



OCEAN  
OBSERVATORIES  
INITIATIVE

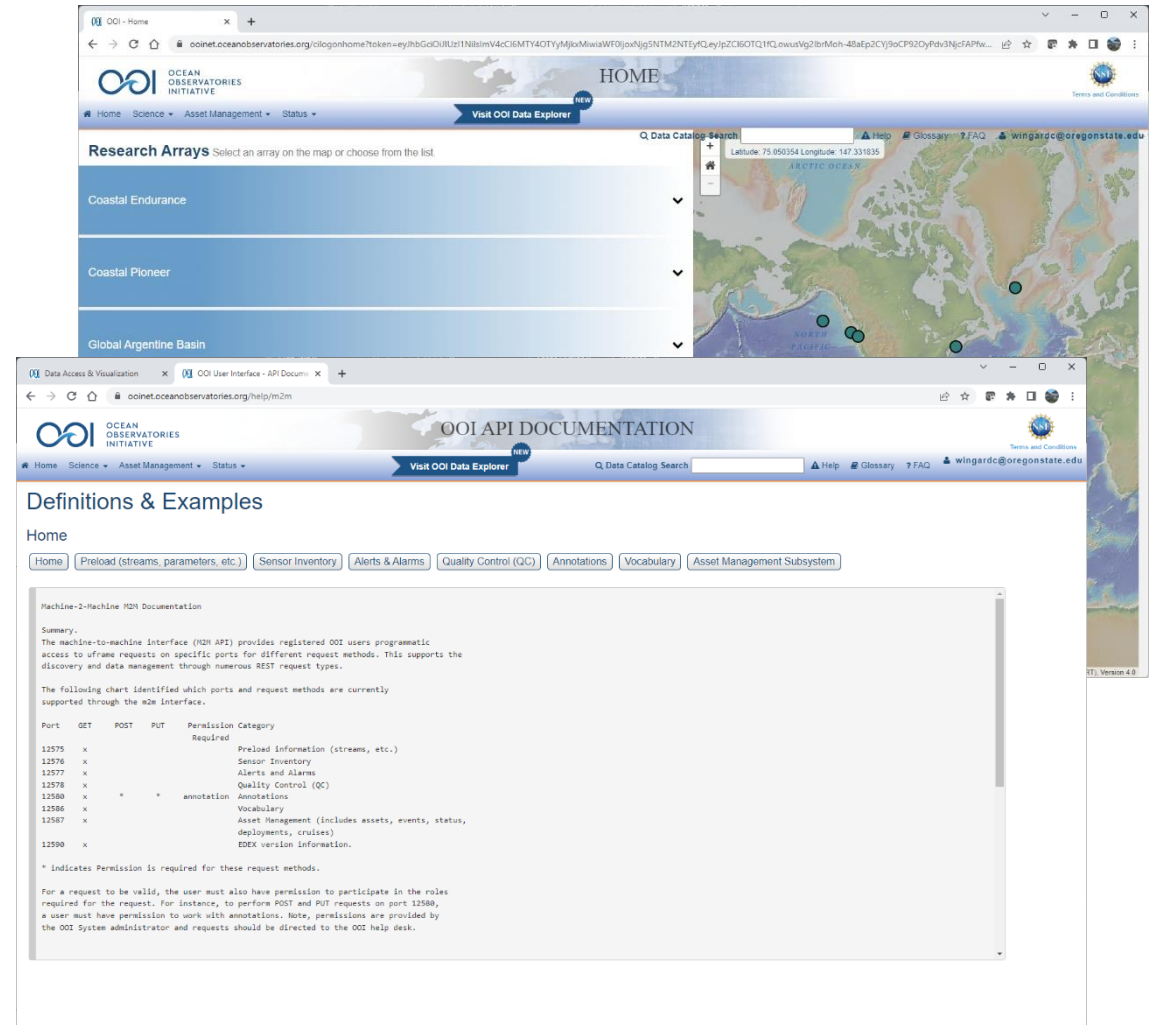
# OOI Data Resources

Wendi Ruef and Christopher Wingard  
Cabled and Endurance Array Data Leads  
July 18, 2023



# OOI Data Portal

- Original website for data access (<https://ooinet.oceanobservatories.org>)
- Slated for deprecation
- Currently the only mechanism for generating and viewing user API keys
- Will remain as the OOI M2M (machine-to-machine) API endpoint
  - Provides programmatic access to both metadata and instrument data (API credentials required)
  - <https://ooinet.oceanobservatories.org/help/m2m> (requires user log to view)
  - <https://oceanobservatories.org/m2m/>



# Terminology (Handout: Decoder)

## Array

- One of 5 major research regions that make up OOI (e.g., Global Station Papa (**GP**) or Coastal Endurance (**CE**))

## Site

- A specific geographic location within an array (e.g., Coastal Endurance Oregon Inshore = **CE01**)

## Platform

- An assembly/infrastructure at a site that hosts a complement of integrated scientific instruments. May be fixed (mooring) or mobile (profilers or gliders) (e.g., Coastal Endurance Oregon Inshore Surface Mooring = **CE01ISSM**)

## Node

- A section of a platform with one or more data loggers and power controllers. Instruments on platforms are plugged into nodes. (e.g., the data logger on the Near Surface Instrument Frame (NSIF, 7 m), part of the mooring riser, of the Coastal Endurance Oregon Inshore Surface Mooring = **CE01ISSM-RID16**)

## Instrument/Sensor

- Terms often used interchangeably. An instrument is a piece of equipment used to collect data. A sensor is a part of an instrument which measures a specific quantity. Each instrument has a Unique ID (e.g., the dissolved oxygen sensor on the NSIF of the Coastal Endurance Oregon Inshore Surface Mooring = **03-DOSTAD000**)

## Reference Designator

- Identifies a particular instrument on a particular node/platform at a particular site
- Example: **CE01ISSM-RID16-03-DOSTAD000** = Coastal Endurance Oregon Inshore Surface Mooring - Mooring Riser - Dissolved Oxygen Sensor

## Data Delivery Method

- **telemetered**: Data returned wirelessly. May be truncated or decimated due to size.
- **recovered\_host/recovered\_inst/recovered\_cspp/recovered\_wfp**: Data downloaded directly from either the platform computer and/or from the instrument after the system is recovered
- **streamed**: Data accessible in real-time, streamed over the fiber optic network from Cabled Array and select Endurance platforms

## Data Stream Name

- Data feed from a sensor that has been read, parsed, and separated based on content (e.g., “engineering”, “science”, “metadata”, etc.) into a named dataset, or “stream”. Stream names are often method specific

## Parameters

- Also sometimes called data variables
- A particular value returned from a sensor (e.g., practical salinity from a CTD). There are multiple parameters in a data stream, some of which may be identified as an OOI Data Product



# OOI Website

- <https://oceanobservatories.org/>
- <https://oceanobservatories.org/research-arrays/>
  - Use this page to navigate to the array of interest (e.g., Coastal Endurance Array)
  - From the array page, select a site of interest (e.g., Oregon Shelf Surface Mooring (CE02SHSM))
  - From the site page, select an instrument of interest (e.g., Bulk Meteorology Instrument Package (METBKA))
  - The instrument list will show the full reference designator for each instrument
  - Selecting an instrument will open that instrument's page in the Data Portal. Clicking on the plus sign in the data catalog will open a table above listing delivery methods and stream names

The top screenshot shows the 'The Arrays' page on the OOI website. It features a world map with various array locations marked, including the Coastal Endurance Array. A red circle highlights the 'Coastal Endurance Array' in the right-hand navigation menu.

The bottom screenshot shows the 'DATA ACCESS & VISUALIZATION' portal. It displays a data catalog for the Coastal Endurance - Oregon Shelf Surface Mooring - Surface Buoy - Bulk Meteorology Instrument Package (CE02SHSM-SBD11-06-METBKA000). A red circle highlights the plus sign in the 'Actions' column of the data catalog table.

Actions	Instrument	Stream Type	Stream Content	Delivery Method	Start	End	Stream
[+]	Bulk Meteorology Instrument Package	Science	Data Products	recovered-host	2015-04-02T20:41:42.030Z	2023-03-19T19:12:36.129Z	metbk_a_dci_instrument_recovered
[+]	Bulk Meteorology Instrument Package	Science	Flux Data Products Calculated From Hourly Averages	recovered-host	2015-04-02T20:41:42.030Z	2023-03-19T19:12:36.129Z	metbk_hourly
[+]	Bulk Meteorology Instrument Package	Science	Data Products	telemetered	2015-04-02T20:41:42.030Z	2023-07-16T18:36:50.039Z	metbk_a_dci_instrument
[+]	Bulk Meteorology Instrument Package	Science	Flux Data Products Calculated From Hourly Averages	telemetered	2015-04-02T20:41:42.030Z	2023-07-16T18:36:50.039Z	metbk_hourly



# Thematic Real-Time Environmental Distributed Data Services (THREDDs)

- M2M Requests (async)
  - User and request specific
  - Temporal organization
  - Limited retention time (currently 6 months, under review)
- Gold Copy THREDDs
  - <https://thredds.dataexplorer.oceanobservatories.org/thredds/catalog/ooigoldcopy/public/catalog.html>
  - Created by internal OOI processes, updated daily
  - Organized by reference designator, delivery method and stream
  - Covers most data sets provided by OOI through the M2M system (exceptions are “engineering” or “metadata” streams)
  - Eliminates the need for most M2M requests
  - ***Basis for the data served via Data Explorer***





# Environmental Research Division's Data Access Program (ERDDAP)

- Gold Copy ERDDAP
  - <https://erddap-goldcopy.dataexplorer.oceanobservatories.org/erddap/index.html>
  - netCDF files from the Gold Copy THREDDS catalogs loaded into ERDDAP datasets
  - Datasets are organized by reference designator, delivery method, stream and deployment number
- Merged ERDDAP (backend to Data Explorer)
  - <https://erddap.dataexplorer.oceanobservatories.org/erddap/index.html>
  - Data from different data delivery methods (and streams) are combined
  - Select parameters available:
    - CF Standard Name (e.g., moles\_of\_oxygen\_per\_unit\_mass\_in\_sea\_water)
    - OOI Data Product (e.g. DOCONCS)
  - High resolution data resampled to max 1 minute resolution
  - Overlapping deployments are combined to create a single timeseries
  - Not fully 1:1 with the Gold Copy THREDDS catalogs. Some of the more complex data sets are still being worked on

ERDDAP - Home Page

erddap-goldcopy.dataexplorer.oceanobservatories.org/erddap/index.html

OCEAN OBSERVATORIES INITIATIVE

Brought to you by NSF

## ERDDAP

ERDDAP (the Environmental Research Division's Data Access Program) is a data server that gives you a simple, consistent way to download subsets of scientific datasets in common file formats and make graphs and maps. This particular ERDDAP installation has oceanographic data (for example, data from satellites and buoys).

### Easier Access to Scientific Data

Our focus is on making it easier for you to get scientific data.

**Different scientific communities have developed different types of data servers.**

For example, OPeNDAP, WCS, SOS, OBIS, and countless custom web pages with forms. Each is great on its own. But without ERDDAP, it is difficult to get data from different types of servers:

- Different data servers make you format your data request in different ways.
- Different data servers return data in different formats, usually not the common file format that you want.
- Different datasets use different formats for time data, so the results are hard to compare.

**ERDDAP unifies the different types of data servers so you have a consistent way to get the data you want, in the format you want.**

- ERDDAP acts as a middleman between you and various remote data servers. When you request data from ERDDAP, ERDDAP reformats the request into the format required by the remote server, sends the request to the remote server, gets the data, reformats the data into the format that you requested, and sends the data to you. You no longer have to go to different data servers to get data from different datasets.
- ERDDAP offers an easy-to-use, consistent way to request data via the OPeNDAP

### Start Using ERDDAP: Search for Interesting Datasets

- Do a Full Text Search for Datasets
- View a List of All 8,582 Datasets
- Search for Datasets by Category
 

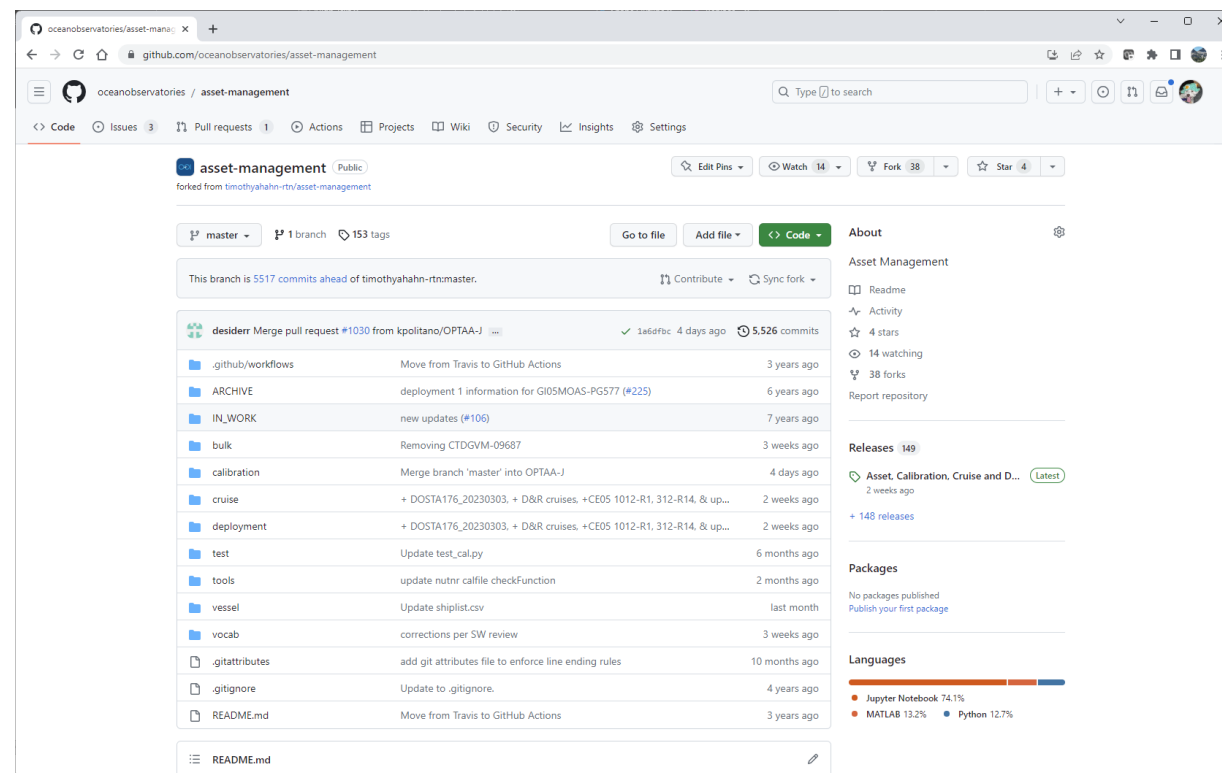
Datasets can be categorized in different ways by the values of various metadata attributes. Click on an attribute (cdm\_data\_type, institution, iaos\_category, keywords, long\_name, standard\_name, variableName) to see a list of categories (values) for that attribute. Then, you can click on a category to see a list of relevant datasets.
- Search for Datasets with Advanced Search
- Search for Datasets by Protocol
 

Protocols are the standards which specify how to request data. Different protocols are appropriate for different types of data and for different client applications.



# OOI Asset Management System

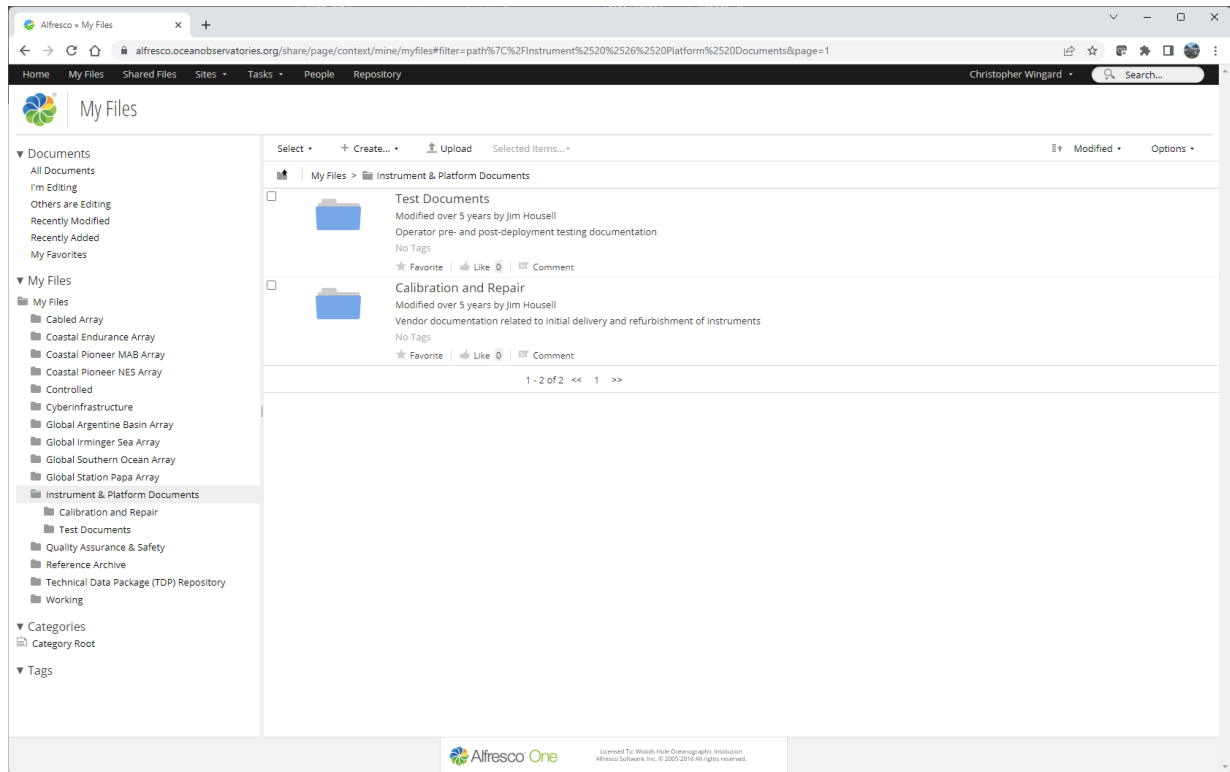
- OOI database system for storing platform configurations, serial numbers deployment information (dates, locations), instrument calibration values, etc.
  - Critical metadata (who, what, where, when)
- GitHub
  - <https://github.com/oceanobservatories/asset-management>
  - Source files for the database in CSV format
  - Includes tools used to create and check calibration values and configurations
  - Information generated and checked by personnel at the MIOs
  - Updated frequently, with updates pushed into the Asset Management database as required
- M2M API
  - Database accessible to users via the OOI M2M API





# OOI Document Management (Alfresco)

- OOI document management system utilizing Hyland's Alfresco software application
  - <https://alfresco.oceanobservatories.org/share/page/>
  - Read-only access via username/password guest
- Instrument & Platform Documents
  - Vendor service and calibration documents under Calibration and Repair
  - Pre-deployment pure-water calibrations under Test Documents/Instruments
- Discrete Samples
  - Organized in folders according to array and deployment cruise names (e.g., My Files > **Coastal Endurance Array** > Cruise Data > **Endurance-18\_SKQ202304S\_2023-03** > Ship\_Data > Water Sampling
  - Samples collected during the deployment and recovery of assets. Highly useful in validating certain parameters.





# OOI Data Explorations

- Developed by OOI Data Team members to facilitate access to different instrument datasets (data reviews, developing QARTOD test limits, conducting deep-dives, etc.)
- Code is freely available to the public via <https://github.com/oceanobservatories/ooi-data-explorations>
- Primarily written in Python with MATLAB and limited R utilities available
  - Key functions for participants in this workshop:
    - `load_kdata`: both MATLAB and Python versions available. Will load data from the `kdata` directory accessible via the OOI JupyterHub.
    - `load_gc_thredds`: both MATLAB and Python versions available. Will load data from the OOI Gold Copy THREDDS catalog
- OPTAA processing functions are currently only available in Python. Core set of utilities can be found in: [https://github.com/oceanobservatories/ooi-data-explorations/blob/master/python/ooi\\_data\\_explorations/uncabled/utilities\\_optaa.py](https://github.com/oceanobservatories/ooi-data-explorations/blob/master/python/ooi_data_explorations/uncabled/utilities_optaa.py)
- Platform specific processing functions are:
  - `cabled/process_optaa.py`
    - `optaa_benthic`: will process data from instruments mounted on the Benthic Experiment Package (burst averaged to 1-hour resolution)
    - `optaa_profiler`: will process data from instruments mounted on the cabled shallow profilers (profiles will be binned into 25 cm depth bins)
  - `uncabled/process_optaa.py`
    - `optaa_cspp`: will process data from instruments mounted on the CSPPs (profiles will be binned into 25 cm depth bins)
    - `optaa_datalogger`: will process data from instruments mounted on the uncabled, fixed depth platforms (burst averaged to ~1-hour resolution)





# Additional Resources

- In-line systems with inherent optical sensors community resource page ([University of Maine](#))
  - MATLAB code for inline systems; covers processing AC-S data among the many other types of optical sensors
  - <https://github.com/OceanOptics/InLineAnalysis>
- **acsPROCESS\_INTERACTIVE**
  - MATLAB code for processing AC-S data following NASA's SEABASS submission standards
  - [https://github.com/JesseBausell/acsPROCESS\\_INTERACTIVE](https://github.com/JesseBausell/acsPROCESS_INTERACTIVE)
- The International Ocean-Colour Coordinating Group (IOCCG) (<https://ioccg.org/>)
  - Community Practices and Protocols through the Ocean Best Practices repository (<https://repository.oceanbestpractices.org/handle/11329/510>)





OCEAN  
OBSERVATORIES  
INITIATIVE

# Questions?

