (*automatically appended*)

Information about the projects within the corresponding regional MPA baseline program (and associated metadata) will be provided by the Ocean Science Trust, and this file will be automatically appended to each data package by the Data Upload Tool. The information file will include all project short codes, project titles, project descriptions, and other information about each project.

SITE/LOCATION INFORMATION LOOKUP TABLE

Required Data Standards

Each data package should include a site/location information file with the site (e.g., research sites) and/or location (e.g., buoy station) information that indicates where the data were collected. For human dimensions and Indigenous Traditional Knowledge projects, this file should include the location that best describes where the data were collected or the location associated with the data (e.g., fishing port or port group, cultural site). This information should be submitted in tabular format, and should include, in separate columns:

- *Project short code*
- *GPS coordinates, in decimal degrees (NAD83 or WGS84 coordinate system)*: For fixed GPS coordinates (i.e., one single point per site), a single set of GPS coordinates is sufficient. However, for survey areas (e.g., fishing footprints, harvesting areas), complete boundary

coordinates should be included so that these data can be mapped in relation to MPA boundaries. However, for survey areas, including fishing footprints, and areas encompassing ITK data collection[6], complete boundary coordinates should be included, so that these data can be mapped in relation to MPA boundaries.

- *MPA name for sites located within MPAs*: Use the MPA name field from the *MPA boundary Information* table as a guide.

For complex areas and/or transects, geographic information system (GIS) shapefiles should be submitted in addition to the table described above. All shapefile submissions should follow the data and metadata standards for geospatial data detailed below.

Recommended Data Standards

Additional site information would enhance comparisons across datasets. Below are some examples of additional site information that could be included in the site/location information lookup table:

- *County*, *State*, *Country*
- *Island name*
- *Other designation* (e.g., National Marine Sanctuary name, Areas of Special Biological Significance name)

Additional location/site information valuable for re-sampling efforts could also be included (e.g., access/trailheads, directions, contacts, permits, etc.).

ORGANISM INFORMATION LOOKUP TABLE

Required Data Standards

All projects that collect organism-specific observation data should include taxonomic information about those organisms in a separate lookup table. This information should be submitted in tabular format, and should include, in separate columns:

- *Project short code*
- *Taxonomy for each organism, as defined on the Integrated Taxonomic Information System (ITIS) website (http://www.itis.gov/) or the World Registry of Marine Species (WoRMS) (http://www.marinespecies.org/index.php)*: When available, the following taxonomic groupings should be provided, each in a separate column: Kingdom, Phylum/Division, Class, Order, Family, Genus, and Species. For a variety of reasons, different projects may use multiple species names

[6] It is the prerogative of the individual tribe to determine the level of spatial resolution represented in geospatial data, to protect and manage the privacy of areas and sites referenced in ITK data/survey collection.

for the same organism (e.g., name changes over time, data entered before/after name change, different protocol for addressing name changes). Because of this, including as much of the taxonomy for each organism as possible will allow data users to compare data across projects and to identify synonymous species due to name changes (e.g., *Asterina* (*Patiria*) *miniata*).

- *"Name" of the organism, as it is called by the survey team*: In many cases, this "name" is defined by the taxonomy for the organism. However, in some cases, species are grouped or "lumped," and the "name" field cannot be easily defined by taxonomy. In these situations, a "name" and accompanying lumping description field should be included in the classification table so that the type of organism can be determined. The lumping description field should include the types of species that are lumped (e.g., the lumping description field for a group named "filamentous red" could read, "includes algal species of *Bangia, Callithamnion, Polysiphonia*"). The lumping description field should also note if a species is always lumped, or if only juveniles are lumped, etc. This information should also be included in the metadata.

Recommended Data Standards

Common names used for specific organisms could be included in species information lookup tables. However, common names are highly variable, and the same common name and/or organism grouping category may mean very different things in different datasets. For ITK data packages, the Tribal/indigenous name for a particular species may also be included. Therefore, metadata associated with the common name field should indicate any additional information that would enable data users to interpret the common name.

SAMPLER/INTERVIEWER INFORMATION LOOKUP TABLE

Recommended Data Standards

If names of the persons associated with a specific data record are included in the data files, sampler or interviewer information should be included in the associated data package. Where and how this information is submitted is left to the discretion of project leads, but it could be submitted within the associated data table, or in tabular format in a separate sampler information lookup table. This information should include the definitions of any sampler codes used in the data files, (e.g., full names for samplers if initials were used in the data). Additional information included on the sampler information table may include but is not limited to role (e.g., recorder, sampler, interviewer), project short code, training level (if applicable), affiliation (e.g., university, research group, tribe, agency), etc.

NON-TABULAR DATA

GEOSPATIAL DATA

Required Data Standards

Even if additional coordinate systems are used and preferred, geospatial data should also be projected in NAD83 or WGS84. All coordinate systems used should adhere to the standard types accepted by EML. See https://knb.ecoinformatics.org/ - external//emlparser/docs/eml-2.1.1/index.html for more information.

For fixed GPS coordinates (one single point per site, for example), a single set of GPS coordinates is sufficient. For survey areas, including fishing footprints and areas referenced in ITK data collection, boundary coordinates should be included.

Geospatial data should include a separate attribute table from the GIS file, with an accompanying metadata file. For prepared map products, files should be saved and uploaded as part of the data package in one of the following formats: pdf, png, jpg, gif, xml, zip, tiff. Coordinate system information should be provided in a separate metadata file, as detailed in the Metadata Standards for Geospatial Information below.

INTERVIEW TRANSCRIPTS

Recommended Data Standards

All interview transcripts, including questions and responses, should include an informative and simple filename, and should be in non-proprietary file formats. Files should also have necessary and documented informed consent from interviewee(s). Common file formats include: csv, txt, pdf.

SURVEY DATA

Recommended Data Standards*:*

Survey data, including individual responses and summarized data from survey questions, should be provided in in tabular format.

DATABASES

Recommended Data Standards

All files in a database should include an informative and simple filename. Files and data should follow existing data requirements for that specific type of file/data type. Refer to existing standards within this document.

PRODUCTS FROM ARCHIVAL RESEARCH

Recommended Data Standards

A wide variety of products could be produced or derived through archival research, such tabular data, media assets, copies of archival documents, and bibliographies of archival documents. Tabular data and media assets should adhere to the standards described above and below, respectively. All products from archival research should include a simple, informative filename.

Copies of archival documents should be provided in a standard, non-proprietary file format, such as pdf, png, jpg, gif, xml, zip, and tiff.

Bibliographies of archival documents should also be provided in a standard, non-proprietary file format, such as pdf, xml, or zip. The citation format should be consistent throughout the document.

MEDIA ASSETS

Media assets include imagery, videography, and audio files. These assets require processing to yield data that could be submitted in tabular format. While monitoring projects will process some of the media produced, most often researchers do not process 100% of the imagery, footage, or recordings. While associated data may not be available for all media assets, metadata should be generated and included within a data package to ensure future researchers can discover the unprocessed media assets and process them to produce data. See Media Asset Metadata section below.

During the MPA baseline monitoring phase, videography and audio files are archived by those who generated them. However, information on how to access these media assets must be included in the metadata files. For long-term monitoring, submission and archiving requirements for media assets may change. If your project is part of the long-term monitoring program, please contact Ocean Science Trust to inquire about media asset submission and archiving requirements.

IMAGERY FILES

Required Data Standards

Imagery files not subject to data use agreements should be saved and uploaded as part of the data package in one of the following formats: pdf, png, jpg, gif, xml, zip, or tiff.

VIDEOGRAPHY FILES

Required Data Standards

Videography files must be stored by the project leader until further instruction is provided. For state-funded projects, access to videography files not subject to data use agreements must be granted upon request.

AUDIO FILES

Required Data Standards

Audio files must be stored by the project leader until further instruction is provided. For state-funded projects, access to audio files not subject to data use agreements must be granted upon request.

METADATA STANDARDS

Each data table, lookup table, and non-tabular data file in every data package should have a corresponding metadata file. That is, each data package will likely contain multiple metadata files.

The metadata files that correspond to ecological data should be submitted in EML format. Metadata files for the MPA boundary information lookup table and the MPA Baseline Program Information lookup table will be provided by California Ocean Science Trust and automatically appended to each data package.

BASIC PROJECT INFORMATION METADATA

Each data package will include a basic project Information metadata file, which will be automatically generated for each project from the information entered into the OceanSpaces Data Upload Tool. The Basic Project Information metadata file will include the following:

Project name is the name of the South Coast MPA Baseline Program project under which the data were collected, available in the MPA Baseline Program Information Lookup Table.

Brief data package name should provide a full description of the data package, and should include enough detail to differentiate it from other similar data packages submitted by the same project.

Temporal coverage indicates the range of time (complete calendar start and end date) covered in the associated data package. Note that these dates may vary among data packages.

Project/data package description should describe the overall project, and the specifics of the particular data package being described.

Creator (s) or dataset owner(s) for each data package are the person(s) that created the data (e.g., the Principal Investigator(s)). There may be more than one creator per data package. Full name, affiliation, mailing address, and email address are required for all data package creators.

Contact(s) for each data package are the person(s) to be contacted with questions about the use or interpretation of the data (may be the same as the owner, or in many cases will be the collector of the data). The same information required for the creator is required for the contact(s).

Associated party(ies) for each data package are the person(s) functionally associated with the data, such as the person who maintains the database (e.g., data manager). The same information required for the creator and contact is required for the associated party.

Ecosystem(s) and Human Uses (as applicable) for which data are provided within the data package

Keywords help to identify the dataset, and enable data packages to be discovered and categorized. Standard keywords are provided by Ocean Science Trust and included in the data upload tool to be used across all projects. Projects should also select additional keywords to help identify a data package.

Geographic coverage (i.e., coordinates of a bounding box for the entire area covered by the associated data package). Note that specific coordinates for each survey site, buoy station, and/or fishing port should be provided separately in the Site/Location Information Lookup Table as part of the data package.

In addition to the metadata that will be pulled from the Data Upload Tool, four files, generated by Ocean Science Trust, will automatically be appended to each data package:

- **Data Use & Attribution Statement**, which defines the usage rights and restrictions for data packages
- **Program Information lookup table** (if applicable)
- **Program Information lookup table metadata** (if applicable)
- **MPA Boundary lookup table** (if applicable)
- **MPA Boundary lookup table metadata** (if applicable)

METHODS AND SAMPLING METADATA

Each data package should include a separate Methods and Sampling metadata file. This file should be used to describe the methods, instrumentation, and/or protocols that apply to all data in the data package. If the methods consist of one or more steps, each step should have a unique title followed by a short description. For data gathered by means of community-based participatory research (e.g., surveys

and interviews), the methods metadata file should include the survey question(s) that were asked (verbatim) to solicit the data gathered.

This short summary will help data users to better understand the data.

Example Method Description:

Step 1: *Species Count Methods - Overview*

Species Count sampling is conducted by recording the number of a specific organism found in a permanent plot at a given site. For some species, size and color are also recorded. The type of plot and number of replicates monitored is dependent on the target species. Algae and invertebrate species counts are recorded. Please note that these data do not reflect all species present in a given area. In addition, the information in this Dataset is limited by the methods described in our survey protocols.

Step 2: *Species Count Methods - Motile Invertebrate Monitoring*

The number of select motile invertebrates are monitored within the photoplots at sites where the monitoring group has sufficient experienced samplers and time to conduct this survey. Sizes of the first ten individuals encountered in each plot are measured to the nearest mm for the following gastropod species: Acanthina *spp.,* Lottia gigantea*,* Nucella *spp.,* Ocenebra *spp.,* Chlorostoma *spp. Specific measured species will vary slightly among regions since only those that are abundant enough to get useful size data should be measured. Limpets less than 5 mm and from 5-15 mm in size are sub-sampled in three 20 x 20 cm quadrats, which are placed in upper left, middle, and lower right corner of each photoplot. Subsampled counts are facilitated by subdividing the 20 x 20 cm quadrat frames into four equal subsections with string. If limpets are extremely abundant (as commonly occurs with the 5 mm category), they are sub-sampled in a 10 x 10 cm section of the 20 x 20 cm quadrat. If no 5-15 mm limpets are counted in the 20 x 20 cm areas and these limpets are present in the plot, then the entire photoplot is counted. Sub-sampled limpet counts will be extrapolated to the full 50 x 75 cm photoplot area (counts in 20 x 20 cm areas are summed and multiplied by 3.125, counts in 10 x 10 cm areas are summed and multiplied by 12.5). Littorines are sub-sampled in a 10 x 10 cm section of the 20 x 20 cm sub-sampling quadrats. As with limpets, counts from subsampled areas will be extrapolated to the full 50 x 75 cm photoplot area. Specific information regarding the motile plot area and number of replicates can be found on the photo plot info table.*

DATA AND LOOKUP TABLE METADATA

Each tabular data file that is part of the data package (e.g., data files and lookup tables) should have an accompanying metadata file, which can be .xml (preferred), .txt, or .csv. These data and lookup table metadata files should include two core components:

1. General Data/Lookup Table Information (i.e., general information about the tabular data file); and
2. Data/Lookup Table Attribute Information (i.e., descriptions for each of the table column headings).

A detailed list of the fields that should be included in each Lookup and Data Table metadata file is provided below. Note that all of the following metadata standards are required:

GENERAL DATA/LOOKUP TABLE INFORMATION

This section of each data and lookup metadata file should include general information about the tabular data or lookup table, including:

- **Project short code:** Available in the Program Information lookup table (if applicable)
- **Table Name:** The actual file name of the data or lookup table

  Example: pisco_cbs_site_table.txt
- **Table File Format:** Note the format of your data file, such as text (.txt), comma separated (.csv), etc. Note that proprietary data formats (e.g., .xls) will not be accepted.

  Example: textFormat
- **Table Description:** A short paragraph describing the table, including information about the data it contains

  Example: Lookup table containing a list of sites and site characteristics for the Coastal Biodiversity Surveys

DATA/LOOKUP TABLE ATTRIBUTE INFORMATION

Each Lookup and Data Table metadata file should include a section with the following additional information about each attribute, or column heading:

- **Attribute Name:** The actual name of the column heading as it appears in the Lookup or Data Table
- **Attribute Label:** A more user-friendly column name, especially useful if the attribute name is difficult to interpret (e.g., spell out acronyms or abbreviations)

- **Attribute Definition:** A precise and descriptive definition of the attribute, used to clarify the data included in the column so data users can accurately understand what the attribute means or represents
- **Datatype:** The datatype for the attribute you are describing (e.g., string, integer, float, double, time, or date). For a full list and definitions of datatypes, please visit the World Wide Web Consortium (W3C) guide, *XML Schema Part 2: Datatypes Second Edition*, *Section 3.2: Primitive Datatypes*, available at http://www.w3.org/TR/xmlschema-2/.
- **Acceptable Data Range:** To allow for basic quality control procedures by future data users, include the acceptable data range for the data attribute that you are describing (e.g., acceptable range of lengths). For data attributes with 'text' datatype, the acceptable data range should be 'all text.'
- **Measurement Scale:** A classification that describes how the data were measured and how values are enumerated and defined. The measurement scale determines the types of statistics that can be used to analyze the data.

Definitions of each potential measurement scale are provided below. Note that each scale is a superset of the one beneath it, so the most restrictive category should be chosen that still accurately describes the data.

- **Nominal:** values are separated into unordered, named categories, such as gender (Male/Female). All text fields should be classified as nominal.
- **Ordinal:** values are separated into ordered categories with a set order. Ordinal data show a particular value's position relative to other values (e.g., low/medium/high, juvenile/sub-adult/adult). Each ordered category should be clearly defined.
- **Interval**: values are arranged on measurement scale of equal-sized units, but with no meaningful zero, such as the Celsius and Fahrenheit temperature scales. If used, units of measurement, precision, and number type (i.e., natural, real, or whole) should be included.
- **Ratio:** an interval scale that does have a meaningful zero point, such as elevation (from sea level), height, and the Kelvin temperature scale. If used, units of measurement, precision, and ratio type should be included.
- **Date_time:** used if the column contains data measured on a date-time scale; if used, format and precision information should also be included.

NON-TABULAR DATA

GEOSPATIAL DATA

Projects submitting geospatial data should provide metadata files that adhere to ISO standard 191**, which is support by the Federal Geographic Data Committee (FGDC) as the geospatial metadata standards), available at http://www.fgdc.gov/metadata/csdgm. A metadata file that adheres to these standards can be generated using ArcCatalog. These geographic metadata files should be submitted as part of the data submission, and should include a separate coordinate system information metadata file.

For reference, links to the complete EML specification for spatial data are available at the following link: https://knb.ecoinformatics.org/ - external//emlparser/docs/eml-2.1.1/index.html.

Note that projects that include geographic data will be required to provide Project Information metadata and Methods and Sampling metadata in addition to the FGDC CSDGM-compliant Data and Lookup Table metadata described above.

INTERVIEW TRANSCRIPTS

Recommended Data Standards

Metadata should be provided for all interview transcripts. These metadata should include the following information:

- **File Name:** The actual file name
- **Project short code** provided by Ocean Science Trust, available in the Monitoring Program information lookup table.
- **File Description:** A short paragraph describing the file, including information about the data it contains.
- **File Format:** A list of the type of file and the format for the file.
- **Date:** The calendar date on which the interview was conducted (in YYYY_MM_DD format).
- **Duration**: Length of interview in 24-hour (HH_MM_SS).
- **Description of interview scope:** This can include the purpose and intention of the interview.
- **Interviewee(s) name and affiliation:** The full name and affiliation of the interviewee.
- **Interviewer(s) name and affiliation:** The full name and affiliation for the person(s) responsible for conducting interviews.

SURVEY DATA

Recommended Metadata Standards

Metadata should be provided for all survey data. These metadata should include the following information:

- **File Name:** The actual file name
- **Project short code** provided by Ocean Science Trust, available in the Monitoring Program information lookup table.
- **File description:** A short paragraph describing the file, including information about the data it contains.
- **File format:** A list of the type of file and the format for the file.
- **Date:** The calendar date on which the interview was conducted (in YYYY_MM_DD format).
- **Description of survey scope:** This can include the purpose and intention of the survey.
- **Surveyor(s) name and affiliation:** The full name and affiliation for the person(s) responsible for distributing or otherwise administering the surveys.
- **Data processor(s) name and affiliation**: The full name and affiliation of the individual or entity who processed the survey data.
- **Survey questions:** The questions that were asked (verbatim) to solicit the data gathered.

DATABASES

Recommended Metadata Standards

Metadata should be provided for all databases produced as part of the work described in the data package. These metadata should include the following information:

- **File Name:** The actual file name
- **Project short code** provided by Ocean Science Trust, available in the Monitoring Program information lookup table.
- **File description:** A short paragraph describing the file, including information about the data it contains.
- **File format:** A list of the type of file and the format for the file.
- **Database description:** A short paragraph describing the database, including information about the data/files it contains.
- **Database creator/publisher:** Entity that created and/or maintains the database.

- **Catalog of files/data records:** A catalog of the files within the database, including the file type (audio, image, etc.) file format (jpg, csv, pdf, etc.)
- **Total number of files/data records:** This is the total number of all files and records contained in the database.
- **Data range:** The data range of data records (in YYYY_MM_DD format).

PRODUCTS FROM ARCHIVAL RESEARCH

Recommended Metadata Standards

Metadata should be provided for all products from archival research. Noted below after each item is the type of product from archival research for which the metadata standard is relevant.

- **File Name:** The actual file name
- **Project short code** provided by Ocean Science Trust, available in the Monitoring Program information lookup table. (ALL)
- **Product description:** A short paragraph describing the document/file/data being accessed. (*copy of archival product, bibliography*)
- **File format:** A list of the type of file and the format for the file. (*ALL*)
- **Donor source**: the primary entity that is hosting/holding/archiving the *original* files/objects/specimens. (*copy of archival product, bibliography*)
- **Donor source contact information**: Include name of donor source, web address or physical address. (*copy of archival product, bibliography*)
- **Creator(s) of the product:** The full name and affiliation (if applicable) for the creator of the product, such as the author(s), artist(s) (e.g., photographer, maker), interviewer(s) & interviewee(s), etc. (*ALL*)
- **Citation format:** The citation format used. (*bibliography*)
- **Date product was accessed:** The date the product was accessed. (*ALL*)

MEDIA ASSETS METADATA

All imagery, videography, and audio files should each be accompanied by a corresponding metadata file. Image and video metadata files should include the following information (as applicable) to aid in linking the image and/or video files to other tabular data. See Audubon Core Multimedia Resources Metadata schema for more information: http://terms.tdwg.org/wiki/Audubon_Core.

- **File Name:** The actual file name

- **Project short code:** Available in the MPA Baseline Program information lookup table
- **Date**: The calendar date on which the image or video was generated (in YYYY_MM_DD format)
- **Time:** The specific time at which the image or video was generated, to the second (in 24-hour HH_MM_SS format)
- **Site/Station/Location**: The location captured in the image or video. Use of location short codes, plot numbers, site name, MPA name, etc. that correspond to data and lookup tables is recommended. Media assets containing ITK data may withhold, or limit the disclosure of, location information to protect and manage the privacy of sites and areas referenced in ITK data collection.
- **Photographer/Videographer Name:** The full name of the photographer or videographer
- **File Description:** A short paragraph describing the file, including information about the data it contains

  Examples:
  Andrew Molera, Chthamalus *Photoplot 1, taken Fall 2001 (for an image)*
  Coastal Biodiversity Survey Protocols, July 2009 (for supporting documentation)
- **File Format:** A list of the type of file and the format for the file

  Examples:
  image/jpeg
  application/pdf
  application/excel
  application/mpeg

UNDERWATER IMAGERY METADATA

Video clip and still imagery metadata files must accompany each underwater video clip and still imagery data file submitted. These metadata should include the following information, in addition to the standards outlined at the top of this section:

- **Camera type:** The type of camera used to capture the video or still imagery.
- **Vehicle type/name:** The type of vehicle onto which the camera was affixed for the duration of the dive, e.g., ROV type/name, camera sled; required for ROV, submersible, and other underwater vehicle surveys
- **Dive number/name**: The dive number, name, or other identifying label for the specific dive on which the video clip or still image was captured.
- **Transect number/name:** The transect number, name, or other identifying label for the specific transect on which the video clip or still image was captured.
- **Camera angle:** The angle of the camera when the video clip or image was taken, e.g., forward or down

- **Photo number** (*if applicable*)

Processed navigation metadata files must be submitted with each video clip metadata file submitted. These files should include the following information, in addition to the standards outlined at the top of this section:

- **Dive number/name**: The dive number, name, or other identifying label for the specific dive on which the video clip or still image was captured.
- **Transect number/name:** The transect number, name, or other identifying label for the specific transect on which the video clip or still image was captured.
- **Transect start time:** The time the transect began in HH:MM:SS
- **Transect end time:** The time the transect ended in HH:MM:SS
- **Latitude:** The latitude, in decimal degrees, of the start and end points of the transect
- **Longitude:** The longitude, in decimal degrees, of the start and end points of the transect
- **Depth:** The depth of the transect, in meters, provided as often as the instruments on board the vehicle recorded it. At *minimum*, this should be the start and end of the transect.
- *Additional information as available, e.g., CTD string*

ABOVE-WATER IMAGERY

Video clip and still imagery metadata files must accompany each above-water video clip and still imagery data file submitted. These metadata should include the following information, in addition to the standards outlined at the top of this section:

- **Camera type:** The type of camera used to capture the video or still imagery.
- **Transect, quadrat, or other identifying number/name:** e.g., permanent bolt reference number
- **Camera angle:** The angle of the camera when the video clip or image was taken, e.g., forward or down
- **Photo number** (*if applicable*)