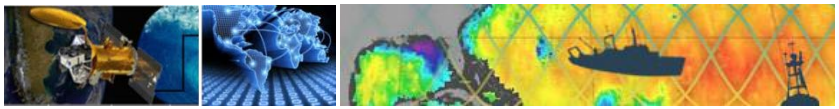


Addressing the Challenges of Satellite and In-situ Oceanographic Dataset interoperability at the PO.DAAC

V. Tsontos & T. Huang

PO.DAAC, NASA/JPL

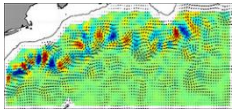
vardis.m.tsontos@jpl.nasa.gov

AGU Conference 2014

© 2014 California Institute of Technology. Government sponsorship acknowledged

Motivation (1/3) - General Considerations

- The earth science enterprise increasingly relies on the **integration** & **synthesis** of **multivariate datasets** from **diverse observational platforms**
- Applications:
  - Science
  - Decision Support
  - Mission Cal/Val



- Increased availability of oceanographic data online and improved access protocols & tools for access
- Outstanding Issues from the user perspective:
  - Inherent complexity/heterogeneity of data
  - Disparate, distributed systems that serve them
  - Overly complex process
  - Need more unified/seamless data access
- Problem Origin:
  - Technical issues: variable adherence to emerging metadata/file format standards, protocols
  - Structural/historical reasons: domain specific data providers
- PO.DAAC's Vision to help address these issues & drive the necessary technical innovation:
  - Traditionally NASA archive for satellite remote sensing physical oceanographic data
  - Additionally support NASA-funded field campaign data

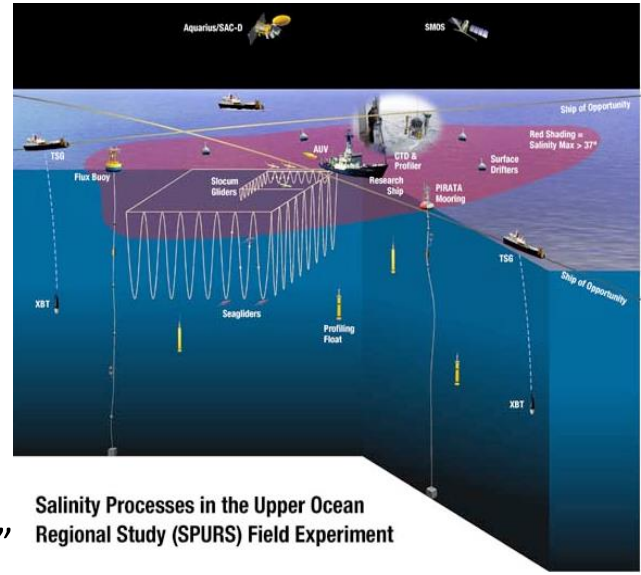
Motivation (2/3) : Aquarius and the SPURS Field Program

- Aquarius-SAC/D NASA salinity mission: http://aquarius.nasa.gov
  - global salinity observations: improved understanding of water cycle
  - 3 years of data archived & distributed by PO.DAAC

- SPURS: NASA-funded salinity oceanographic field campaigns: http://spurs.jpl.nasa.gov
  - Inter-agency & International effort/support
  - SPURS I: salinity max region N. Atlantic (5 cruises 2012-13):
  - SPURS II: high rainfall ITCZ area E. Tropical Pacific (2015-16)

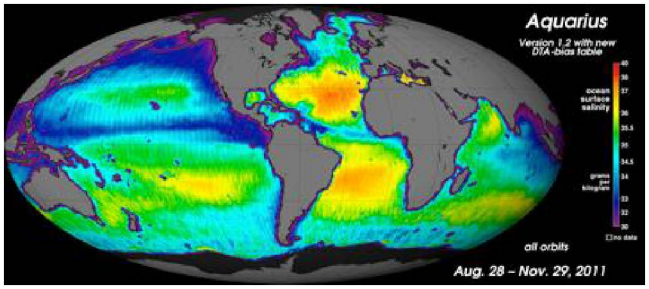
- Objectives:
  - salinity process study
  - synergies with Aquarius: sub-footprint SSS variability, water column structure & surface salinity field characterization

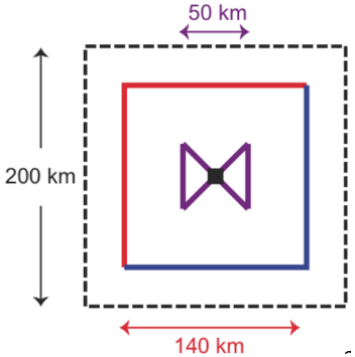
- Sampling:
  - Adaptive survey design: feature tracking, ROMS model
  - Range (novel) platforms & instruments: Multi-scale "Sensor-Web"

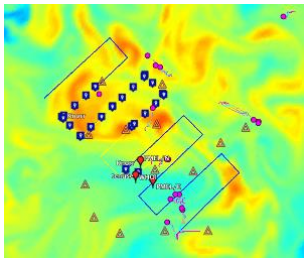


Salinity Processes in the Upper Ocean Regional Study (SPURS) Field Experiment

Aquarius/SAC-D







Motivation (3/3) : SPURS Sampling & Data

- SPURS I Sampling Platforms & Sensors by Project Activity:
  - ❑ Near surface salinity gradients: **Surface Salinity Profiler** (towed CTD chain 0.5-2m)
  - ❑ Horizontal surface T/S advection & eddy fluxes: 174 **Drifters SSS & ADCP current velocity**
  - ❑ **SPURS Mooring**: surface Met/Radiation Package,<5m CDT HR-currents,0-90m CTD/ADCP chain
  - ❑ Multi-scale Upper ocean salinity characterization: **AUVs & Wave Gliders**
  - ❑ Sustained Observations: **XBT transects, profiling floats, TSG, Pirata Mooring, CTD casts & underway, drifters**
  - ❑ Thalassa Cruise (FR) & Midas Cruise (SP):: **Drifters, CTD, ADCP, underway sensors, SEASOR, ASIP profiler**
  - ❑ **Profiling floats**: CTD, Wind, Rain
  - ❑ 3 **Gliders**: 2x 6mo deployments, 200x200km2 area, upper 1000m : CTD
  - ❑ 2 **microstructure gliders**: 20 and 50 km areas
  - ❑ Upper **ocean profiling & drifts with Floats**: CTD, Pressure, Radiation, Ambient Noise

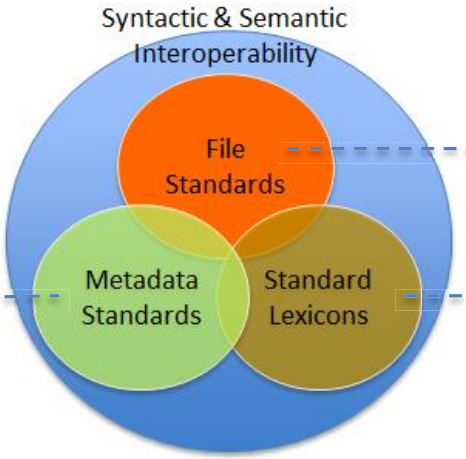
- SPURS I Data: Diverse & heterogeneous
  - Initially in native "unstructured" file formats & metadata *(no requirements on PIs)*
  - Approx. 15 datasets overall (platform/instrument)
  - 750mb, 195 data files for 13 core data types
  - Outstanding Datasets: salinity microstructure, modified ARGO float, SPURS moorings data

- SPURS archival request to PODAAC: initial 2011/12 & 2014
  - **Lack of file/metadata standards : a critical archival constraint & central interoperability challenge**
  - Determines: extent data discoverability & tool/service integration

Interoperability

The ability of coupled systems to:

- Communicate and exchange data via common formats & protocols
- Meaningfully interpret and reproducibly act on exchanged data



- Self-describing file formats
  eg. netcDF, HDF
  NODC Feature Class .nc templates

Metadata models •

eg. Climate Forecast (CF)
Marine Metadata MMI
WaterML (OGC/CUAHSI)
SensorML (OGC)

- Common vocabulary & semantic relationships
  eg. CF Standard Names/Units
  GCMD keywords
  SWEET ontology
  MMI Vocab & Ontology Repository

<u>Key interoperability challenges for Field Datasets</u>: exemplified by SPURS

- Inherent diversity, complexity, heterogeneity
- Lack of adherence to metadata & data format standards that are particularly acute for *in-situ* data
- Insufficient technical capacity, resources, interest by data producers

Global Change Master Directory
Discover Earth science data and services
OGC®
Marine Metadata Interoperability

Towards a PO.DAAC *in-situ* Support Capacity

- Blueprint developed for a ROSES 2013/ACCESS A36 Proposal

  F.Bingham (SPURS/UNCW), V.Tsontos (JPL), T.Huang (JPL), S.Arms (Unidata), J.Vazquez (JPL), D. Seidof (NOAA), K.Baynes (NASA/EOSDIS)

  - ✓ Practical/evolutionary approach: with emphasis on SPURS as case study
  - ✓ File format & metadata standardization via NODC .nc Feature Class templates
  - ✓ Application/enhancement of UNIDATA's Rosetta generalized File conversion Tool
  - ✓ Extend metadata model of PO.DAAC "DMAS" catalog
  - ✓ Evaluation of changes for external metadata interfaces (ECHO/GCMD)
  - ✓ Extension to Portal Faceted Search capabilities
  - ✓ Integration with <u>existing</u> PODAAC tools/services (eg. OPeNDAP, THREDDS, LAS)
  - ✓ New services: satellite-*in situ* collocation/matchup services

- Focus: - exclusively NASA Field-Campaign Data in support of science & NASA missions
  - broaden infrastructure & help drive technical innovation in area of satellite-*insitu* interoperability

podaac SPURS unidata NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)

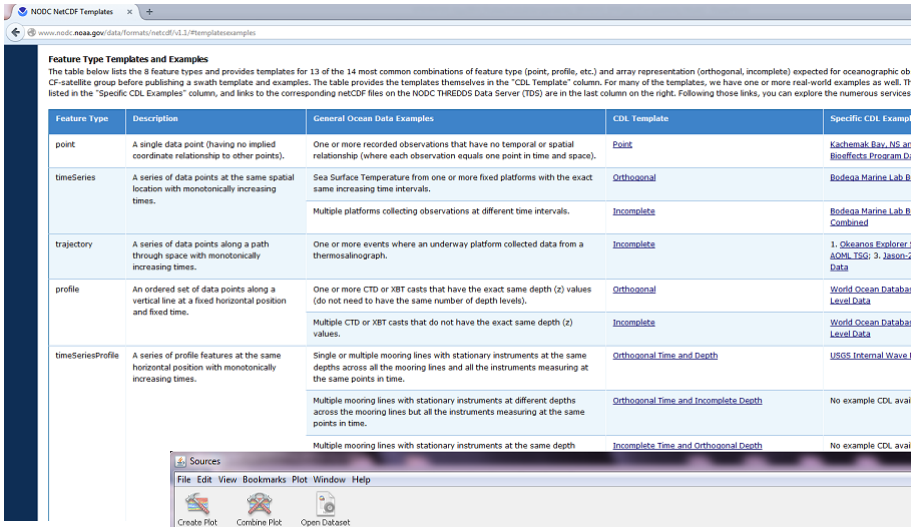
NOAA NATIONAL OCEANOGRAPHIC DATA CENTER (NODC)
UNITED STATES DEPARTMENT OF COMMERCE

Welcome to NODC | Access Data | Submit Data | Public Outreach | About NODC

NOAA Satellite and Information Service

You are here: Home › Formats and Codes › NetCDF Templates v1.1

NODC NetCDF Templates v1.1



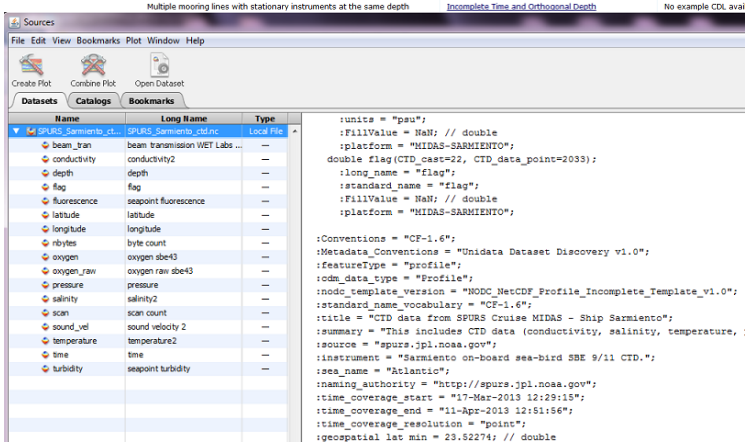
NODC NetCDF Templates

www.nodc.noaa.gov/data/formats/netcdf/v1.1/#templatesexamples

Feature Type Templates and Examples

The table below lists the 8 feature types and provides templates for 13 of the 14 most common combinations of feature type (point, profile, etc.) and array representation (orthogonal, incomplete) expected for oceanographic ob
CF-satellite group before publishing a swath template and examples. The table provides the templates themselves in the "CDL Template" column. For many of the templates, we have one or more real-world examples as well. Th
listed in the "Specific CDL Examples" column, and links to the corresponding netCDF files on the NODC THREDDS Data Server (TDS) are in the last column on the right. Following those links, you can explore the numerous services

Feature Type	Description	General Ocean Data Examples	CDL Template	Specific CDL Exampl
point	A single data point (having no implied coordinate relationship to other points).	One or more recorded observations that have no temporal or spatial relationship (where each observation equals one point in time and space).	Point	Kachemak Bay, NS ar Bioeffects Program Da
timeSeries	A series of data points at the same spatial location with monotonically increasing times.	Sea Surface Temperature from one or more fixed platforms with the exact same increasing time intervals.	Orthogonal	Bodega Marine Lab B
		Multiple platforms collecting observations at different time intervals.	Incomplete	Bodega Marine Lab B Combined
trajectory	A series of data points along a path through space with monotonically increasing times.	One or more events where an underway platform collected data from a thermosalinograph.	Incomplete	1. Okeanos Explorer AOML TSG; 3. Jason-2 Data
profile	An ordered set of data points along a vertical line at a fixed horizontal position and fixed time.	One or more CTD or XBT casts that have the exact same depth (z) values (do not need to have the same number of depth levels).	Orthogonal	World Ocean Databa Level Data
		Multiple CTD or XBT casts that do not have the exact same depth (z) values.	Incomplete	World Ocean Databa Level Data
timeSeriesProfile	A series of profile features at the same horizontal position with monotonically increasing times.	Single or multiple mooring lines with stationary instruments at the same depths across all the mooring lines and all the instruments measuring at the same points in time.	Orthogonal Time and Depth	USGS Internal Wave
		Multiple mooring lines with stationary instruments at different depths across the mooring lines but all the instruments measuring at the same points in time.	Orthogonal Time and Incomplete Depth	No example CDL avai
		Multiple mooring lines with stationary instruments at the same depth	Incomplete Time and Orthogonal Depth	No example CDL avai



Sources

File Edit View Bookmarks Plot Window Help

Create Plot | Combine Plot | Open Dataset

Datasets | Catalogs | Bookmarks

Name	Long Name	Type
SPURS_Sarmiento_ct...	SPURS_Sarmiento_ctd.nc	Local File
beam_tran	beam transmission WET Labs ...	—
conductivity	conductivity2	—
depth	depth	—
flag	flag	—
fluorescence	seapoint fluorescence	—
latitude	latitude	—
longitude	longitude	—
nbytes	byte count	—
oxygen	oxygen sbe43	—
oxygen_raw	oxygen raw sbe43	—
pressure	pressure	—
salinity	salinity2	—
scan	scan count	—
sound_vel	sound velocity 2	—
temperature	temperature2	—
time	time	—
turbidity	seapoint turbidity	—

```
    :units = "psu";
    :FillValue = NaN; // double
    :platform = "MIDAS-SARMIENTO";
  double flag(CTD_cast=22, CTD_data_point=2033);
    :long_name = "flag";
    :standard_name = "flag";
    :FillValue = NaN; // double
    :platform = "MIDAS-SARMIENTO";

:Conventions = "CF-1.6";
:Metadata_Conventions = "Unidata Dataset Discovery v1.0";
:featureType = "profile";
:cdm_data_type = "Profile";
:nodc_template_version = "NODC_NetCDF_Profile_Incomplete_Template_v1.0";
:standard_name_vocabulary = "CF-1.6";
:title = "CTD data from SPURS Cruise MIDAS - Ship Sarmiento";
:summary = "This includes CTD data (conductivity, salinity, temperature,
:source = "spurs.jpl.noaa.gov";
:instrument = "Sarmiento on-board sea-bird SBE 9/11 CTD.";
:sea_name = "Atlantic";
:naming_authority = "http://spurs.jpl.noaa.gov";
:time_coverage_start = "17-Mar-2013 12:29:15";
:time_coverage_end = "11-Apr-2013 12:51:56";
:time_coverage_resolution = "point";
:geospatial_lat_min = 23.52274; // double
```

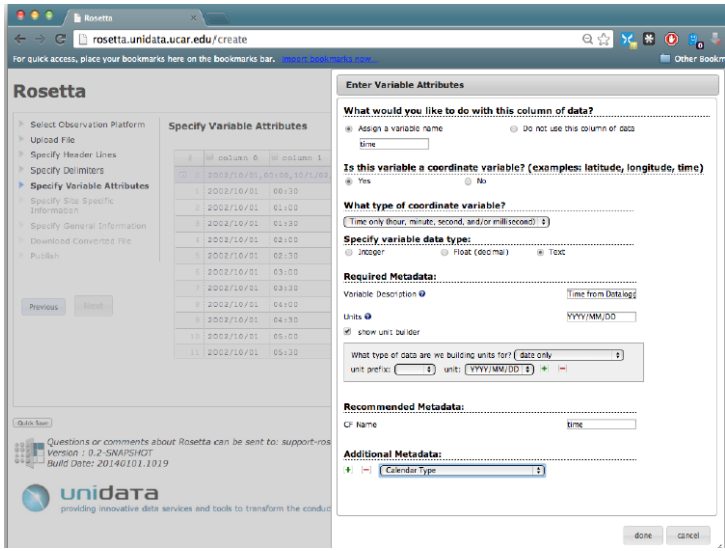
- NODC netCDF "spatial feature class" templates (8)
- Represent distinct sampling geometries:
  eg. point, profile, trajectory, time series etc.
- Standard Self-Describing file format (netCDF)
- CF & ACDD compliant metadata:
  share common, standard global & variable attributes
- Well documented with good examples
- Launched in 2012 after period of public comment

- PODAAC's assessment:
  - Important standardization framework for oceanographic field data
  - NODC Templates central to *in-situ* data archival solution at PODAAC
  - Recommended to SPURS & adopted
  - Provided guidance to SPURS for implementation
  - Caveat: Widespread adoption constrained by lack of an associated conversion tool

Rosetta: generalized purpose web-based format converter

http://rosetta.unidata.ucar.edu

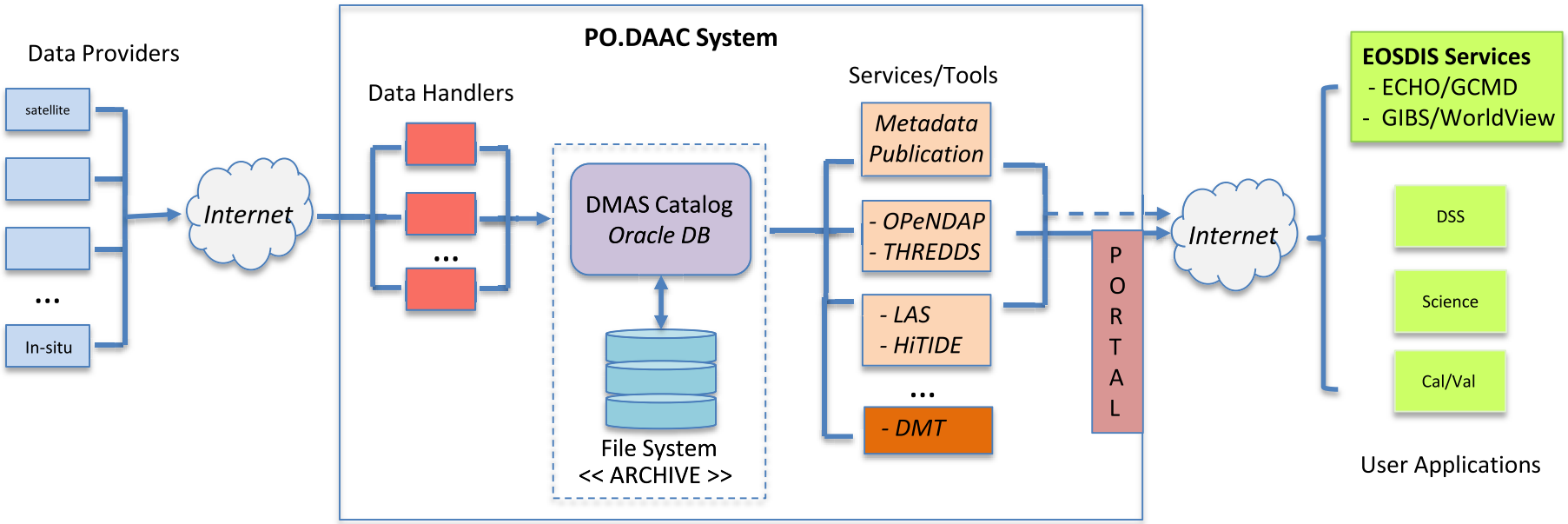
- Developed by Sean Arms (UNIDATA)
- Generalized web-based tool for conversion of arbitrary, unstructured ASCII data files to CF compliant netCDF files (and Excel)
- Developed & Applied under NSF-ACADIS project "Advanced Coop. Arctic Data & Information Service"
- GUI wizard – guided, step-wise process for conversion & augmentation of file metadata by user
- Burden of conversion mechanics removed from data producer
- Comprehensive, Robust, Consistent translation framework
- Java Web-App Tech. Stack: Java, JS, Spring, Tomcat



- Proposed Enhancements in 2013 ACCESS-SPURS proposal:
  - extend support for NODC .nc templates
  - implement as a RESTful web service interface enabling automated/programmatic data transformation
  - integration within PO.DAAC DMAS

Adapting the PO.DAAC Infrastructure for In-Situ Support

- **DMAS**: PO.DAAC's **D**ata **M**anagement & **A**rchival **S**ystem
- Extensions of DMAS data model & components to support *in-situ* metadata
  - Handlers
  - Catalog
  - Portal
  - Metadata services



Developing a Generalized Satellite-In Situ Collocation Service

- 2014 ROSES AIST proposal to develop a Distributed Matchup Service (DOMS)

  S. Smith (FSU), M. Bourassa (FSU), T.Huang (JPL), V.Tsontos (JPL), B. Holt (JPL), S. Worley (NCAR)

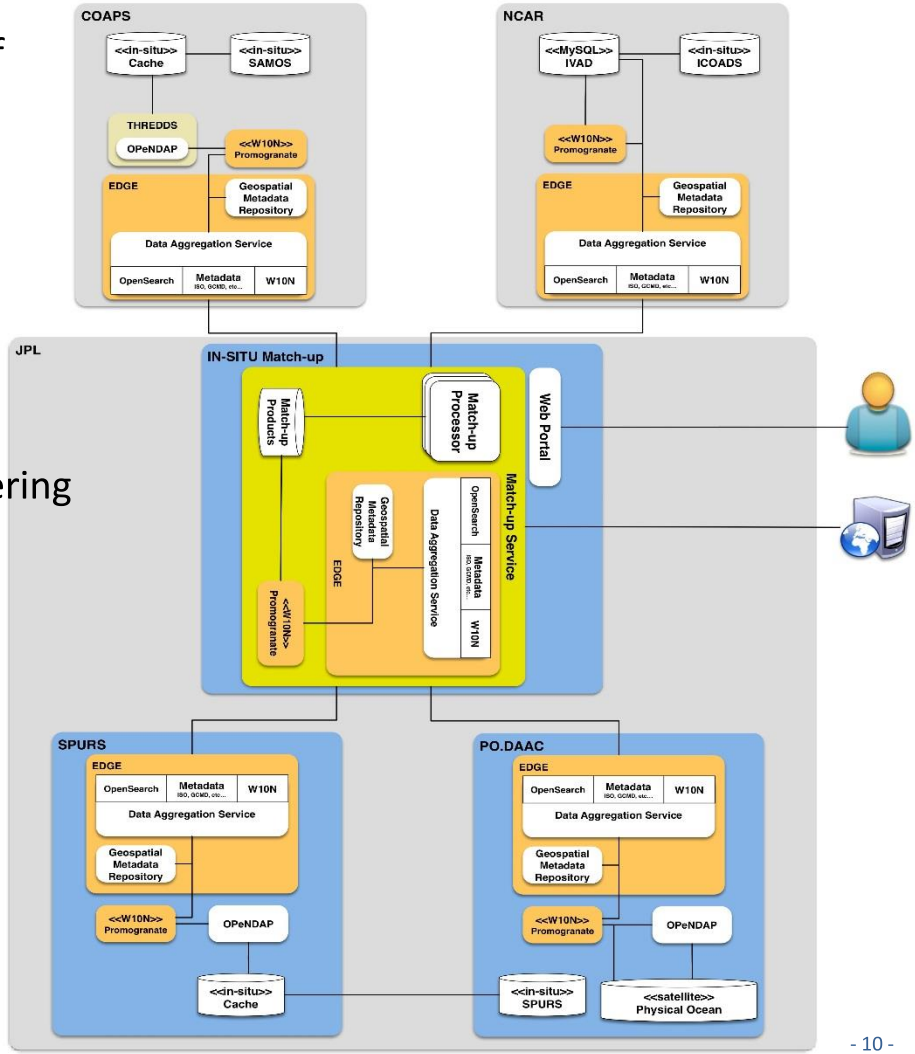
- Generalized, publically accessible service in support of research & mission cal/val

- Hosted @ PO.DAAC but distributed data/providers
  - NCAR: ICOADS/ARGO
  - FSU COAPS: SAMOS
  - PO.DAAC: S PURS

- Web-service & Form-based querying with flexible filtering
  - by instrument, sensor, parameter
  - spatio-temporal collocation criteria (x,y,z,t) & search radii

- Range of use cases established:
  - drive requirement & development

- 2 year project: 2015 kick-off



Conclusions

- General need for more seamless access to diverse, distributed datasets for earth science applications: science, decision support, mission cal/val

- PO.DAAC vision:
  - Work with the community to address related technical issues & help drive necessary innovation
  - Support for NASA field campaign data (SPURS I & II)

- Central Interoperability Challenge for In-situ Datasets
  - Inherent diversity, complexity, heterogeneity
  - Acute lack of adherence to metadata & data format standards
  - Insufficient knowledge/technical capacity, resources, interest by data producers

- Tractable: Leverage existing standards & frameworks to address metadata/interoperability issues to extend the PO.DAAC system
  - eg. NODC netCDF templates, ROSETTA, MMI

- Develop new capabilities facilitating improved integrated access/usage of satellite & in-situ data
  - AIST Matchup Service (DOMS)