



OCEAN  
OBSERVATORIES  
INITIATIVE

# Introduction to Curated Dataset 1 (and 2)

Chris Wingard, Endurance Array Data Lead

2023-07-17

OOIFB Summer 2023 Workshop





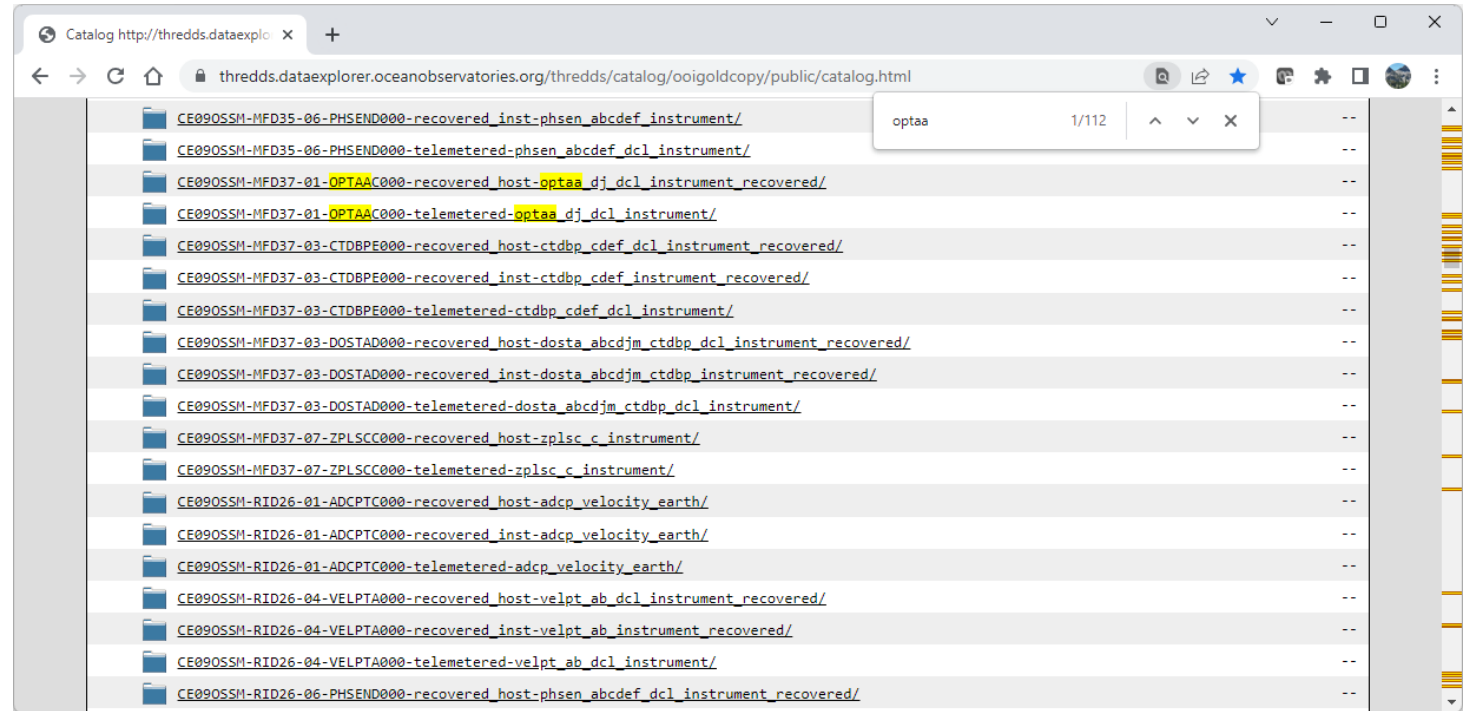
# Curation Defined

- “Data curation is the process of creating, **organizing** and maintaining data sets so they can be accessed and used by people looking for information. It involves collecting, **structuring**, indexing and cataloging data for users in an organization, group or the general public. Data can be curated to **support** business decision-making, academic needs, scientific research and other purposes.”  
(<https://www.techtarget.com/searchbusinessanalytics/definition/data-curation>)
- “Data curation includes ‘all the processes needed for principled and controlled data creation, maintenance, and management, together with the capacity to **add value** to data’”. ([https://en.wikipedia.org/wiki/Data\\_curation](https://en.wikipedia.org/wiki/Data_curation))
- Or restated: data curation is the process of organizing and structuring (and potentially extending) data in support of an operational goal.



# Processing Goals

- Organizing (collecting multiple data sources for subsequent analysis)
- Structuring (ensuring consistency across data sets)
- Supporting (HITL assessments, annotations and automated QC testing)
- Value-Added (adding intermediary data products, cross-validation, pure-water calibrations, cross-deployment comparisons)

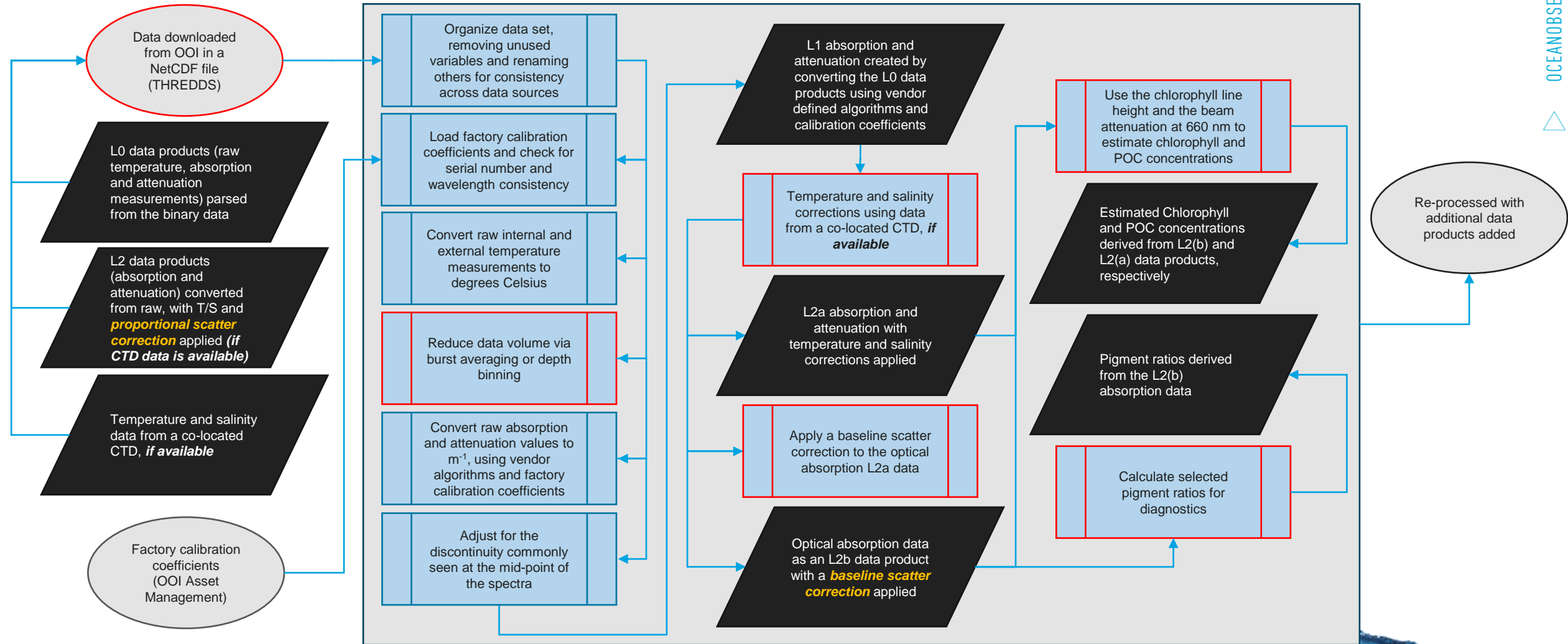


<https://github.com/oceanobservatories/ooi-data-explorations>  
See python/uncabled/process\_optaa.py and utilities\_optaa.py, and python/cabled/process\_optaa.py





# Processing Workflow



<https://github.com/oceanobservatories/ooi-data-explorations>  
 See python/uncabled/process\_optaa.py and utilities\_optaa.py, and python/cabled/process\_optaa.py



# Structure of the Curated Data

- Raw variables (L0)
  - Raw signal, reference and temperature measurements (counts)
- OOI Processed Variables (L2)
  - The optical\_absorption and beam\_attenuation arrays with temperature and salinity corrections applied and a proportional scatter correction applied to the absorption measurements ( $m^{-1}$ , NaN-filled if co-located CTD data is not available)
  - The a and c channel wavelength arrays (nm)
  - Temperature ( $^{\circ}C$ ) and salinity (psu) from the co-located CTD (NaN filled if not available)
- Re-processed Variables (L1, L2a and L2b)
  - Converted temperature sensor measurements ( $^{\circ}C$ )
  - Converted (L1) absorption (apg) and attenuation (cpg) measurements using factory calibration coefficients ( $m^{-1}$ )
  - Temperature and salinity corrected (L2a) absorption (apg\_ts) and attenuation (cpg\_ts) values ( $m^{-1}$ )
  - Scatter corrected (L2b) absorption (apg\_ts\_s) using the baseline correction method ( $m^{-1}$ )
- Additional Variables
  - Estimated chlorophyll and POC concentrations and selected pigment ratios
- Factory Calibration Coefficients (json formatted file)

/home/jovyan/acs\_workshop/data/reprocess/ce02shsm/nsif/optaa

The screenshot shows a Panoply window titled "Sources" with a menu bar (File, Edit, View, History, Bookmarks, Plot, Window, Help) and toolbar (Create Plot, Combine Plot, Open, Remove, Remove All, Hide Info). Below the toolbar are tabs for "Datasets", "Catalogs", and "Bookmarks".

Name	Long Name	Type
ce02shsm.nsif.optaa.deploy03.reprocess...	Data produced by Stream Engine version 1.20.6 for...	Local File
a_jump_offsets	Absorption Channel Holographic Grater Jump Offset	ID
a_reference	Absorption Channel Reference Measurements	2D
a_reference_dark	Absorption Channel Dark Reference	ID
a_signal	Absorption Channel Signal Measurements	2D
a_signal_dark	Absorption Channel Dark Signal	ID
apg	Particulate and Dissolved Absorbance	2D
apg_ts	Particulate and Dissolved Absorbance with TS Corre...	2D
apg_ts_s	Particulate and Dissolved Absorbance with TS and S...	2D
beam_attenuation	Optical Beam Attenuation Coefficient	2D
c_jump_offsets	Attenuation Channel Filter Offsets	ID
c_reference	Attenuation Channel Reference Measurements	2D
c_reference_dark	Attenuation Channel Dark Reference	ID
c_signal	Attenuation Channel Signal Measurements	2D
c_signal_dark	Attenuation Channel Dark Signal	ID
cpg	Particulate and Dissolved Attenuation	2D
cpg_ts	Particulate and Dissolved Attenuation with TS Corre...	2D
deployment	deployment	ID
elapsed_run_time	Elapsed Run Time	ID
estimated_chlorophyll	Estimated Chlorophyll Concentration	ID
estimated_poc	Estimated POC Concentration	ID
external_temp	External Instrument Temperature	ID
external_temp_raw	Raw In-Situ Temperature	ID
internal_temp	Internal Instrument Temperature	ID
internal_temp_raw	Raw Internal Instrument Temperature	ID
optical_absorption	Optical Absorption Coefficient	2D
ratio_carotenoids	Carotenoid to Chlorophyll Absorbance Ratio	ID
ratio_cdcom	CDOM to Chlorophyll Absorbance Ratio	ID
ratio_phycobins	Phycobins to Chlorophyll Absorbance Ratio	ID
ratio_qband	Chlorophyll Q Band to Soret Band Absorbance Ratio	ID
sea_water_practical_salinity	Practical Salinity	ID
sea_water_temperature	Seawater Temperature	ID
time	time	ID
wavelength_a	Absorption Channel Wavelengths	2D
wavelength_c	Attenuation Channel Wavelengths	2D
wavelength_number	Wavelength Number	ID

File "ce02shsm.nsif.optaa.deploy03.reprocessed.nc"  
File type: Hierarchical Data Format, version 5

```
netcdf file:/C:/Users/wingardc/ooidata/m2m/47B47D/47B47D/47B47D/c
dimensions:
  time = 1274;
  wavelength_number = 100;
variables:
  int external_temp_raw(time=1274);
  :_FillValue = 65535; // int
  :comment = "Raw measurements, reported in counts, from the
  :precision = 0; // int
  :coordinates = "time lat lon";
  :long_name = "Raw In-Situ Temperature";
  :units = "count";
  :data_product_identifier = "OPTTEMP_L0";

double elapsed_run_time(time=1274);
:long_name = "Elapsed Run Time";
:units = "ms";
:coordinates = "time lat lon";
:_FillValue = 0.0; // double
:comment = "Time in milliseconds since the instrument was p
:precision = 0; // int

int internal_temp_raw(time=1274);
:_FillValue = 65535; // int
:comment = "Raw measurements, reported in counts, from the
:long_name = "Raw Internal Instrument Temperature";
:precision = 0; // int
:coordinates = "time lat lon";
:data_product_identifier = "OPTTEMP_L0";
:units = "count";

int c_reference_dark(time=1274);
:_FillValue = 65535; // int
:comment = "Attenuation channel reference detector dark cou
:precision = 0; // int
:coordinates = "time lat lon";
```





# Curated Dataset 1 (and 2)

- Dataset 1

- Oregon Shelf Surface Mooring (CE02SHSM), midwater platform (Near-Surface Instrument Frame, NSIF) at 7 m (site depth 80 m)
- Data from Deployment 3 (Endurance 5), spanning 2016-05-17 to 2016-07-09 (instrument failed shortly thereafter)
- Simple example, with reasonable data for Homework 1. Processing workflow is hard-coded, but users can alter the dataset loaded to explore other sites/deployments
- [https://nbviewer.org/github/oceanobservatories/ooi-data-explorations/blob/master/python/examples/notebooks/optaa/process\\_kdata\\_optaa.ipynb](https://nbviewer.org/github/oceanobservatories/ooi-data-explorations/blob/master/python/examples/notebooks/optaa/process_kdata_optaa.ipynb)

- Dataset 2

- Oregon Shelf Coastal Surface-Piercing Profiler (CSPP, CE02SHSP)
- Data from Deployment 19, spanning 2021-04-06 to 2021-04-29 (76 profiles from ~70 m to the surface)
- Processing workflow is broken apart to show the different steps involved offering users an example for how they could reprocess the data themselves, altering decision points to suit their own needs
- Includes additional steps using pre-deployment pure-water calibrations to correct the shape and magnitudes of the spectra
- [https://nbviewer.org/github/oceanobservatories/ooi-data-explorations/blob/master/python/examples/notebooks/optaa/processing\\_workflow\\_using\\_the\\_cspp\\_optaa.ipynb](https://nbviewer.org/github/oceanobservatories/ooi-data-explorations/blob/master/python/examples/notebooks/optaa/processing_workflow_using_the_cspp_optaa.ipynb)







OCEAN  
OBSERVATORIES  
INITIATIVE

# Questions?

