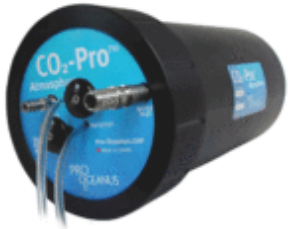
The Ocean Observatories Initiative (OOI) Approach to Data and Metadata

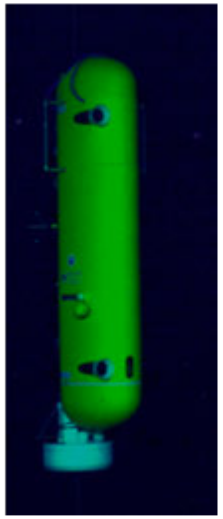
- Wendi Ruef, University of Washington
- Al Plueddemann, Woods Hole Oceanographic Inst.
- Ed Dever, Oregon State University

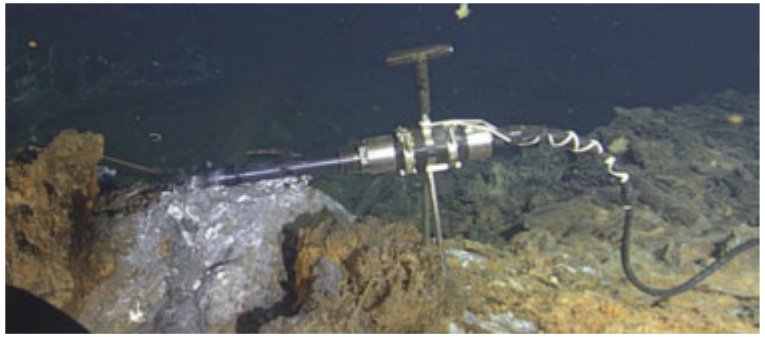
NSF EarthCube Workshop for Ocean Time Series Data
September 13-15, 2019 (Univ. Hawai'i)

Types and Numbers of Instruments in OOI

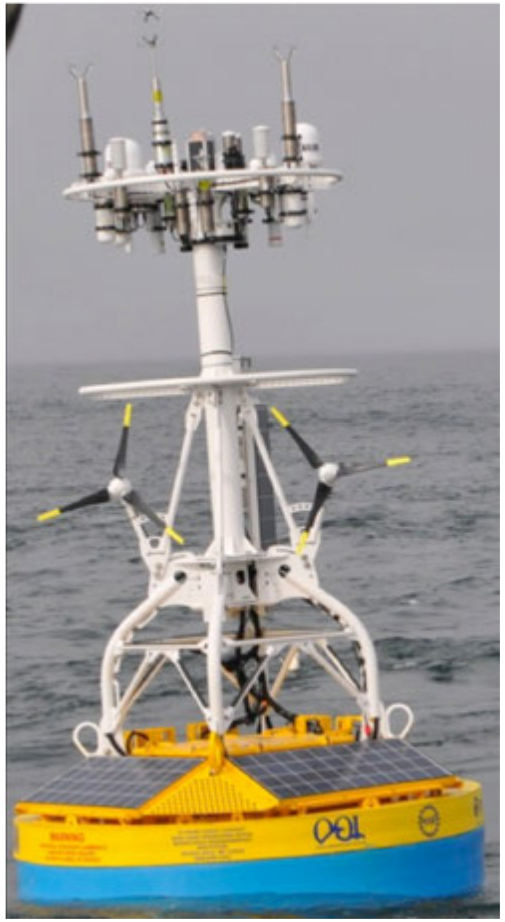
- Instruments and sampling derived from traceability matrices based on OOI science goals
- ~75 instrument models and 200 data products
- Over 800 instruments total
- The diversity of instrument types, platforms and data delivery methods present unique challenges.







Infrastructure



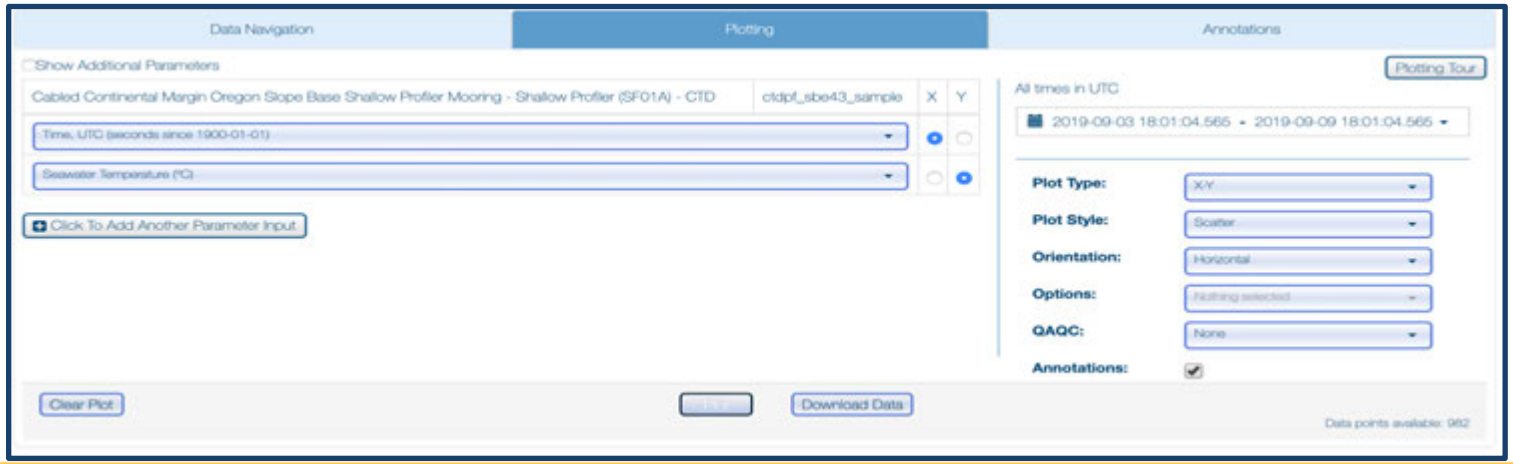
- Turn cycles
  - Regional Cabled and Global Scale Arrays: once per year.
  - Coastal Pioneer and Endurance Arrays: twice per year
  - Coastal gliders and surface profilers: nominal three month cycle
  - Pioneer AUV's: deployed on an expeditionary basis.
- Sparing
  - Two sets of instruments and platforms – one deployed and one being refurbished.
- Instrument preparation
  - Most instruments calibrated by manufacturers; testing and validation by OOI

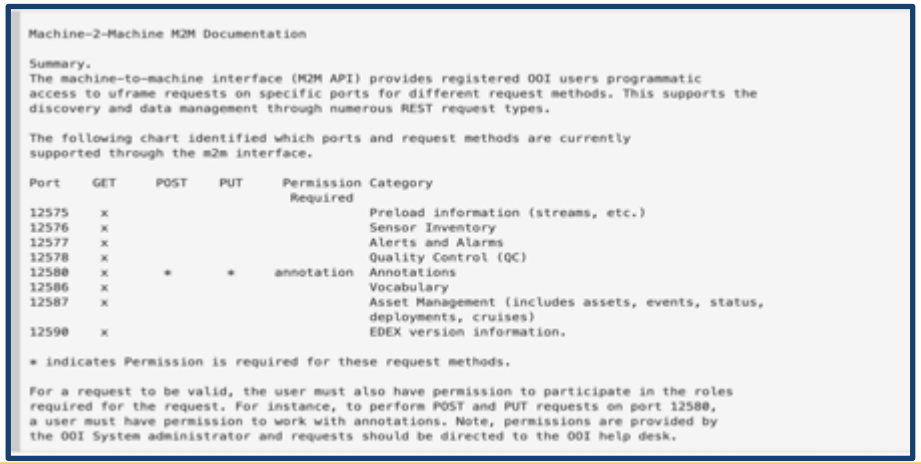
System and Instrument Metadata

oceanobservatories.org	• Raw data archive • Descriptive pages for instruments and platforms • Community tools for data processing and plotting
GitHub Repos	• Configuration sheets for each deployment • Calibration coefficients used in data processing • Raw data parsers and data product algorithms
OOINet	• HITL annotations of data streams • Instrument and deployment metadata • Data product access and downloads
Alfresco	• SOPs, cruise plans, shipboard data and samples • Sensor vendor documentation, calibration history • Program documentation

OOI Data Products

- Unique data delivery models with on demand data processing based on AWIPS (the Advanced Weather Interactive Processing System) display, and analysis package developed by the National Weather Service and Raytheon
- Data product algorithms used to convert raw data into scientific units written in Python from code provided by manufacturers
- Raw and processed data download through GUI and M2M





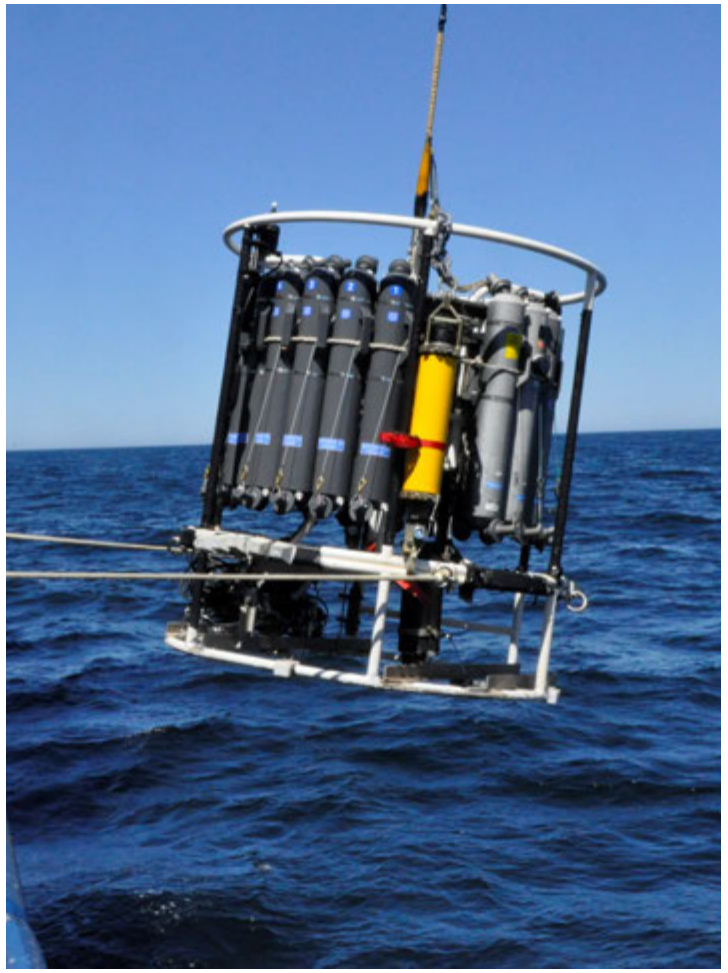
```
Machine-2-Machine M2M Documentation

Summary.
The machine-to-machine interface (M2M API) provides registered OOI users programmatic
access to uframe requests on specific ports for different request methods. This supports the
discovery and data management through numerous REST request types.

The following chart identified which ports and request methods are currently
supported through the m2m interface.

Port    GET     POST    PUT     Permission Category
                                 Required
12575   x                       Preload information (streams, etc.)
12576   x                       Sensor Inventory
12577   x                       Alerts and Alarms
12578   x                       Quality Control (QC)
12580   x       *       *       annotation  Annotations
12586   x                       Vocabulary
12587   x                       Asset Management (includes assets, events, status,
                                deployments, cruises)
12590   x                       EDEX version information.

* indicates Permission is required for these request methods.

For a request to be valid, the user must also have permission to participate in the roles
required for the request. For instance, to perform POST and PUT requests on port 12580,
a user must have permission to work with annotations. Note, permissions are provided by
the OOI System administrator and requests should be directed to the OOI help desk.
```

OOI Data Products

- Physical sampling records, cruise plans and reports on the OOI document management system (Alfresco)
- Compilation of shipboard and physical samples into common format
  - Parse-able and human-readable (CSV) data summary for each OOI cruise
  - Bottle and discrete data assigned individual QC data flags (based on WOCE)
  - Summary posted along with sampling SOPs, summary README, all original data logs, cast files, and sample sheets

Data Validation and Physical Sampling

- Overlap old and new deployments where possible
- Underway shipboard measurements (e.g. met, T-Sal, ADCP)
- CTD profiles with bottles for chemical analyses (salt, O2, Chl, nutrients, carbon)
- Bottle data used to verify/calibrate CTD sensors (when possible) and deployed assets
- Validated CTD traces used to validate deployed assets

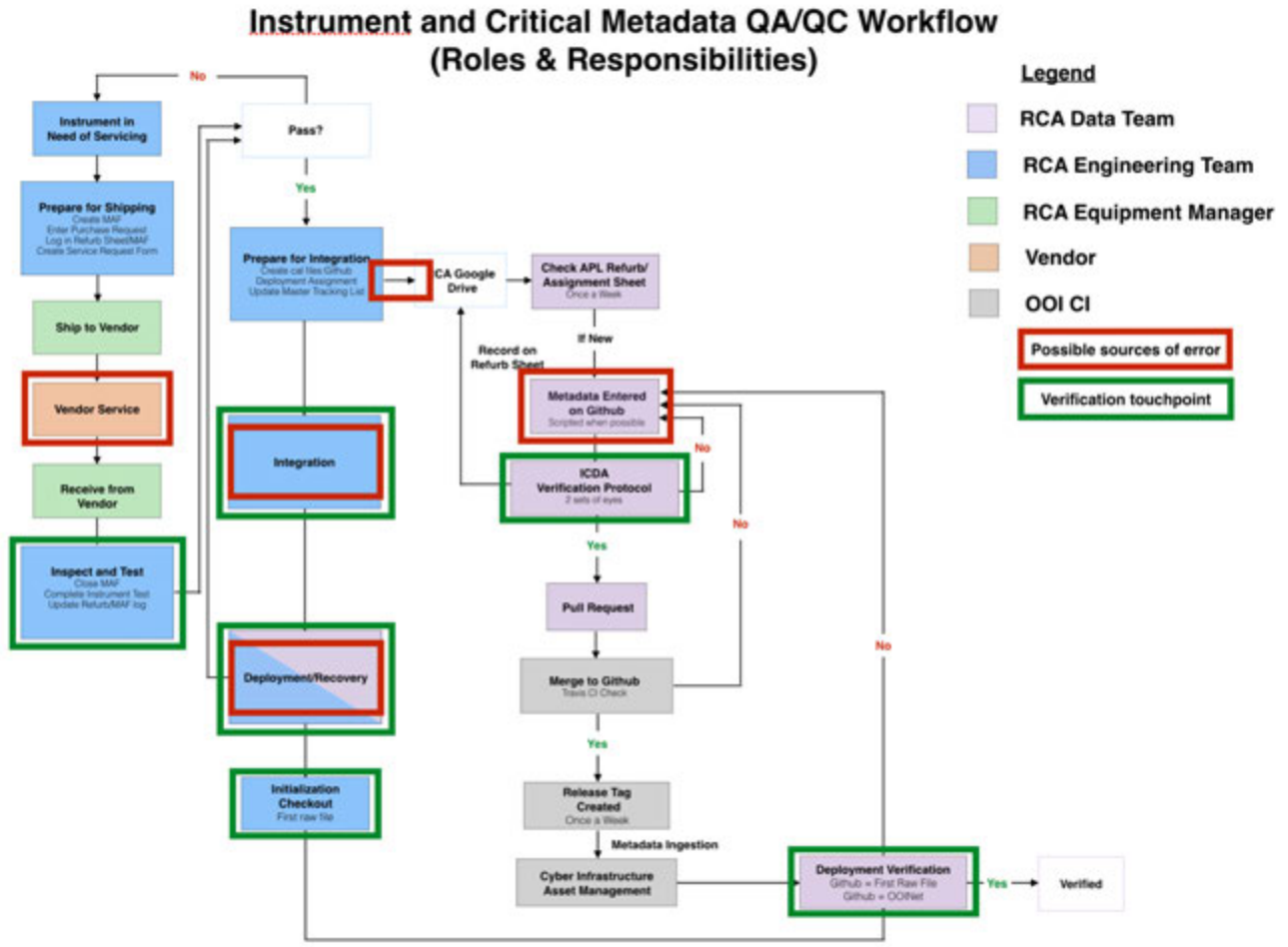


QA/QC Procedures:

- Quality conformance testing of incoming sensors
  - Basic functional checks on instruments are performed upon receipt.
  - Non-conforming instruments are returned to the vendors.
  - Additional calibration procedures may be performed in house
- Follow community practices for biofouling mitigation
  - Manufacturer-provided mitigation, plus copper tape, zinc ointment, etc
  - Recently implemented third party UV light mitigation where needed (O2 optodes, spectral irradiance, and digital still cameras).
- Integration and burn in before deployment
  - Instruments integrated onto platforms and tested to greatest extent possible before deployment

QA/QC Procedures:

- Comprehensive 2i-HITL review of metadata (ongoing)
- Verification documentation to increase user confidence in metadata accuracy (in process)
- Ongoing implementation of QARTOD standards
- Review by operators of data and annotation of performance issues
- Active ticketing and helpdesk system to track data and system issues



FAIR Data Principles

- Findable: *Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier*

  - OOI:
    - All data associated with appropriate metadata that contains all key knowledge about the data record
    - Metadata included in header of downloaded NetCDF files as well as Asset Management tables in OOI Net data portal

  - Challenges:
    - On demand processing makes versioning of data and creation of DOIs difficult
    - Inconsistent parameter naming makes data discovery harder

FAIR Data Principles

- Accessible: *Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.*

  - OOI:
    - Manual plotting and netCDF/csv data download through a GUI
    - Machine to machine (M2M RESTful API) interface and scripts for on demand processing
    - Select datasets served via other data provisions (e.g., raw data, ERDDAP server, IRIS, glider DAC, NANOOS, NDBC, GOA-ON, R2R)

  - Challenges:
    - Real-time data, on-demand processing, large datasets difficult to serve to outside repositories
    - Incomplete documentation of M2M

FAIR Data Principles

- Interoperable: *Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.*

  - OOI:
    - Metadata follows CF 1.6 standard, with additional metadata types and fields specific to OOI as necessary

  - Challenges:
    - Current CF standard is 1.7
    - some OOI science parameters do not have assigned standard names (in violation of CF standards)
    - Moving towards more commonly used ISO standards

FAIR Data Principles

- Reusable: *Data and collections have a clear usage license and provide accurate information on provenance*

  - OOI:
    - OOI Data Usage Policy available online: https://oceanobservatories.org/usage-policy/
    - All OOI data are publicly available for usage in proposals and the scholarly record, free registration for data download
    - Provenance and annotations included with data downloads (user opt-in)

  - Challenges:
    - Notifying users of changes in algorithms, provenance and data products
    - Implementing automated QC as part of on-demand processing makes it difficult to track and communicate QC statistics.

Acknowledgements