

Towards Improved Satellite-In situ Oceanographic Data Interoperability and Associated Value Added Services at the PO.DAAC

Vardis Tsontos

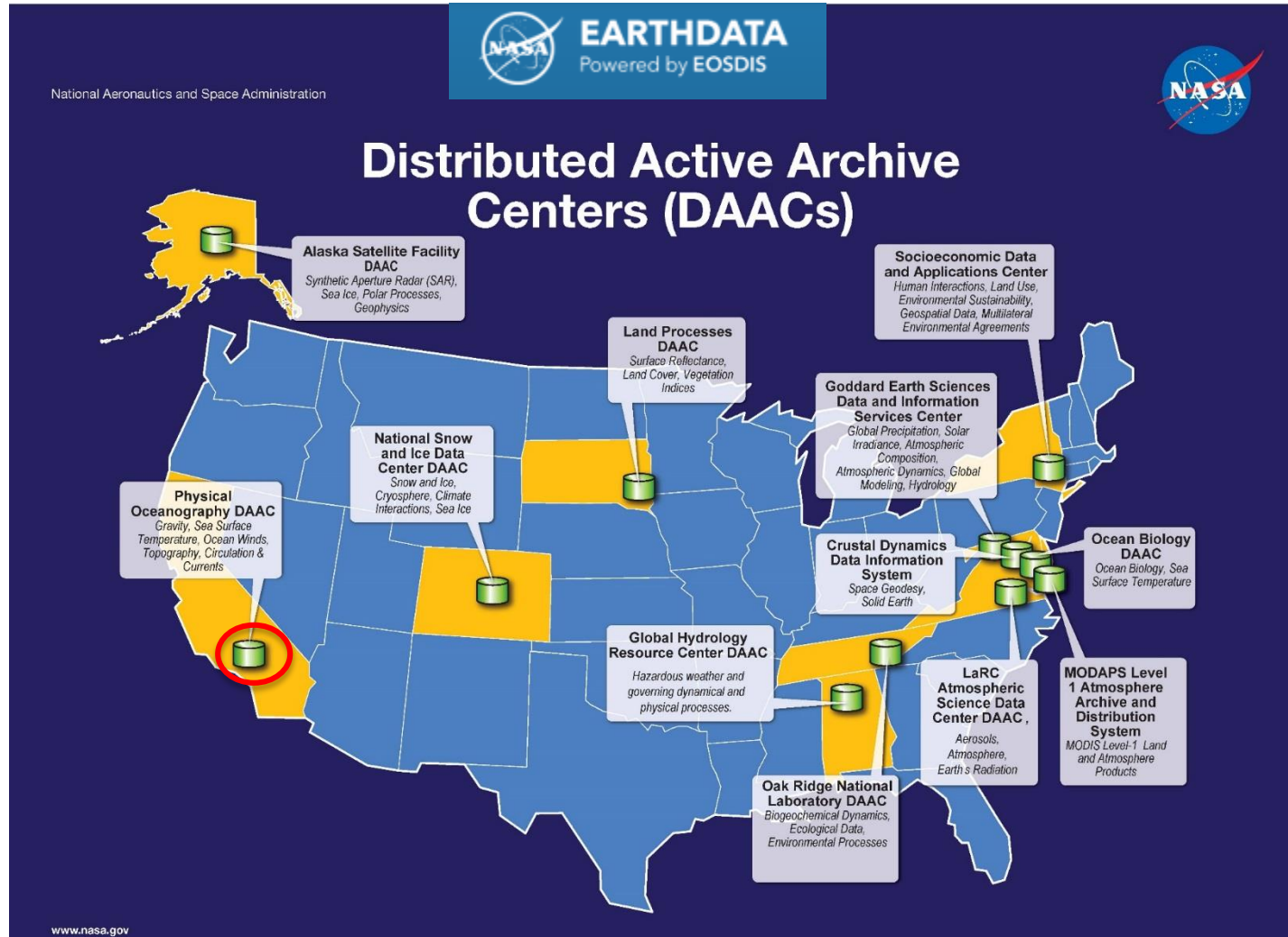
NASA PO.DAAC

Jet Propulsion Laboratory, California Institute of Technology

vardis.m.tsontos@jpl.nasa.gov

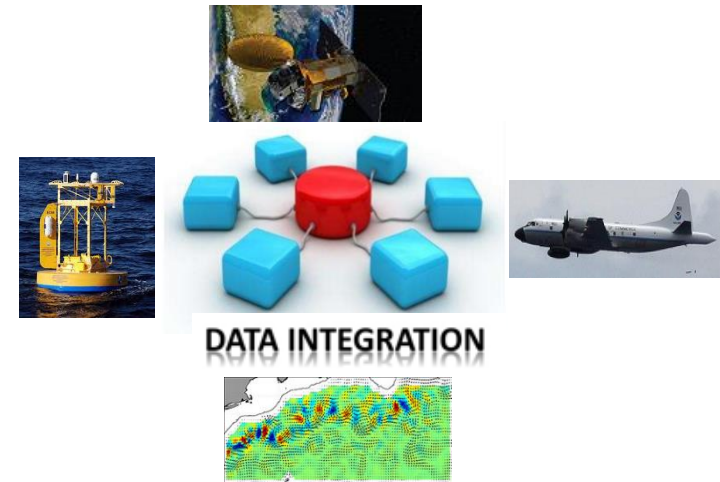
ESA Living Planet Symposium 2016

- Designated NASA archive for Physical Oceanographic Satellite Data
Parameters: SST, SSS, Ocean Winds, Altimetry, Gravity, Sea Ice
- Recently also provide NASA field campaign support: SPURS, OMG



Motivation - General Considerations

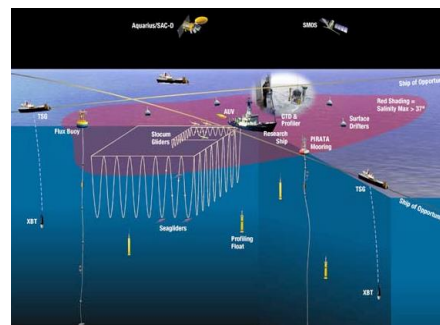
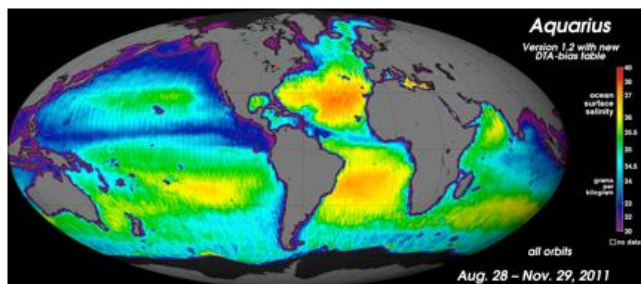
- The earth science enterprise increasingly relies on the **integration & synthesis** of **multivariate datasets** from **diverse observational platforms**
- Applications:
 - Science
 - Decision Support
 - Mission Cal/Val
- Increased availability of oceanographic data online and improved access protocols & tools for access
- Outstanding Issues from the user perspective:
 - Inherent complexity/heterogeneity of data
 - Disparate, distributed systems that serve them
 - Need more unified data access
- Problem Origin:
 - Technical issues: variable adherence to emerging metadata/file standards & protocols
 - Structural/historical reasons: domain specific data providers
- PO.DAAC's Vision to help address these issues & drive the necessary technical innovation to more seamlessly support both NASA satellite mission and field campaign data



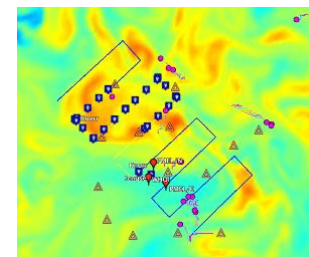
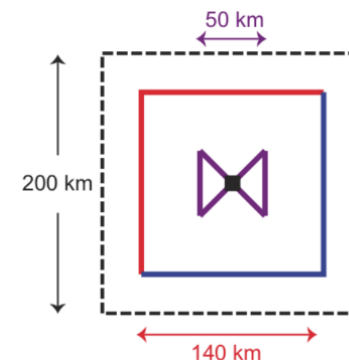
Aquarius and the SPURS Field Program

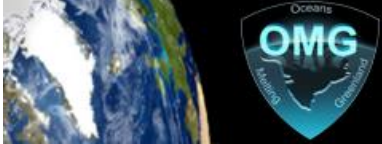
- **Aquarius-SAC/D NASA salinity mission:** <http://aquarius.nasa.gov>
 - Global surface salinity observations: improved understanding of water cycle
 - 3 years, 9 months of data from Aug. 2011 archived & distributed by PO.DAAC
- **SPURS:** NASA-funded salinity oceanographic field campaigns <http://spurs.jpl.nasa.gov>
 - Salinity process study and synergies with Aquarius:
 - sub-footprint SSS variability column structure & surface salinity field characterization
 - Inter-agency & International effort/support
 - SPURS I: salinity max region N. Atlantic (5 cruises 2012-13) <http://podaac.jpl.nasa.gov/spurs>
 - SPURS II: high rainfall ITCZ area E. Tropical Pacific (2016-17)
 - Range (novel) platforms & instruments: Multi-scale “Sensor-Web”
 - Diverse/Heterogeneous Data:
 - 15 core data sets (1GB): *CTD, UCTD, Seasoar, ADCP, TSG, MET, Mooring (2), glider (4), drifter, float (3)*
 - Initially in native “unstructured” file formats & metadata (*no requirements on PIs*)

Aquarius/SAC-D

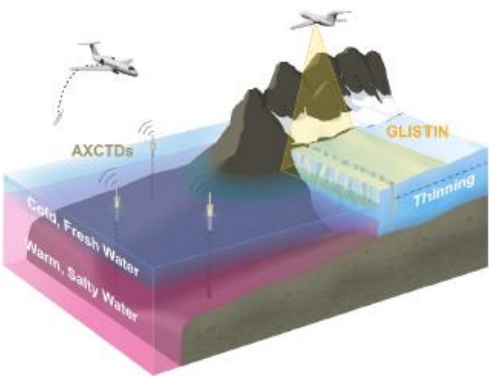


Salinity Processes in the Upper Ocean
Regional Study (SPURS) Field Experiment

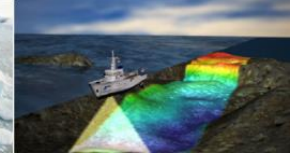
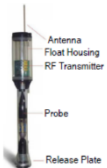
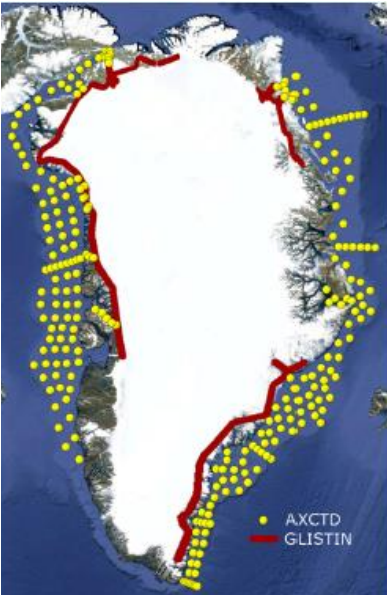




- NASA 5-year airborne & oceanographic field campaign (initiated July 2015)
- Aim: understand role of ocean-ice interactions in contributing to loss of ice from the Greenland Ice Sheet
 - observe changing water temperatures on the continental shelf surrounding Greenland
 - glaciers response to the presence of warm, salty Atlantic Water
 - role of complex shoreline & bottom topography on flows/melt rates
- Observations & Data: Archival/Distribution via PO.DAAC

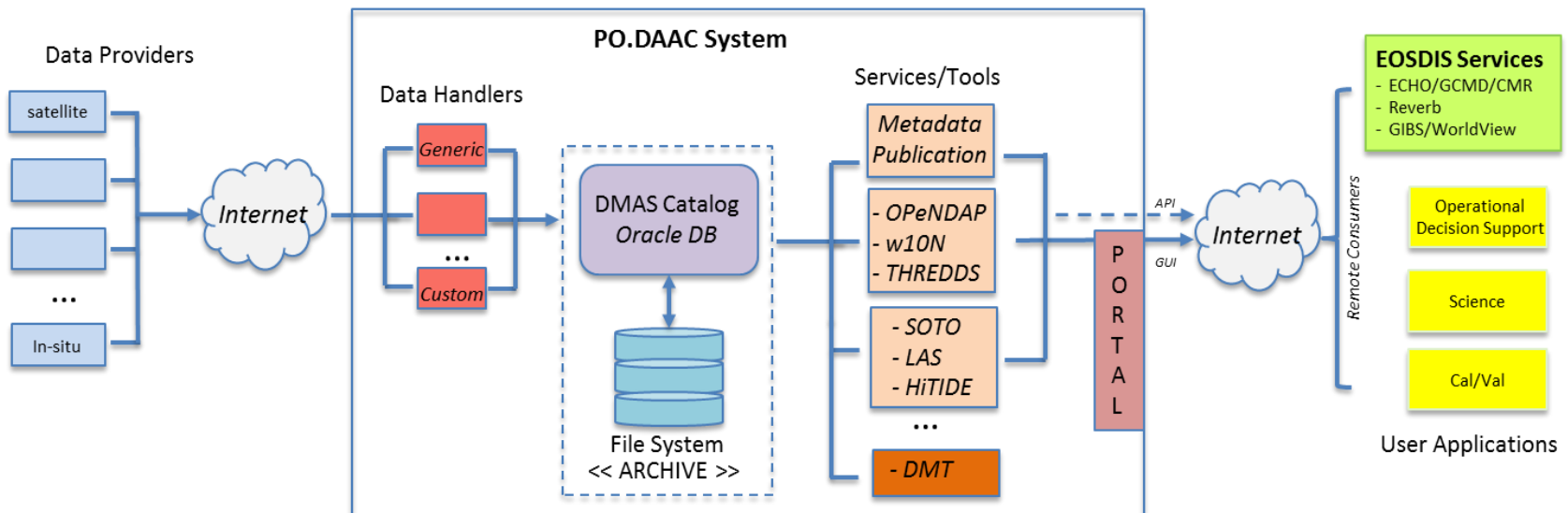
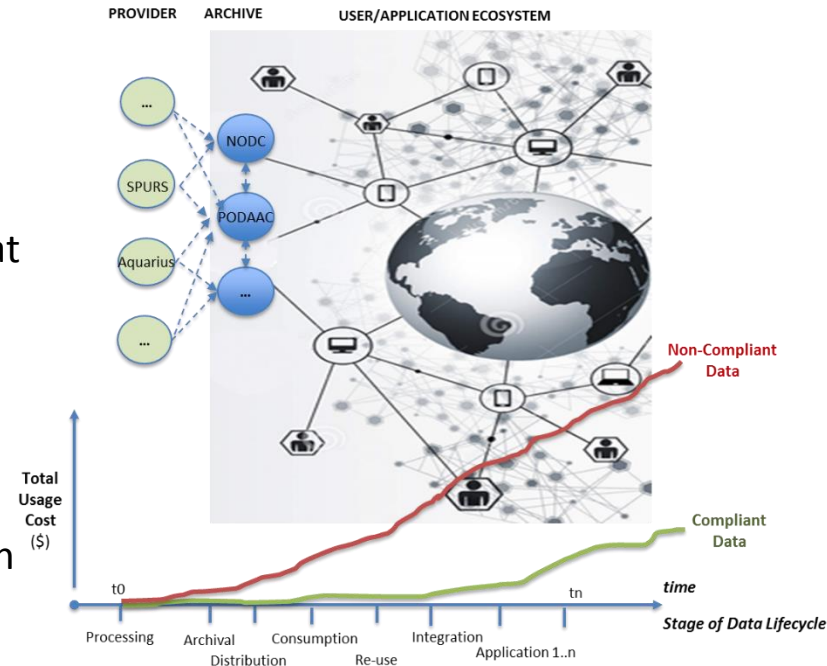


Venue	Instrument	Level	Data Product
Ocean	AXCTD	0	AXCTD raw data
		1	T & S Profiles with GPS
		2	Iso- T & S Profile Maps
ICE	GLISTIN-A	0	GLISTIN raw data
		1	SLC data
		2	DEM of Ice Surface
Bathymetry	AIRGrav	0	AIRGrav raw data
		2	Gravity maps by SGL
		3	Bathymetry maps (Gravity based)
	Ship Survey	0	Sonar survey raw data
		2	Bathymetry data
		3	Bathymetry maps
OMG Science	Science	4	Ocean State / Ice Loss / Sea Level Prediction



INTEROPERABILITY: IMPLICATIONS

- Inherent diversity, complexity, heterogeneity of field data
- Lack of adherence to metadata & data format standards that are particularly acute for *in-situ* data
- Impacts archival, discoverability & tool/service integration
- Exponentially increasing costs for non-compliant data across the dataset lifecycle -> importance of early mitigation



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

- ✓ Practical/evolutionary approach: with emphasis on SPURS & OMG as use cases
- ✓ 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 (NASA-CMR)
- ✓ Extension to Portal Faceted Search capabilities
- ✓ Integration with existing PODAAC tools/services (eg. OPeNDAP, THREDDS, LAS, SOTO)
- ✓ New value-added services: eg. satellite-*in situ* collocation service

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



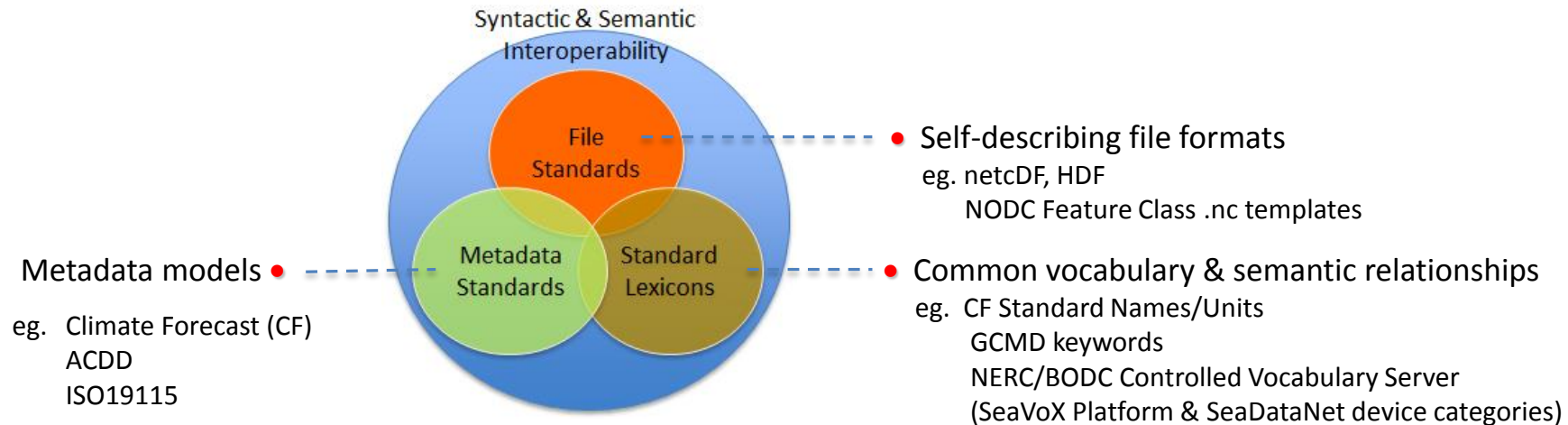
NATIONAL OCEANOGRAPHIC
DATA CENTER (NODC)



Interoperability

The ability of coupled systems (machine-to-machine) to:

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



Marine Metadata Interoperability
Ontology Registry and Repository

Help | Terms of Use | Create account | Sign in

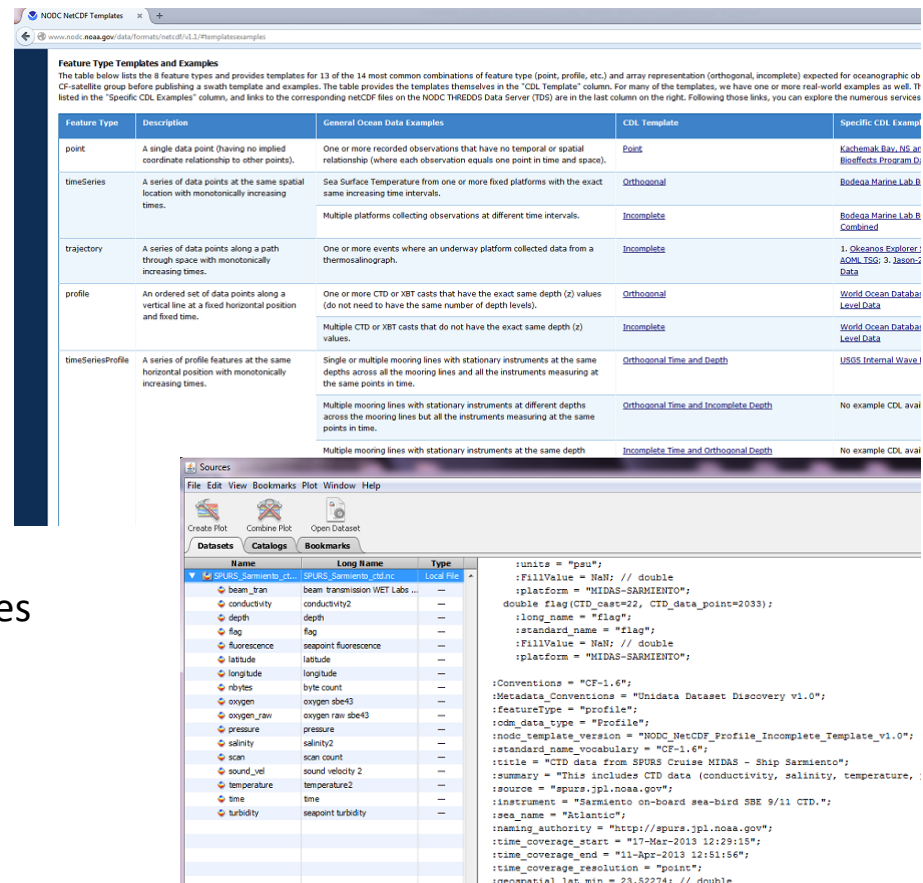
Search terms

Refresh	Search ontologies	URI	Name
All ontologies			
Type			
mapping	1	http://mmiew.org/ont/ios/ios_core_cf_map	IOOS Core Variables to Climate Forecast Standard Name Map
vocabulary	2	http://mmiew.org/ont/ios/map_ios_cf	IOOS Parameter to Climate Forecast Standard Name Map
Authority	3	http://mmiew.org/ont/ios/platform_map_ios_mmi	IOOS Platform to MMI Platform Ontology Map
age	4	http://mmiew.org/ont/ios/platform_map	IOOS Platform to IOOS Platform Hierarchy
avars	5	http://mmiew.org/ont/ios/map_ios_ios	IOOS Parameter to IOOS Parameter
argo	6	http://mmiew.org/ont/ios/map_ios_variable	IOOS Core Variable Map
bodc	7	http://mmiew.org/ont/ios/map_societal_areas	IOOS Societal Benefit Areas Map
catfish	8	http://mmiew.org/ont/ios/map_category	IOOS Category Map
cdip	9	http://mmiew.org/ont/ios/map	OBCC and CF Mapping
cercoos	10	http://mmiew.org/ont/ios/device_map	Map of ECIS Device to MMI Sensor

NODC NetCDF Templates v1.1

- 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 & OMG: adopted
 - Provided guidance to SPURS for implementation
 - Caveat & Lessons Learned from SPURS:

Widespread adoption constrained by lack of an associated conversion tool



The screenshot shows the NODC NetCDF Templates v1.1 web page. The page lists 13 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 data. 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 data.

Feature Type	Description	General Ocean Data Examples	CDL Template	Specific CDL Examples
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 and Bioeffects Program Data
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. Multiple platforms collecting observations at different time intervals.	Orthogonal Incomplete	Bodega Marine Lab B 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; ACMS 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). Multiple CTD or XBT casts that do not have the exact same depth (z) values.	Orthogonal Incomplete	World Ocean Database Level Data World Ocean Database 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. Multiple mooring lines with stationary instruments at different depths across the mooring lines but all the instruments measuring at the same points in time. Multiple mooring lines with stationary instruments at the same depth	Orthogonal Time and Depth Orthogonal Time and Incomplete Depth Incomplete Time and Orthogonal Depth	USOS Internal Wave No example CDL avail No example CDL avail

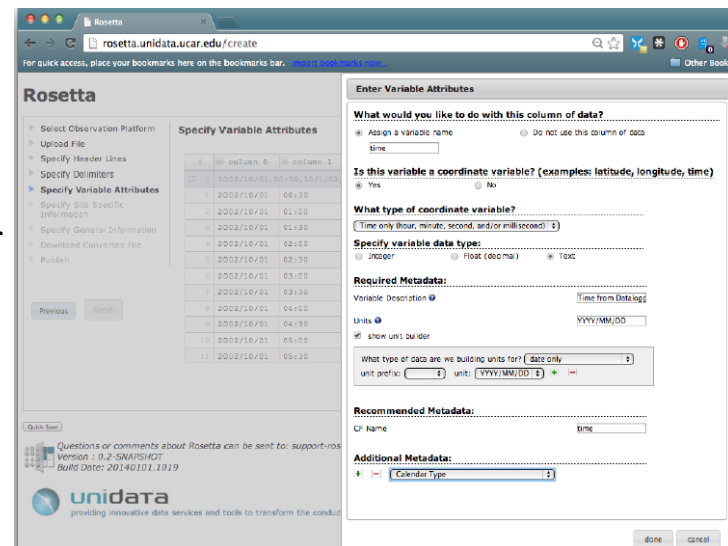
The screenshot also shows a netCDF metadata editor window. The editor displays the following metadata:

```

:units = "pau";
:FillValue = NaN; // double
:platform = "MIDAS-SARMIENTO";
double flag(CTD_cast=24, 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";
:nc_data_type = "Profile";
:ncdc_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 12:29:19";
:time_coverage_end = "11-Apr-2013 12:51:56";
:time_coverage_resolution = "point";
:geospatial_lat_min = 23.52274; // double
  
```

- Developed by Sean Arms (UNIDATA)
- Generalized web-based tool for conversion of arbitrary, unstructured ASCII data files to CF compliant netCDF files
- 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
- Web-App. Tech. Stack: Java, JS, Spring, Tomcat
- Planned Enhancements:
 - extend support for NODC .nc templates
 - implement as a RESTful web service interface enabling automated/programmatic data transformation
 - integration within PO.DAAC DMAS

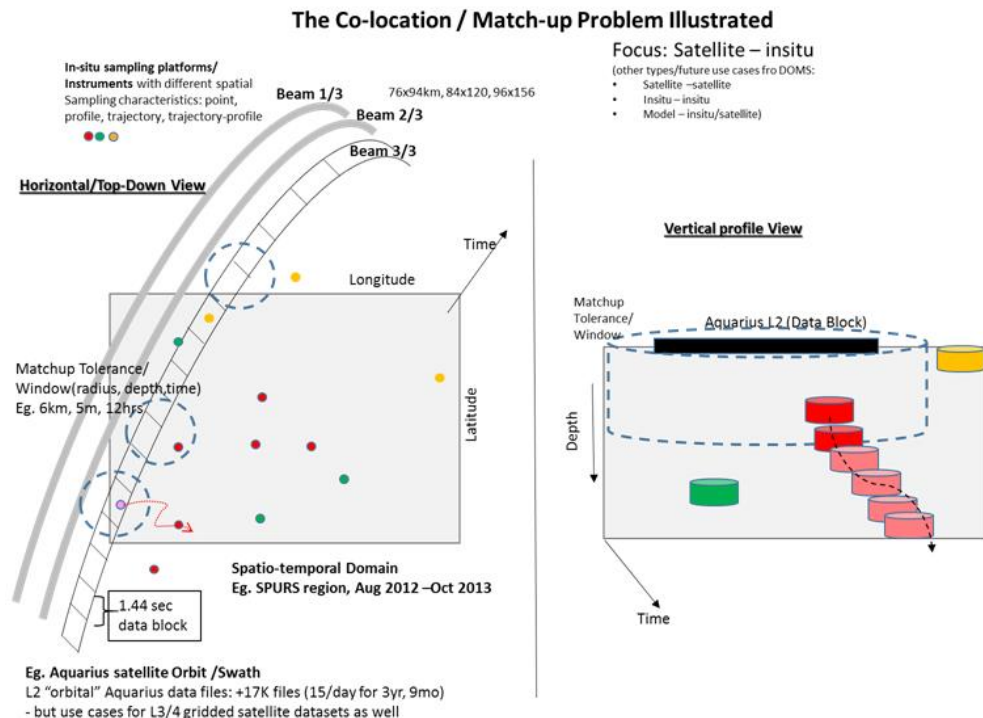


Developing a Generalized Satellite-In Situ Collocation Service

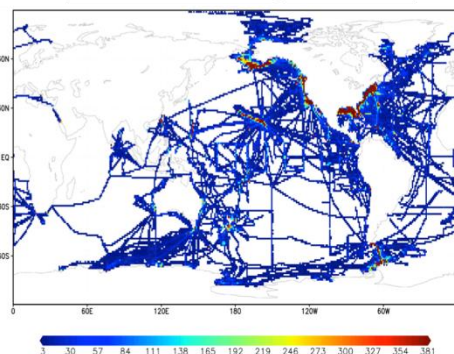
2014 NASA/AIST Project to develop a Distributed Matchup Service (DOMS) for Satellite-Insitu Data

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

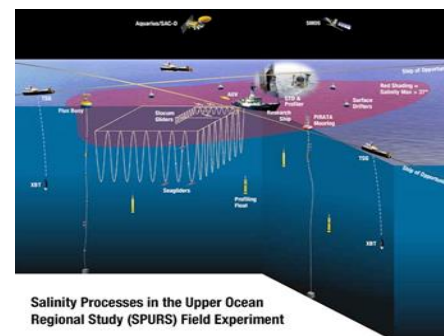
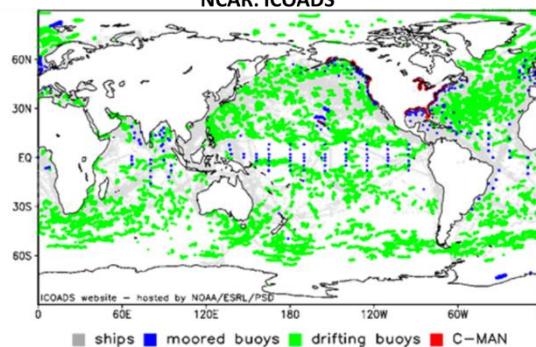
- 2 year tech development project: 2015 kick-off
 - Aim: develop generalized, scalable, publically accessible matchup service in support of research & mission cal/val
 - Distributed data/providers
 - NCAR: ICOADS/ARGO
 - FSU COAPS: SAMOS US Research Cruises
 - PO.DAAC: SPURS
- Satellite SST, SSS, Winds (L2/3/4)



SAMOS Data Density: 2005-2014



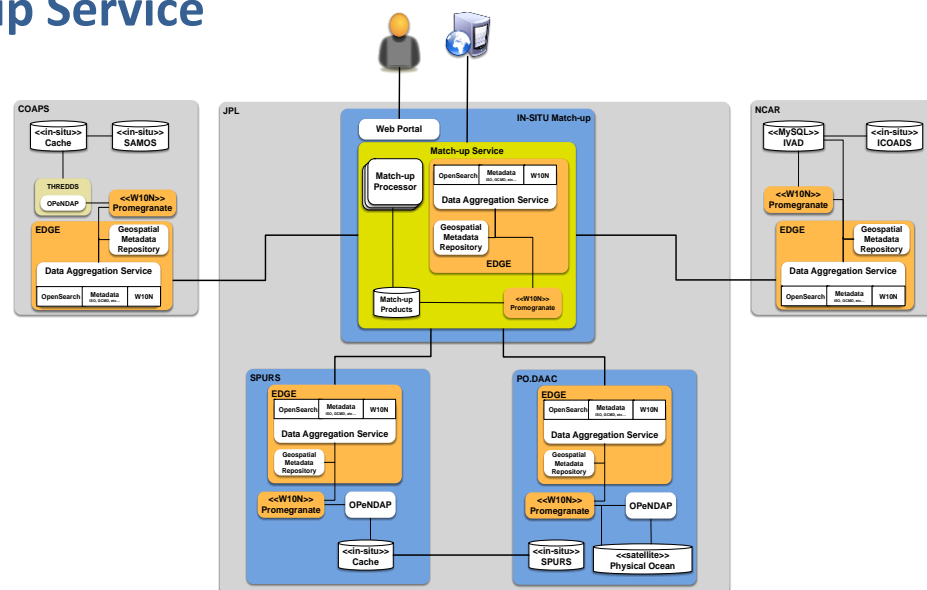
NCAR: ICOADS



DOMS: Distributed Oceanographic Matchup Service

Architecture

- Common software stack (eg. Edge, W10n)
- supports heterogeneous backend repositories (RDB, File Archive/THREDDS, Cassandra NoSQL)
- Distributed queries
- Computationally demanding parallel KD-Tree matchup operations close to high volume satellite data at PODAAC
- Service hosted & publicly accessible via PODAAC



DOMS interfaces: Web-service & Form-based querying with flexible filtering

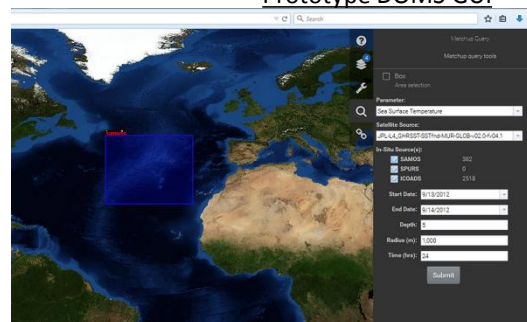
- by instrument, sensor, parameter
- spatio-temporal domain & matchup window criteria (x,y,z,t)

Example DOMS metadata query URI

```
SPURS Data Service
- Sample query: curl
  "https://deepdata.jpl.nasa.gov/ws/search/spurs?startTime=2012-08-
  01T00:00:00Z&endTime=2013-10-31T23:59:59Z&box=-45,15,-30,30"
- Processed datasets
  • SPURS1_CTD
  • SPURS1_ADCP
  • SPURS1_ARGO
  • SPURS1_SEAGLIDER
  • SPURS1_MOORING_WHOI
- 13,817,392 documents (each with sst, wind, and sss measurements)

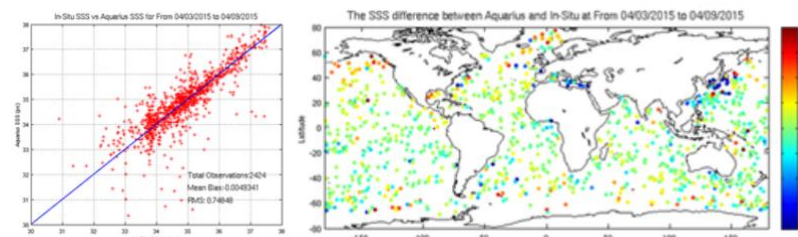
Satellite Data Service
- Sample query: curl -X GET
  "https://deepdata.jpl.nasa.gov/nexus/doms/datainbounds?ds=JPL-L4_GHRSST-
  SSTInd-MUR-GLOB-v02.0-fv04.1&env=DEV&minLat=23.93&maxLat=25.26&minLon=-
  39.74&maxLon=-37.33&startTime=1345280400000&endTime=1345366800000"
- Dataset: GHRSST Multi-scale Ultrahigh Resolution (MUR) - Over the SPURS region
  • Metadata index - 585,792 documents
  • Data tiling - 591,887 tiles
```

Prototype DOMS GUI



Outputs:

- Tabular output files (csv, .nc)
- Graphical summaries





Conclusions

- General need for more seamless/integrated 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, OMG)
- 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 by data producers
- Tractable: Leverage existing standards & frameworks to address metadata/interoperability issues to extend the PO.DAAC system
 - eg. NODC netCDF templates, ROSETTA, BODC controlled vocabs
- Value-Added services facilitating improved integrated access/usage of satellite & in-situ data
 - eg. DOMS Matchup Service