OCEAN EXPLORATION COOPERATIVE INSTITUTE | EXPLORING THE NATION'S BLUE FRONTIER
OCEAN EXPLORATION TRUST | UNIVERSITY OF NEW HAMPSHIRE | UNIVERSITY OF RHODE ISLAND
UNIVERSITY OF SOUTHERN MISSISSIPPI | WOODS HOLE OCEANOGRAPHIC INSTITUTION

Ocean Exploration Cooperative Institute

Data Management Plan

DRAFT 1.0.0

Drafted by the OECI Data Management Working Group Members:

Brian Calder (Chair)	UNH
Matt Koskela	OET
Mashkoor Malik	NOAA
Megan Cromwell	NOAA
Henry Jones	USM
Derek Sutcliffe	URI
Vicki Ferrini	LDEO
Stace Beaulieu	WHOI
Tina Haskins	WHOI
Lindsay Gee	OET
Joshua B. HIll	USM
Shawn Kalis	USM
Ben Ausdenmoore	USM
Dwight Coleman	URI
Adam Soule	URI
Jason Fahy	URI

Last updated on 2021-03-12
Approved by NOAA Office of Ocean Exploration and Research on 2021-07-07

Introduction

Purpose

This document contains the high-level description of the data management plan (DMP) for the Ocean Exploration Cooperative Institute (OECI). Due to the complexity and variety of the types of data collected by the OECI participants, and to allow for more flexibility in revision of these policies as new data types are added and methods improve, this document provides policies for data management at a high level of abstraction, and defers the details for individual data types to type-specific data plans (TSDPs) provided as appendices.

It is expected that component institutes of the OECI will establish more detailed DMPs for themselves, or for individual expeditions, using this DMP as a policy guide, and in particular, any projects that are generating data should include a project-specific data management plan as part of their proposal, and should include information on data storage and availability in their annual reports. All DMPs for OECI sponsored expeditions *must* be consistent with this DMP. For multi-institute expeditions, the DMP *shall* explicitly establish the responsibilities for data management for each of the institutes. To allow for clear communication in planning expeditions, all OECI component institutes are *strongly encouraged* to make available, through the OECI office at University of Rhode Island, their current and evolving DMPs.

Precedence

This document provides high-level policies for all data types, which therefore form a **minimal requirement** for the TSDPs. TSDPs *may* add further detail, but *must* remain consistent with these policies.

Data Collection Summary

Basic Information

This plan is for the period 2021-2025, and covers all data collected by members of the OECI during OECI-funded field expeditions on an on-going basis through this time range. Anticipated data types that will routinely be acquired include:

- Data catalog that provides metadata that describes the data package, formats, and naming conventions (Appendix A).
- Multibeam echosounder mapping data, to include bathymetric, acoustic seafloor backscatter, and water column acoustic backscatter (Appendix B).
- Sub-bottom acoustic reflection data (Appendix C).
- Optical imagery, video and still, from remotely-operated, optionally crewed, and autonomous vehicles in, on, and over the water (Appendix D).
- Oceanographic data, for example, sound speed profiles, weather observations, acoustic doppler current profiler data, etc. (Appendix E).

- Event logs and other associated metadata for cruise, and particularly dive, operations (Appendix F).
- Physical samples (Appendix G).
- Vehicle navigation and sensor data (Appendix H).

(Data types being collected by the various instruments are detailed in the appendices.)

Spatial Extents

This data collection is expected to be primarily in the U.S. EEZ, but may have occasional data collections in the Extended Continental Shelf, in international waters, or elsewhere (with appropriate permits).

Data Sources

The data will be collected primarily through ships of exploration, but may also include use of University-National Oceanographic Laboratory (UNOLS) vessels, or remotely-operated, optionally crewed, and autonomous vehicles deployed from those ships, or from shore to suit the requirements of the data collection and Expedition Principal Investigators (ePIs). Within this DMP, ePI is taken to mean the lead scientist/engineer in charge of effort associated with a data type, collection asset, or designated by the observing platform. Where multiple ePIs will be present on an expedition, they *shall* agree among themselves prior to the expedition starting, the duties and responsibilities for data management during the expedition.

NOAA Observing Systems Used

Data from the following NOAA Observing Systems of Record may be used:

- Argo Profiling Floats (Argo)
- Commercial Fisheries Dependent Data Surveys (NMFS CFD)
- Continuously Operating Reference Stations (CORS)
- Ecosystems Surveys (Ecosystems Surveys)
- Ecosystems and Fisheries-Oceanography Coordinated Investigations (EcoFOCI)
- Fishery Independent Surveys (Fish Surveys)
- Global Sea Level Observing System (GOOS GLOSS)
- Habitat Assessment (Hab Surveys)
- Hydrographic Survey (HYDRO)
- Landsat Series (Landsat)
- NOAA Ships (NOAA Ships)
- National Coral Reef Monitoring Program (NCRMP)
- National Current Observation Program (NCOP)
- National Water Level Observation Network (NWLON)
- Ocean Noise Reference Station Network (NRS)
- Physical Oceanographic Real-Time System (PORTS)
- Shoreline Mapping (Shoreline)
- Voluntary Observing Ship (VOS)

Responsible Party

The responsible party for development of this document is:

OECI Data Manager (TBD)
Graduate School of Oceanography
University of Rhode Island
215 S. Ferry Road
Narragansett, RI 02882
+1 401 874 6222
oeci-data-management@etal.uri.edu

The responsible party for top-level data management (TLDM) under the OECI is:

OECI Data Manager (TBD)
Graduate School of Oceanography
University of Rhode Island
215 S. Ferry Road
Narragansett, RI 02882
+1 401 874 6222
oeci-data-management@etal.uri.edu

While the individual participants within the OECI *shall* provide data managers for their own data collection efforts, in coordination with the TLDM, and *shall* provide resources within their OECI budget requests to support this role.

Data Stewardship

Data Maintenance

Data collected under this DMP *shall* be recorded to stable media in the field as appropriate for the data type (see TSDP appendices), and *shall* include a backup process to ensure that raw data from the instrument is preserved without any modification. PIs *shall* establish, document, and maintain an archival logical data structure (ALDS) appropriate to the data type and media (see metadata requirements), and preserve this for archival.

Data Processing

Where possible, processing of data *shall* be done in the field using a workflow appropriate to the data type (see TSDP appendices). Data products constructed in the field *shall* be maintained in the ALDS separate from the raw data.

Quality Assurance and Control

All data collected for the OECI *shall* undergo quality assurance and control (QA/QC) as appropriate for the data type (see TSDP appendices). Ideally, this would be done at sea, or as soon as possible after reaching shore. Results of QA/QC activities, and associated methods, *shall* be documented in the associated metadata and preserved with the data in the ALDS.

Metadata Construction and Preservation

Basic Principles

ISO 19115-compliant metadata *shall* be generated for the expedition-level data package, to include at least the minimal required metadata, and ideally the optional metadata described below and in the relevant TSDPs. ISO metadata *shall* be constructed post-cruise at the OECI Data Management facility, but the ePI *shall* ensure that all required metadata is collected in order to support this activity. Where possible, metadata *shall* be generated automatically from the data using appropriate scripts or programs. The metadata *shall* be maintained in the ALDS and transferred to the appropriate archive for the given data type. The template used for the ISO metadata *shall* be agreed with the relevant archive endpoint before submission.

All ePIs are *strongly encouraged* to establish a formal sign-off procedure for their data and metadata prior to leaving the ship, and prior to submitting data to the archive. The purpose of this process is to provide an assurance that the data is complete, and authorized for distribution and archival.

Minimal Required Metadata

At a minimum, the developed metadata *shall* contain:

- A unique identifier for the expedition (e.g., an expedition name, or DOI).
- Spatial and temporal extents of the data collection.
- A responsible party to answer detailed questions about the data collection.
- A responsible party to answer detailed questions about the archival of the data.
- A short description (abstract) for the data collection.
- A description (ideally machine-readable) of the data collection event.
- A comprehensive description of data lineage from raw to processed products (if any), including any corrections applied.
- The results of any QA/QC applied to the data.
- A machine-readable specification for contents of the ALDS, including indication of whether data was recorded in each category (Appendix A).
- File manifest and checksums to ensure the completeness and integrity of data files.

Optional Metadata

The generation of metadata specific to the data type (see TSDP appendices) is strongly encouraged in all data collection.

Data Archival and Dissemination

Archive Endpoints

In general, all raw data collected by OECI member institutions and PIs *shall* be deposited at NOAA's National Centers for Environmental Information (NCEI) for long-term archival, if acceptable to the archive. Due to the range of data types expected to be collected, it is expected that different data packages may be submitted to different sub-archives of NCEI; for example, bathymetric mapping data would be sent to the Boulder, CO facility. In some cases, different archives *may* be required for a single data type due to pre-existing collections or financial constraints (e.g., for physical samples). In this situation, details of which archives are being used *shall* be provided in the metadata. Details are provided in the TSDP appendices.

Products derived from raw data collected by OECI member institutions and PIs *may* be transferred to the appropriate NCEI archive at the discretion of the archive's data managers. Otherwise, OECI member institutions and PIs *shall* make any products available via a domain repository, an institutional repository, or other server, ideally through discoverable, standards-based protocols (e.g., Web Map Services [WMS], or a RESTful interface).

Availability

All data collected under this DMP *shall* be made available to the public unless national security restrictions or other legal restrictions prevent this.

Accessibility

Raw data *shall* be archived in the format generated by the instrument, without modification. Derived data (e.g., processed data, as well as final products) at any level of processing *shall* be archived in a machine-readable format that is fully documented. Preference *shall* be given to data formats which are open-source, or for which an open-source access library is available. Details are provided in the TSDP appendices. The format used *shall* be identified in the associated metadata, and a format document *shall* be provided in the ALDS if available.

Any PDF documents included in the ALDS should be Section 508 compliant wherever possible.

Intermediate Storage

Between data collection and archival, all data *shall* be maintained in a secure data facility that includes regular backups and offsite storage of the backup media. Ideally, this would be done by the ePI's institution, but if that is not possible, the Inner Space Center at URI *shall* act as the intermediate storage and preservation site for OECI data until it is archived. In addition, a full copy of all expedition data *shall* be sent to ISC at the end of data collection to provide a "backup of last resort" until the data is submitted to a long-term archive.

Ideally, data *shall* be maintained on the collection platform until it is confirmed to be safely transferred to intermediate storage, and *shall* be maintained in the intermediate storage facility at least until it is confirmed as accepted into the long-term archive. Where the collection platform cannot maintain data (e.g., for ePI instruments embarked on a host platform), at least two copies of the data *shall* be made on board, and *shall* be transferred ashore by separate routes (where possible) to ensure data continuity.

Timelines

Raw data collected under the OECI *shall* be made available to the archive endpoint within 60 days of the end of the data collection event, taken in this case to mean the specific expedition during which the data was collected, rather than the end of the prevailing OECI grant. The archival object *shall* be the ALDS, and *shall* include the appropriate metadata.

Data subjected to QA/QC *shall*, ideally, be provided to the appropriate archive endpoint within six months of the end of the data collection event, unless there are extenuating circumstances. If a second data submission is required, it *shall* use the same ALDS and allow for merging of the datasets at the archive. Metadata in the ALDS *shall* be updated to indicate any further processing, and the results of the QA/QC efforts.

Derived products shall be made available as soon as possible after the end of the data collection event, and in no case more than two years afterwards. If an auxiliary data submission is required, it *shall* use the same ALDS and allow for merging of the datasets at the archive. Metadata in the ALDS *shall* be updated to indicate any further processing.

Appendix A: Data Catalog

Introduction: The Data Catalog serves as documentation of the archival logical data structure (ALDS), and provides machine-readable metadata that describes the directory structure, naming conventions, and processing status. It is intended to provide data consumers with the information necessary to navigate the data and directory structures efficiently either with software tools (e.g., data managers) or through visual inspection of the data distribution (e.g., science users) accompanied by a human-readable data report generated directly from the Data Catalog to describe the data package.

Appendix version: 0.1.0 (2021-02-02).

Required Metadata:

The Data Catalog *shall* be provided in a JSON format that adheres to an agreed upon schema (described in a Data Catalog supplemental document (TBD)) for each sensor used during the expedition. This *shall* include:

- a list of the file types produced
- a description of the naming convention(s) used
- the absolute path to the data files
- confirmation of whether or not the sensor was used during the expedition or dive
- a summary of data type(s), format(s), and file extensions
- processing status (e.g., raw vs processed)
- a link to a file manifest with filename and checksums (e.g., md5deep)
- Should also include the path to the cruise report and clarification of its format

Optional Metadata: The Data Catalog will be able to accommodate additional metadata that is *strongly encouraged* but not required. This includes:

- Sensor position and offsets
- Sensor calibration information
- Device serial number

Appendix B: Multibeam Echosounder Data

Introduction

This appendix describes the data types collected during routine multibeam echosounder (MBES) mapping, the expected data workflow, QA/QC procedures, and derived product details required to complete the Type-specific Data Plan (TSDP) for MBES data. The focus is primarily on surface ship mapping, but specialization sections for subsurface mapping are provided where required. A number of other acoustic systems may be collected during an expedition (e.g., acoustic Doppler current profilers, EK80-style fisheries echosounders, etc.). Those data types are not covered by this appendix.

Note that different types of survey may require different data, and different processing methods. For example, surface mapping to support ROV/AUV operations may do significantly less processing than would be normal for primary mapping, and therefore may not require data for post-processing motion effects, etc. Expedition PIs (ePIs) *shall* therefore identify processing and data requirements prior to the expedition, and record their choices as part of the metadata.

Appendix version: 1.0.0 (2021-03-08).

Data Types Collected

Routine mapping with MBES *shall* include collection of at least:

- Bathymetric data observations in the instrument manufacturer's native format. This data *shall* contain sufficient data to re-solve for the sounding (e.g., two-way travel time, beam launch angle, surface sound speed, etc.) if required. Ideally, this format would be a published format.
- Seafloor backscatter observations in the instrument manufacturer's native format. This data *shall* contain at least one observation of seafloor backscatter per bathymetric observation (typically a mean or boresight acoustic backscatter measurement), but by preference should include the full time-series of backscatter per bathymetric beam ("snippet" or "time-series" backscatter).
- Watercolumn backscatter observations in the instrument manufacturer's native format. This data *shall* include at least the magnitude of the backscatter per time sample in each beam, but by preference should include complex data (i.e., phase and amplitude data).
- Motion time series for the observing platform in the instrument manufacturer's native format, or embedded into the MBES native format. The time series shall be at a frequency sufficient to capture the natural dynamics of the platform, normally at least 100Hz. Ideally, sufficient data shall be recorded to allow for post-processing of kinematics.
- Positioning information for the observing platform in the instrument manufacturer's native format, or embedded into the MBES native format. The time series of position shall be at least 1Hz. Ideally, sufficient data shall be recorded to allow for post-processing of

kinematics. This data might be collected in common with, or in addition to, ship's positioning data (Appendix E).

- Vertical profiles of the sound speed variations at the observing platform in the instrument manufacturer's native format. Sufficient data shall be recorded to allow for post-processing of the sound speed profile, including positioning information. The sound speed profile (SSP) *shall* be recorded to either the full depth of the water, or until the profile becomes pressure dominated, whichever is shallower. This data might be collected in common with, or in addition to, oceanographic measurements during the expedition (Appendix E), but should remain with the MBES data, since it is essential for (re)processing.
- Draft measurements for the observing platform (for water level-based vertical correctors, if necessary).
- A static calibration ("patch test") dataset to allow for resolution of static offsets between the components of the survey system. The dataset *shall* include, if at all possible, the latest positional survey of the sonar system and its components.
- Any and all available self-tests run for the MBES system (e.g., Kongsberg BISTs), and [a swath extinction curve](), if available.

Data *shall* be written into at least the instrument manufacturer's raw data format, e.g., Kongsberg KMAll (bathymetry, backscatter, auxiliary data), and WCD (watercolumn); Reson S7K (all data). Watercolumn data *shall* be recorded into a separate file from the bathymetry and backscatter data; positioning data *shall* be written into this file in addition to any other recording. File naming conventions *shall* follow the manufacturer's standard practice, and *shall* be preserved into the Archival Logical Data Structure (ALDS) without modification. Additional data formats may be used for real-time conversion of data (e.g., Generic Sensor Format, GSF), but *shall not* be the only capture of raw data.

Subsurface Mapping: It may be impossible to provide all standard data types with subsurface assets, particularly full positioning information (except in post-processing), sound speed profiles, and draft measurements. Where possible, however, the appropriate equivalents (e.g., resolved USBL position, DVL, computed sound speed, and pressure sensor depth) should be recorded in the ALDS and documented in the metadata. Patch test data is often measured directly on subsurface systems, but *shall* be confirmed at least once per season, but ideally on each expedition, through direct observation, and recorded in the metadata.

Best Practice: Many MBES implementations attempt to capture all data sources into a single data file. Prefer, however, to record the watercolumn data into a separate file in order to limit file sizes for bathymetric data, and thereby speed up processing. Motion data recorded in MBES files often lack the data required for post-processing, so a separate file, in the manufacturer's native format, in addition to what is recorded by the MBES in real time is recommended if available. Similarly, SSPs recorded by MBES that apply them in real time can often lack data for post-processing, and therefore should be recorded separately. Frequency of SSP capture is a complex question which depends strongly on the local oceanographic environment, but for deep water should be no less than once every six hours or if the difference between the surface sound speed and the sound speed profile is greater than 5 m/s. The offsets in position and angle ("patch test") results can often be recorded or applied in multiple locations within the dataset, which can lead to confusion or double-application. Prefer to record the values in the MBES data files, and apply the offsets in the sonar hardware, where possible, and carefully document whichever method is used in the metadata. Optimization of data collection for backscatter can take some effort; see the GeoHab Backscatter Working Group guidelines for further detail. Multibeam systems are complex and need continual maintenance and monitoring for best effect; the UNOLS Multibeam Advisory Committee has many guides on this topic, and is recommended for advice and/or consulting on multibeam issues.

Data Workflow

Bathymetric Data

MBES data *shall* be processed during acquisition in order to ensure that any difficulties are caught and resolved during the expedition, and then refined on shore. For subsurface mapping, processing *shall* be done after each dive. All OECI MBES "at sea" bathymetric data processing *shall* consist of:

1. Conversion to a processing format. This will depend on the software in use, but is typically a manufacturer's proprietary format.
2. Application of static and dynamic correctors. This *shall* include at least static offsets in position and orientation between the components of the MBES system, refraction corrections using the collected SSPs, and motion dynamics of the observing platform. In regions where the expected tide is more than 0.1% of the observed depth, water level vertical corrections *shall* be made using either a predicted or observed water level data, or through ellipsoidal referenced survey techniques.
3. Assessment of the total propagated uncertainty of the individual soundings.
4. Generation of a "working" data product. This can take different forms depending on the processing system in use; see ***best practice*** below.
5. Iterative quality assurance inspection of the data, or "working" data product, as required, to resolve any extraneous or mis-solved soundings and/or their effects.
6. Resolution of all operator interventions to provide a preliminary "at sea" data product.
7. Iteration of steps 5-6 until all issues are resolved.
8. Quality assurance of the preliminary "at sea" data product against any extant data in the area, or cross-checks against contemporaneous "cross line" data if none exists. The results of the QA checks *shall* be recorded in machine-readable form in the ALDS, and *shall* be described in the metadata.
9. Creation of finalized "at sea" data products as outlined below.

All OECI "on shore" data processing *shall* follow the same general trend as the "at sea" processing, starting from the finalized "at sea" data product, which may be the final product at the discretion of the ePI (i.e., no "on shore" processing is done). Post-processed solutions for motion, positioning, water level corrections, and SSPs *shall* be applied if available. Finalized data products shall be generated as described below, and recorded in the ALDS. Details of the processing methodology (include “none done” if no extra work was completed), and especially any variations of this general plan *shall* be recorded in the data lineage minimal metadata.

Subsurface Mapping: For subsurface assets, the sequence of events in the processing sequence may be different (e.g., the correct positioning information may only be available after significant post-processing), which will reorganize the workflow. Any deviations from the workflow as well as all the non-standard data employed (e.g., Kalman filters backward solution, seafloor mooring tide sensors, etc.) *shall* be documented in the metadata (typically in a report of survey or cruise report).

Best Practice: Details of workflow will vary with processing software. The workflow used, and software type and version should be recorded in the metadata in order to assist in tracking down potential problems later. In all cases, ePIs should attempt to preserve the ability to reverse the processing flow, or reprocess the data, whenever possible. If the ePI’s institute does not already have a similar practice in place, it is recommended that ePIs use the NOAA OER Deepwater Exploration Mapping Procedures Manual (or other similar practice) as a guide to standardize expectations and products. Expedition PIs are **strongly** encouraged to use grid-based processing methods whenever possible, and to take advantage of automated or semi-automated methods that are ubiquitous in modern processing software in order to provide more objective and efficient processing, and more useful output products. Grid resolution should either be automatically determined by the software, derived from the mean depth in the area, or allowed to vary according to the data in software which supports it (very strongly recommended), and should not be set by arbitrary operator choice.

Backscatter Data

“At sea” backscatter processing *shall* consist of constructing a backscatter mosaic using the finalized “at sea” bathymetry. This *may* be refined in “on shore” processing if the bathymetric model is changed to accommodate post-processed data.

Best Practice: Configuration and processing methods for backscatter have been extensively studied by the GeoHab Backscatter Working Group, who have published a detailed report on the matter. Use of these procedures for backscatter is strongly recommended. Depending on the MBES in use, optimizing for backscatter can impact bathymetry data; all expedition PIs are recommended to assess data needs before the expedition and document these decisions in the metadata.

Watercolumn Data

Unless required by the expedition, “at sea” processing of watercolumn data is expected to consist of ensuring that the data is being recorded and preserved in the ALDS unless there are targets of scientific interest in the area (e.g., gas/water seeps).

Best Practice: Availability of tools for water column processing, and staffing to support it will vary depending on the platform. Currently, routine processing of all water column data is not always feasible at sea. However, for limited areas, processed with tools such as QPS FMMW, it is possible to search for scientific targets of interest. Products from this type of processing can include the location targets of interest (spreadsheet, *.csv/*.tsv, *.shp file), screen capture from WC processing and possibly image of location with associated bathymetry or seabed backscatter.

QA/QC Procedures

Bathymetric data *shall* be inspected for quality control and assurance during “at sea” and “on shore” processing. This will consist of comparing the data collected against archive holdings in common areas, and/or collecting orthogonal lines of data during the data collection event and comparing areas of common ensonification. A number of commercial and open source tools are available for this purpose, and will be used as appropriate to the data. Results of the analysis *shall* be recorded in the metadata and archived, if possible, in the ALDS.

Product Data Types

The following data types *may* be made for product objects:

- Processed sounding-scale bathymetric data in a well-defined format, for example Generic Sensor Format (GSF).
- Processed grid-based bathymetric data in a well-defined format, for example Bathymetric Attributed Grid (BAG).
- Processed acoustic backscatter mosaics in a well-defined format, for example Geographic Tagged Image File Format (GeoTIFF).

At a minimum, each format used *shall* have a PDF format description corresponding to the version of the data format being written, which *shall* be preserved in the ALDS if available. Ideally, it will be an open-source format with a support software library to read and write the data. For grid-based outputs, a suitable alternative is the availability of a Geospatial Data Abstraction Library (GDAL) driver. The data format *shall* be specified in the metadata, and a format document *shall* be provided in the ALDS if available.

Best Practice: Avoid the use of plain ASCII data for outputs, since it generally causes significant data loss. Document the software being used (name and version) to make the outputs, since significant variability in output format has been found between different systems and versions. Document the version of the format being generated. Prefer binary data formats, particularly for backscatter. Since auxiliary (e.g., metadata) files can readily be lost, prefer formats that can encode their metadata internally so that this cannot happen. Prefer common formats (e.g., BAG, ESRI Grid) over more obscure formats (e.g., S-102) or ones that do not preserve metadata well (e.g., TIFF for bathymetric grids) unless there is a compelling need. Prefer output products that can continue to be manipulated (e.g., BAG) over ones that cannot (e.g., a PDF hardcopy of a map). Prefer machine-readable structure formats for metadata (e.g., XML, JSON) over plain text, Word document, or PDF.

Additional Metadata

For MBES data, the following metadata *shall* be included in the description:

- All data format documents for data recorded in the collection event, or constructed from that data.
- Documentation of the calibration ("patch test") data, and any available engineering survey of the observing platform that specifies linear and angular offsets between the components of the survey system.
- Full specification of the horizontal and vertical coordinate reference frames used in the survey (whether geodetic frames, or local frames constructed from the data, e.g., for subsurface mapping).
- Software packages and their version numbers used for processing, or data product generation.
- Version information for the firmware running on all components of the survey system.
- Make, model, and serial number of all components of the survey system.
- Specification of the model used for TPU estimation, and the component measurement uncertainties used to feed the model.
- Description of the data processing strategy used, and the workflow specific to the software package in use.
- Results of all final QA/QC checks conducted on the data "at sea" and "on shore".
- Method used for vertical correction (e.g., water level, ellipsoid, pressure sensor) and associated processing methods and data sources.
- Selection method for SSP used to process the data, and any processing parameters used for the SSPs (e.g., extended using an oceanographic atlas, converted from Expendable Bathythermograph [XBT] data using a fixed salinity, etc.)
- Especially from subsea assets, any and all parameters used in data processing algorithms (if used) and product construction.
- Operational parameters of the MBES system, unless they are encoded in the data files being written by the system.

Best Practice: Record more than you think you need to. Data processing packages and MBES firmware have been known to have bugs that can be fixed after the fact, but only if the specific version of software/firmware is known. Recording specifics of the data workflow is critical in avoiding common mistakes (e.g., applying the same corrector twice). Prefer to record the workflow directly in the metadata to avoid transcription errors, or having the details hidden in another report.

Archival Logical Data Structure Components

The specific structure of the ALDS will vary according to the needs of the particular data collection event. Therefore, no mandatory structure is specified.

Best Practice: Maintaining a separate directory structure for raw and processed data is recommended. Some ePIs prefer to structure data files by date of collection (e.g., a hierarchical directory tree of year-month-day), but most current processing systems do not require this, and in deeper water this structure tends to be cumbersome. A simple directory containing all of the raw data files (by type) is usually simpler. Provide a separate directory for data format documents. The naming scheme for directories is not critical, but consistency in the naming of products is. Since names are readily changed or mangled, however, prefer internal metadata (i.e., metadata encoded in the file itself) rather than attempting to encode some level of metadata in the name of the file. Prefer a single metadata file in a separate directory over small pieces of metadata scattered throughout the ALDS. Working copies of data, and working directories for processing, should be held externally to the ALDS to avoid the potential for a processing reset removing raw archivable data. Prefer ALDS storage that implements at least RAID5/6, and when possible have a network-connected duplicate RAID doing incremental backups of the main ALDS. Send ALDS copies ashore through different routes. Avoid long pathnames and deep directory trees where possible to avoid occasional legacy operating system limitations; avoid special characters, and especially spaces and case-sensitive characters, in filenames for the same reasons. Prefer an automated data management system to ensure consistency in data placement.

Archive Endpoint

All raw and processed MBES mapping data shall be archived at NCEI Boulder.

Appendix C: Sub-bottom Acoustic Reflection Data

Introduction

This appendix describes the data types collected with Chirp acoustic reflection systems, sometimes called subbottom (SBP) profilers, and the expected data workflows, QA/QC procedures, and derived products under the Type-specific Data Plan (TSDP) for SBP data. For recommended “Best Practices” for Chirp Acquisition and Processing, refer to Saustrup *et al*. (2018).

Appendix Version: 0.1.0 (2021-03-08).

Data Types Collected and Workflow

Standard chirp systems generate as primary output a match-filtered full-waveform record of sinusoidal aspect (i.e., both positive and negative values). However, it is common practice with chirp systems to transform full-waveform records into envelope records, which provide a higher contrast spectrum but contain positive values only. Usually this last output is considered the SBP “raw” data, and is archived either in the proprietary format or directly converted to the Society of Exploration Geophysicists SEG-Y format (SEGY). Our recommended best practice is to acquire, if it is possible, the full waveform output in order to have greater confidence in the stratigraphic interpretation at all scales. Amplitude envelope records can always be derived from the full-waveform in post processing.

It is noted that most SBP data, after being matched filtered and SEGY converted, are directly usable and do not undergo any further processing. However, a quick and automated level of processing of chirp data can render the envelope data much more interpretable, and the underutilized, higher-resolution, full-waveform data more accessible. The following processing steps are only suggested and optional.

Survey Geometry Correction

Recording Delay Correction: The data are corrected for any recording delay (nonzero start recording time, also called deepwater delay) that may have been used in the field by applying the appropriate static correction.

Towfish Depth: A time series for towfish depth is recorded in the field and is used to correct the observed data to a sea-surface datum. This depth can be estimated using a variety of methods, including cable length/angle, a pressure sensor mounted on or integrated into the towfish, or ascertained with a USBL system. For best results, this step should be performed before the seafloor picking. Note this correction is different for the Static Correction applied to AUV-borne chirp survey, that will be specified below.

Heave Compensation: Heave artifacts are present when the towfish is pulled up and down as the ship responds to wave motion. They are at once the most destructive factors in image quality.

AUV-borne chirp survey

SBP data collected on AUVs *must* contain the information to correct for the time shift due to the AUV changing depth during survey. This correction is usually automatically performed in any visualization/interpretation software (i.e. IHS Kingdom, SonarWiz, etc.) during the import data step, as long the SEG Y Trace Headers correctly contains the following information:

- Lag Time (Time Zero) in milliseconds, bytes 105-106
- Additional Start Time in microseconds, bytes 181-184 or otherwise specified.

Signal Processing

Signal processing includes minimal operations such as frequency filtering, deconvolution, gain correction, and water column muting. Some operations such as frequency filtering and deconvolution, can only be applied to the full-waveform data traces.

Frequency Filtering. Full-waveform data are bandpass-filtered using a filter comparable to the source wavelet band (e.g., 700- to 12,000-Hz Butterworth Filter, with a filter length of 91 samples). This step primarily removes low-frequency towing noise.

Deconvolution. Ideally, if the data are matched-filtered by the outgoing pulse, chirp data should not require additional signal processing. However, the match filter is not perfect, presumably owing to differences between modeled and actual outgoing pulse waveforms; this results in ringy reflections in the full-waveform record. Image quality can therefore be significantly improved with a standard predictive deconvolution technique.

Gain Correction. Amplitudes are corrected to account for transmission loss and spherical divergence. This step is done using a water-velocity spherical divergence correction followed by a windowed lateral trace balance.

Water Column Muting. Data may be muted above the picked seafloor arrival time to remove any water column noise. This is done primarily for generating a cleaner display for publication. This step should not be performed, however, if there are features of interest in the water column (e.g., active gas seeps).

Product Data Types

The subbottom data management pipeline at NCEI relies heavily on SEGY as a means of extracting navigation necessary to generate tracklines that display the location of the data in the Trackline Geophysical Data Viewer. Data submitted in unsupported formats will still be accepted but will not be discoverable through the web services provided at NCEI.These data can only be accessed from the archive upon request to trackline.info@noaa.gov. If your data are not SEGY, the ePI *shall* email NCEI to discuss the options available for their data.

Best practice for archiving
Archive the following:
(1) raw acquisition system file format containing all data channels (e.g. .jsf format for EdgeTech systems, .kea and .kab for Knudsen systems);
(2) SEG-Y files of both envelope and full-waveform data. Adopting the standard SEG Y Trace Header structure, examples: FFID in bytes 9-12; Position (Lat, Long, East, North) in bytes 73-76 and 77-80; for AUV-borne survey Lag Time (T0) in bytes 105-106; additional start time in bytes 181-184 (suggested).
(3) cruise report;
(4) cruise logs;
(5) lineage (all step in regards to how data has been collected, processed, or manipulated).

Reference:
Saustrup, S., J.A. Goff, and S.P.S. Gulick. (2018). Recommended “best practices” for chirp acquisition and processing. U.S. Dept. of the Interior, Bureau of Ocean Energy Management, Gulf of Mexico OCS Region, New Orleans, LA. OCS Report BOEM 2019-039. ix + 16 pp.

Society of Exploration Geophysicists Technical Standards - https://seg.org/Publications/SEG-Technical-Standards

Archive Endpoint

All Subbottom data shall be archived at NCEI Boulder.

Appendix D: Optical Imagery (Video and Still)

Appendix version: 0.1.0 (2021-03-08).

It is recommended that deck-to-deck video be recorded and saved so that watercolumn video will be preserved, but it is recognized that this is dependent on available digital storage onboard the vessel.

By-Vehicle/Sensor Applicable Format Recommendations

Remotely Operated Vehicles

- Full-dive recordings, typically the ROV's main HD forward-facing camera lens, should be **compressed video files**. Choose a level of compression for the amount of storage.
- Special highlights, depending on the research vessel's storage capabilities, may be much larger losslessly compressed or **uncompressed video files** if necessary for post-production use or sensitive graphical processing of small objects, such as plankton or marine snow.
- Specific small objects studies may require **uncompressed video files**, but may use **uncompressed stills** as well when motion isn't required
- Ancillary cameras should be **compressed video files** if they are recorded. Typically ROVs have several cameras and the main forward camera is the highest resolution.

Ship-Mounted Cameras

If used and available:

- Operational camera views (e.g. aft deck, over the side) should be saved as **compressed still image files**, if images are exposed natively
- Video files should be saved as **compressed video files**, if video is exposed natively

Autonomous Underwater Vehicles

- Long time series video and still images should be saved in **compressed** formats.
- Photomosaics should be saved as raw still images, associated navigation, and compressed composite images.

Video Plankton Recorders

- Solutions typically capture and process H264 video black and white hi-res imagery packed in MP4 for processing - save the MP4s as **compressed video** or save the images as **compressed images**.

Format Recommendations

Compressed Video Files

The choice of amount of video compression, in Megabits per second, is important to operations. In the case of a smaller research vessel with less storage space, they may opt to record their master full-dive recording in 8 Mbps. This is small enough for the video file to be used on the web in Video-On-Demand applications if recorded with the right codec and wrapper. A larger research vessel may have a full-dive recording in 150 Mbps and create a proxy resembling the smaller research vessel's master, at 8Mbps.

From OER Data Management Team Report - Video Data Management Modernization Initiative (VDMMI) Video Data Management Best Practices: *Video should be captured at the highest quality levels with the lowest level of compression consistent with available resources and expected downstream processing and use.*[1]

- Capture video at the native resolution and framerate of the camera source in use
- Prefer a capture codec of H264 and wrapper of MP4 if possible
- Avoid recording to proprietary formats and avoid recorders that do not support H264 MP4 recording
- Adjust the compression (bitrate) of the video capture device to the least compression possible *consistent with available resources and expected downstream processing*
- Date and time of video start should be made part of the video filename to aid discovery and to enable metadata location and extraction from other sources
- Date and time should be captured in frame metadata in the video file if possible
- Creation of proxy videos should use a codec of H264 and a wrapper of MP4 whenever possible. 5Mbps is a sensible bitrate for quality and portability of 1080p proxy
- Creation of proxy videos should prefer to retain the original frame size (resolution), but drop the bitrate
- Audio should be compressed onto video at 48 kHz if recorded
- Prefer to record audio separately at the native resolution of its source, if possible, before being embedded onto video. E.g. an intercom should be recorded separately if possible if the recording is valuable on its own, otherwise it's trapped with the video stream

Uncompressed, losslessly uncompressed, and visually uncompressed Video Files

If there is a use case requiring pixel-specific accuracy, it is usually for something limited in time duration, and could be replaced with a series of captures as uncompressed image files, or an uncompressed/losslessly compressed/visually-losslessly-compressed file instead. Some examples of raw captures being useful:

- Counting very small creatures in a large frame that cannot be zoomed in on individually
- Signals sent from very high resolution cameras with low framerates, 360, or light field capture capability

File Naming Conventions

Recommended file naming convention:
UNIQUEID_ISO8601time_free*_free*.<ext>

[1] https://www.ncei.noaa.gov/sites/default/files/2020-04/VDMMIProjectDocumention_FINAL_Dec2016.pdf

- “UNIQUEID” could be {Survey/Asset/Deployment} or {[cruiseID/vessel]_[cameraID]}, and “free” could be a content code, frame number or camera ID.
- Where “ISO8601time” is the UTC start time of the recording in the format: yyyymmddThhmmssZ, where yyyy = four-digit year, mm = two-digit month, dd = two-digit day, hh = two-digit hour, mm = two-digit minute, ss = two-digit second, “Z” = UTC timezone.

E.g. X1907_VID_20191107T135000Z_ROVHD.mov

- Video files should include a single start time, and the start time should be stamped onto the file in its filename. ISO 8601 is the most preferred format for the timestamp, but in Windows file systems, colons are not supported in filenames, and can be replaced with the underscore character or a hyphen.
- As recording softwares may not let you choose how to write the filename, it is important that ISO 8601 time can at least be reconstructed from the actual timestamp during processing.

Video Timecode Metadata

Time-syncing is of paramount importance with underwater video as it is the key metadata field that enables linking video to other data and metadata. Priorities with respect to time-syncing are: (1) for everyone to use Network Time Protocol (NTP) and synchronized clocks, (2) use GPS Coordinated Universal Time (UTC), (3) use a single master clock synced to GPS and providing NTP. External verification of time-sync is recommended.

- It would be ideal if the ISO 8601 timestamp included milliseconds in the filename when possible. ISO 8601 allows extension past seconds with a ‘.’ or ‘,’ followed by decimal seconds. This extra precision will allow the synchronization of higher-resolution data with the video such as video overlays and non-video data with highly accurate timestamps.

It is recommended that there is redundancy with respect to metadata embedded in video files, and that the metadata is provided in companion files, and that there are multiple methods for accessing metadata. Tools, such as QTchange, are currently available to inject/read time-code and should be utilized wherever possible. Operators are encouraged to report on the method and accuracy/confidence of time-syncing in accompanying metadata.

Metadata Priorities

Short-Term

- Standardized file naming convention and file-level metadata (listed below)
- Acquire and provide critical (meta)data, identifying resources and standards to automate metadata insertion
- Discontinue use of burned-in video overlays and audio timecode
- Embed metadata on a closed captioning (CC) channel based on recommendations

- Gather available resources to develop an accepted vocabulary for annotation
- Time-stamp annotation with UTC, GPS time-synced, and make available with video files
- Develop recommendations for IGSN (International Geo Sample Number, http://www.geosamples.org/) documentation for samples
- Include audio channel information in metadata for operators that currently record audio

Medium-Term

- Add capabilities for metadata on CC channel where it's not yet available.
- Add capabilities for operators to implement audio narration on video that do not currently record audio

Long-Term

- Identify standards to enhance optional metadata
- Embed additional information into video files (e.g. timecode and geospatial data)
- Develop resources/tools to more easily use annotation files with video
- Address mechanisms/frequency for updating authoritative ISO (International Organization for Standardization) records for video resources with annotation summaries
- Pursue use of additional audio channels for other (meta)data?

Accompanying Metadata

- Dive metadata
    - Dive ID and Cruise ID, possibly Project Name and PI(s)
    - Vehicle Name
    - Geographical Area (e.g. feature or study site name(s))
    - Copyright/licensing (or point of contact for such information)
- Vehicle metadata
    - Temporal extent (start/stop time),
    - Geospatial extent (min/max lat/lon),
    - Vertical extent (min/max depth),
    - Camera tilt angle, if available
    - Camera information (make, model, serial number, native format, frame rate, field of view, zoom),
    - Scaling laser separation and date verified.
- Descriptors:
    - "Packing list" describing data distribution contents/format so scientists/data managers know what they are receiving.
    - Point to cruise metadata, dive metadata, dive summary, cruise reports, guidelines/how-to documents, etc.
    - Point to companion data
    - Audio: availability (yes/no), which channel includes which content (science, technical, or other)
    - CC: is CC metadata included (yes/no)? Guidelines on how to use
- Companion data
    - All time-series data should be UTC time stamped & GPS time-synced
    - Checksums should be provided for all data files

    - Critical Companion Data
        - Platform navigation (vehicle, vessel, or diver)
        - Include navigation data status (raw, processed, etc.)
        - Vehicle attitude (heading, depth, altitude, pitch, roll
        - Annotation (EventLog) [details below]
    - Optional Companion Data
        - Environmental data (temperature, salinity, O2, etc.)
        - Camera attitude (pan, tilt, zoom)
        - Vehicle lighting configuration?
        - Offsets (sensors, cameras, etc.)
        - Sample IDs (SESAR recommended)
        - Transcripts of audio

Metadata on Closed Caption Channel

Metadata embedded on the Closed Caption (CC) channel is recommended as a means for displaying key metadata for situational awareness, as well as providing a visual QA/QC for technical purposes. Embedded metadata will not provide visual display of audio channels.

Recommended metadata to include on the CC Channel:

- Dive ID
- Date/time (yyyymmddThhmmssZ)
- Latitude/Longitude (meter precision)
- Heading (integer, units)
- Depth (integer, units)
- Altitude (integer, units)
- Optional text field

Embedding metadata in CC requires hardware/software and may not be currently achievable by all operators. Other community members who have already implemented the embedding of metadata can provide advice on implementation (i.e. Ocean Networks Canada, ROPOS).

Embedded Geolocation Metadata

- Embed geolocation metadata if possible or provide companion files
- Additional information is needed regarding the possibility of embedding geolocation data in video files, available standards that could be adopted, and potential future uses (e.g software tools) to extract and utilize the information

Subtitles / Watermarks / Video Overlay

- Prefer not to “burn-in” a logo, metadata, or text on top of the full-dive video recording
- Burning-in graphics for live broadcast or situational awareness means the shoreside recordings will contain these burn-ins, so they should be avoided unless needed

Annotation/Event-Log

- Because the eventlog is time stamped and the video should be timestamped (via filename or intra-frame metadata) it isn't required to embed eventlog entries into video
- If possible, it is beneficial to record eventlog timestamps and text inside the recording files to which it is applied (a copy as embedded metadata)
- It is beneficial to have the event logger able to record still frames of video when an event is fired. This can be used for QC or context upon later use

Onboard Recording Media

There is no recommendation for short-term storage or recording media, and adhering to a common archive media type is not critical. It is, however, critical that video content be preserved on media that will persist, and that redundant copies are maintained.

Requirements for Transporting Video

- Maintain two copies of all files on vessel
- Create one copy for the home institution/Principal Investigator for distribution
- Create one copy for URI Inner Space Center for OECI archival
    - Provide on LTO tape *if available*
- Once the two copies are verified, they may be deleted from the vessel

Telepresence and Video Recording

- Even if streaming to shore and recording on shore, at least save a small copy on ship
- Video sent to the ISC via satellite are recorded on Wowza units and served from a SAN
- These video files are available from the SAN for local participating scientists as well as production for during-cruise needs

Requirements for Archival and Access

There is currently no centralized public video and imagery repository. Therefore, a best effort shall be made to provide access to the data until a more formalized long-term solution can be determined. Until further notice, the URI Inner Space Center shall serve as the archival facility for the OECI.

At a minimum, the data manager *shall*:

- Provide proxy video data in the cruise data package to expedition principal investigator
- Provide proxy video data and full resolution video data in the cruise data package to the home institute
- Provide proxy video data and full resolution video data in the cruise data package to the URI Inner Space Center.
    - If the home institute does not have an effective way to make video data available, the URI Inner Space Center can assist with this task by providing data on hard drive or via cloud.

Appendix E: Vessel Navigational, Meteorological, and Oceanographic Data

Introduction

This Appendix describes collection and archiving of Vessel Navigational, Meteorological, and Oceanographic data. We distinguish ship-provided data from PI-provided data (from PI-provided instruments), and we distinguish data collected while underway (i.e., hull-mounted or flow-through) from data collected over-the-side (e.g., lowering an instrument on wire). Some data types collected by the research vessel are described in other appendices (e.g., Appx. B Multibeam). For navigation and sensor data from vehicles that operate beyond the research vessel, see Appendix H.

Data Types Collected

Data types will differ between vessels. Typical data types for UNOLS vessels are listed by the R2R repository: https://www.rvdata.us/data. NOAA OER's Data Access landing page has a 'List of Archived Data Types' following a cruise on NOAA Ship Okeanos Explorer: https://oceanexplorer.noaa.gov/data/access/access.html. At the end of this section, we provide an example from the Okeanos Explorer Oceanographic Data Pipeline.

Underway

- Ship navigation data
- Shipboard meteorological sensor data
- Sound velocity probe (note: sound velocity profiles may also be contributed with Multibeam data)
- Shipboard oceanographic flow-through sensor data (e.g. TSG, fluorometer)
- Water column sonar data (e.g., ADCP, EK80) (distinguish from Watercolumn and Backscatter data in Appx. B Multibeam)
- Trackline geophysics: Gravimeter
- Data from PI-provided instruments (e.g., instruments deployed in line with the flow-through seawater that are in addition to the ship-provided instruments)

Over-the-side/towed

- CTD cast data[2]
- XBT cast data[2]
- Sound velocity profiler cast data[2]
- Winch data (e.g., CTD, sled or net tow, coring or dredging)

---

[2] Note: sound velocity profiles may also be contributed with Multibeam data.

- Data from PI-provided instruments (e.g., instruments deployed on CTD frame or wire that are in addition to the ship-provided instruments)
- Magnetometer
- Sensors on net tows (e.g. MOCNESS, bongo, trawls)

Processing

Some processing is expected, e.g., conversion to scientific units based on instrument calibrations. However, it should be noted that some of the oceanographic data types may need further processing to be science-ready (e.g., CTD cast data).

Archiving collected* data:

We recommend archiving the collected data in its original form, and to also archive when possible the data in non-proprietary formats.
The collected (* some, processed) ship-provided data shall be submitted to an archive through one of the following mechanisms:

1. R2R, applicable for most UNOLS vessels,
2. if R2R not applicable, then we recommend specialized domain repository:
    a. Example: Water column sonar data - NOAA https://www.ngdc.noaa.gov/mgg/wcd/
    b. Example: Meteorological - SAMOS: Shipboard Automated Meteorological and Oceanographic System https://samos.coaps.fsu.edu/html/
3. if no domain repository then contribute directly to NOAA NCEI.

For PI-provided data, R2R would not be applicable. To decide where to submit/archive these data, first consider the data policy of the funding agency for collection of those data. If the funding agency does not direct the deposition of those data, then follow steps 2 and 3 above.

Appendix F: Event Logging

Overview

Events that take place during a cruise or dive should be logged. An 'event' is any observation, whether scientific or operational in nature, that is worth recording and that may not already be directly captured in textual form by other data logging systems. Examples of scientific events include in-situ biological and geological observations, deployment and recovery of sensors, and start/end of surveys. Examples of operational events are things such as cruise/dive milestones (I.e. 'on-station', 'vehicle-in-water', 'vehicle-on-bottom', 'start-of-survey', etc). An example of an event relevant to both science and operations would be the collection of a sample and its placement on a sample tray.[3]

Event logs provide a more systematic record of what occurred during a cruise or dive and can help recreate daily operations that might be written into a cruise report. Metadata for the event logger should include (but not limited to) the cruise ID, dates of the cruise, information about the cruise, name of the primary investigator.

Each entry should include the following:

- Date and time (UTC)
- Position (latitude and longitude)
- Type of Instrument or Platform (e.g. CTD, net, buoy or ROV, ASV, AUV)
- Deployment type (e.g. cast, dive, deployment, recovery)
- Author

Additional useful information include but is not limited to:

- Dive ID
- Cast ID
- Sample ID
- Comment field for anecdotal information

Events can be edited or entered at a later time than they occur in the event that the data logger missed an entry, or if a logger wants to add additional comments to an existing event. For example, users can add information such as species ID, sample ID's and other useful information as the operations are reviewed and samples are processed.

Vessel Events

Vessel is defined as the ship or main platform for data collection. Each use of a sampling device (profiling instrumentation, corers, nets, drifter deployments, etc.) should be logged with a unique ID and ancillary information to describe the event, such as UTC time and position. It is recommended to log vessel events separately from vehicle events because each data set will go to a different archival repository.

---

[3] https://usermanual.wiki/Document/sealogUserGuide.1333645115

The event log for the vessel should be started when the vessel leaves port, and finalized when arrived in port. The onboard project primary investigator or other designee is responsible to review the log periodically during a cruise to ensure completeness and accuracy. The vessel event log is utilized for over-the-side deployments, sonar surveys, and any other activities that affect data quality of the shipboard integrated sensors.

Example Events:
- CTD in the water
- CTD cast start
- CTD at depth
- CTD cast end
- CTD on deck

Vehicle Events

Vehicles are defined as platforms operating independently from the main vessel. This includes (but not limited to) remotely operated, autonomous surface and underwater vehicles. In the case of ROVs, events are typically logged to identify when observations are made, samples are collected, and if there are any issues during the course of the dive. When setting up the relevant metadata for a dive, the dive ID shall be provided.

Example events:
- ROV in the water
- ROV on bottom
- OBS: Biology
- OBS: Geology
- SAMPLE: Core 1
- SAMPLE: Bio Box A
- ROV off bottom
- ROV on deck
- START Transect
- END Transect
- PROBLEM: description

Export for Archival

The final event logs for the vessel and vehicle(s) shall be exported as a plain text delimiter separated value (e.g. *.csv or *.tsv) file in the ALDS. The event logs shall be ingested at the appropriate archive, for example vessel events at R2R[4] and vehicle events MGDS[5]. Location of the event logs shall be maintained in the Data Catalog (Appendix A).

---

[4] https://www.rvdata.us/data
[5] https://www.marine-geo.org/submit/

Logging System Recommendations

The OECI does not require a specific system with which to log events, but can make some recommendations:

- R2R Shipboard Sampling Event Logger
- Sealog

Appendix G: Physical Samples

Introduction

This appendix describes the description, processing, and archiving of physical samples collected during OECI activities. It is the responsibility of the operator of the collecting platform (i.e., vehicle or vessel) to ensure that samples are properly documented and delivered to an endpoint archive.

Naming

Samples should be named according to a standardized format that produces a unique sample ID and that is consistent with vessel or vehicle operator or platform used. Sample names should be consistent throughout a cruise and indicate the sequential order of their collection.

Sample Processing

It is expected that for samples that require processing, the methods will be well documented and linked to samples via cruise reports. If samples are split (i.e., subsampled), the parent sample should be documented first, and if desired child samples can be given a new ID based on the parent ID. (e.g., parent = NA122_012, children = NA122_012a; NA122_012b). Sample naming convention should be determined by the ePI and operations group. An example sample id scheme from OET includes: CruiseID_Sample#_Subsample_Preservation_Recipient, where Preservation is indicated by a standard code (e.g., a=70% ethanol, b=95% ethanol, c=dry). Naming convention should be described in a readily accessible document.

Sample Archiving

It is expected that samples collected during OECI activities will be transmitted to established repositories for archiving. Some samples may be distributed to the laboratories of individual scientists. If possible, those should be sub-samples and the parent sample should be archived. If this is not possible (e.g., samples are consumed during analysis), this should be noted in the metadata. The location and name of the archive and contact information for the archive curator should be documented in sample metadata.

Sample Description

Sample ID shall be linked to metadata that follows a standard vocabulary as defined by the System for Earth Sample Registration (Vocabularies | System for Earth Sample Registration. Metadata categories that are unknown or impossible to define, should be left blank. If possible, samples should be registered with an IGSN with an OECI suffix to ensure they are discoverable. A sample spreadsheet will be provided to assist in metadata generation and curation.

Sample Type

- **Core** - long cylindrical cores
- **Core Half Round**- half-cylindrical products of along-axis split of a whole round
- **Core Piece**- material occurring between unambiguous [as curated] breaks in recovery.
- **Core Quarter Round** - quarter-cylindrical products of along-axis split of a half round.
- **Core Section** - arbitrarily cut segments of a “core”
- **Core Section Half** - half-cylindrical products of along-axis split of a section or its component fragments through a selected diameter.
- **Core Sub-Piece**- unambiguously mated portion of a larger piece noted for curatorial management of the material.
- **Core Whole Round** - cylindrical segments of core or core section material.

- **Cuttings** - loose, coarse, unconsolidated material suspended in drilling fluid.
- **Dredge**- a group of rocks collected by dragging a dredge along the seafloor.
- **Experimental Specimen**- a synthetic material used during an experiment
- **Grab** - a sample (sometimes mechanically collected) from a deposit or area, not intended to be representative of the deposit or area.
- **Hole** - hole cavity and walls surrounding that cavity.
- **Individual Sample**- a sample that is an individual unit, including rock hand samples, a biological specimen, or a bottle of fluid.
- **Oriented Core** - core that can be positioned on the surface in the same way that it was arranged in the borehole before extraction.
- **Other** - a sample that does not fit any of the existing type designations. It is expected that further detailed description of the particular sample will be provided.
    - Profile Water Sample - water collected in a Niskin bottle via CTD cast or vehicle
    - Seawater System Water Sample - water collected from the shipboard continuous surface seawater system
    - Suction or Slurp - water collected via a vehicle suction system (e.g. hydrothermal fluid)

Descriptive Metadata

- **IGSN:**Leave blank if you want SESAR to assign the IGSN (recommended)
- **Parent IGSN:**Leave blank if a parent IGSN does not exist
- **Release Date:** Date when sample metadata is publicly accessible and searchable. If null, defaults to date of registration (recommended)

- **Material:** material that the sample consists of “Rock”, “Mineral”, “Liquid>aqueous” [list]
- **Field Name (informal classification):**Taxonomy (field name) Informal classification of sample e.g. basalt, amphibole, sea water
- **Classification:**Taxonomy (formal classification) Formal categorization of sample e.g. igneous, volcanic; IMS mineral name [rock classification] [mineral classification]
- **Sample Description:** Free text to describe features of a sample such as its components, texture, color, shape, etc. Ex: “dredge with 50 pieces of basalt and mud”; “euhedral specimen, variety: chiastolite”, “reference quality, medium crystal size.”
- **Other name(s):** Other name(s) used for the sample. Provide multiple names delimited by semi-colons.
- **Age (min):** Numerical value for the minimum age of a sample “4.2”
- **Age (max):** Numerical value for the maximum age of a sample “4.6”
- **Age unit:**Unit for the age provided “Ma” (for million years); “years”
- **Geological Age:** Age of a sample as described by the stratigraphic era, period, state, etc. “Cretaceous”; “Upper Miocene”
- **Collection method:** Method by which a sample was collected e.g. “dredging”; “Coring>PistonCorer” (http://www.geosamples.org/help/vocabularies#collection)
- **Collection Method Description:**Additional information about the collection method, e.g. if special equipment or procedures were used
- **Size:** Size of the registered object, such as the dimension of a specimen, the length of a core, or the weight of a dredge. e.g. “2x4” (enter”cm” into ‘size unit’); “45” (enter “kg” into ‘size unit’)
- **Size Unit** (i.e., cm, in, m, kg): Unit for the numerical value provided for ‘size’. e.g. “meters”; “kg”
- **Comment:**Any additional comment about the sample that does not fit into the existing fields.
- **Purpose**: The purpose for collecting the sample. e.g. “paleomagnetism”

Geolocation Metadata

- **Latitude:** Latitude of the location where the sample was collected. (‘Start latitude’ for linear sampling features such as dredges.) Needs to be entered in decimal degrees. Negative values for South latitudes. “-24.7852” (=24.7852 S); “5.89634” (=5.89634 N)
- **Longitude:** Longitude of the location where the sample was collected. (‘Start longitude’ for linear sampling features such as dredges.) Needs to be entered in decimal degrees. Negative values for ‘West’ longitudes. e.g. “-103.785 W); “68.9045” (=68.9045E)
- **Northing (m) (Coordinate system: UTM NAD83):**Geographic Cartesian coordinate of where the sample was collected, in meters, as in UTM. “4111279”
- **Easting (m) (Coordinate system: UTM NAD83):** Geographic Cartesian coordinate of where the sample was collected, in meters, as in UTM. “305294”
- **Zone (e.g., 11R):**UTM zone, use a number from 1 to 60, followed by a letter between A and Z to specify UTM zone with no spaces in between. Letter must be capitalized.

- **Elevation:** Elevation at which a sample was collected. Use negative values for depth below sea level. Minimum elevation if a range is provided for the elevation/depth below sea level. “678.5”; “-4536” (=4536 meters depth below sea level)
- **Elevation Unit:**Unit in which elevation start and/or end are provided in (must be either feet, meters, miles, or kilometers).
- **Latitude(end), Longitude(end), and Elevation(end)**are specified for Dredges.
- **Navigation Type:**The method used to determine the location of the sample, e.g. “GPS”, “DVL”. See the MGDS navigation type list .
- **Physiographic Feature:** Type of physical feature that your sample was collected from. e.g. “volcano”; “mid-ocean ridge”; “lake” [list]
- **Name of physiographic feature:** Name of the physiographic feature that you entered. “Mauna Loa”; “East Pacific Rise”; “Dead Sea”
- **Location description:** Free text description of the location
- **Locality:** Name of the specific place where your sample was collected. This could be the name of a mine, a volcanic field, a vent field, or similar. e.g. “Franklin Mine”; “Craters of the Moon National Park”; “MARK area”
- **Locality description**: Additional information about the specific place where your sample was collected.
- **Country:** Country where the sample was collected. (if applicable) e.g. “Sweden”; “United States” [list]

Collection Metadata

- **Field Program/Cruise**: Name or identifier of the field program (cruise or expedition), during which the sample was collected. (if applicable) e.g. “HLY0102”; “ICDP-05/09”; “ODP Leg 73”
- **Platform type:**Type of platform used for the collection of the sample. (if applicable). e.g. “Ship”; “Drill rig” [list]
- **Platform name:** Name of platform used for the collection of the sample. (if applicable). e.g. “R/V Roger Revelle”; “GLAD200”
- **Platform description:**Any further information about the platform.
- **Launch Type:** The type of launch used to collect the sample. e.g. “HOV”, “ROV”. See the MGDS launch type list .
- **Launch Platform Name:**The name of the launch used to collect the sample. e.g. “Jason II”, “Alvin”
- **Launch ID:** Further identifying information about the launch. e.g. for an Alvin Dive, the dive number “3969”
- **Collector/Chief Scientist:** Name of the person who collected the sample. In case of larger field programs, name of chief scientist responsible for sample collection. e.g. “John Smith”
- **Collector/Chief Scientist Address:** Institution, address, & email of the collector or chief scientist. e.g. “Woods Hole Oceanographic Institution, Woods Hole, MA; xxxxx@whoi.edu ”
- **Collection date:** Date when the sample was collected. YYYY-MM-DD

- **Collection time** (i.e., 13:00 GMT): Time when the sample was collected. Please provide GMT as hh:mm:ss. e.g. “04:33.54”
- **Collection date (end):** Date when the sample collection was finished (if a date range is provided). YYYY-MM-DD
- **Collection time (end):** Time when the sample collection was finished.
- **Collection Date Precision:** The temporal precision of the collection date and time (e.g. “second”, “minute”, “hour”, “day”, “month”, “year”)

Curation Metadata

- **Preservation Method:** e.g., 70% ethanol, dry
- **Current Archive:** Name of institution, museum, or repository where the sample is currently stored.
- **Current Archive Contact:**Address and/or email of the person who should be contacted for information about or access to the sample.

Sample Hierarchy Metadata

- **Depth in Core (min):** Minimum depth at which a sample was collected from its parent core.
- **Depth in Core (max):**Maximum depth at which a sample was collected from its parent core.
- **Depth scale**(i.e., MBSF, MCD): Unit in which the depth is provided. e.g. “Meters below seafloor”; “meters composite depth”

Appendix H: Vehicle navigation and sensor data

Introduction

This Appendix describes collection, processing, and archiving of data associated with vehicles that operate beyond the research vessel including ASVs, AUVs, ROVs, and UAVs. Some data types collected by these vehicles are described in other appendices (e.g., Acoustic Mapping, Video and Still Imagery). This Appendix focuses on data products including vehicle navigation, attitude, along-track oceanographic data (e.g., temperature, oxygen), and other along-track sensor data (e.g., pressure).

Data Types Collected

Data types will differ between platforms. Best practice is to include data from all available sensors. For every dive a file shall be created containing a list of all sensors installed on the vehicle. This file may include model, serial number and placement on the vehicles for each item listed as well as date of last calibration and link to calibration results or contact to obtain calibration. This information should be reviewed at the beginning of each expedition to ensure it is up to date and updated throughout the expedition as equipment is replaced.

Example sensor types (derived data in parens):

USBL range and bearing (position)
Compass (heading)
IMU (pitch, roll, yaw)
DVL (altitude)
Pressure (depth)
Conductivity (salinity), Temperature, pressure (depth)
Dissolved Oxygen (% saturation)
Magnetic intensity

Vehicle navigation and sensor data shall be archived at its native collection rate and a summary of data sampled at a minimum of 1Hz shall be created and archived.

Processing

Navigation

Navigation data may be in the form of GPS for surface vehicles, INS, USBL, or dead reckoning for submerged vehicles. In many cases, multiple positioning streams are merged and may be modified/corrected with the aid of sensor-derived data (e.g., SLAM). Navigation data should indicate the level of processing (e.g., level 0 = raw, level 1 = filtered, level 2 = merged and corrected) and the processing methods should be described.

Sensor Data

Units of measurement for individual sensors should be documented (e.g., ˚C, dB, voltage) and any processing of sensor data should be documented (e.g., methods of converting of pressure to depth).

Time

Time from a common time base should be recorded with each sensor reading in order to merge multiple data streams post-deployment. Times should be recorded as UTC at a fidelity commensurate with the sensor data rate.

Products

The following data products are recommended to be produced following each deployment and each cruise:

1. An ASCII data file with time and raw sensor data shall be produced.
2. An ASCII data file with navigation and sensor data sampled at 1Hz shall be produced.
3. Documentation of the sensor name, manufacturer, serial number, their placement, and calibration information for each deployment shall be provided in the Data Catalog (Appendix A).

Archive

Data products from vehicle deployments should be submitted to a publicly accessible permanent archive, MGDS.