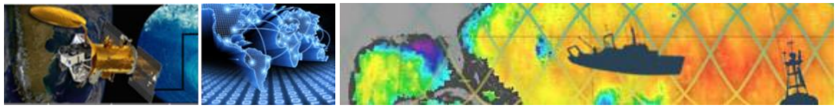


Towards Improved Satellite-In situ Oceanographic Data Interoperability and Associated Value Added Services at the PO.DAAC

Vardis Tsontos

NASA PO.DAAC

Jet Propulsion Laboratory, California Institute of Technology

vardis.m.tsontos@jpl.nasa.gov

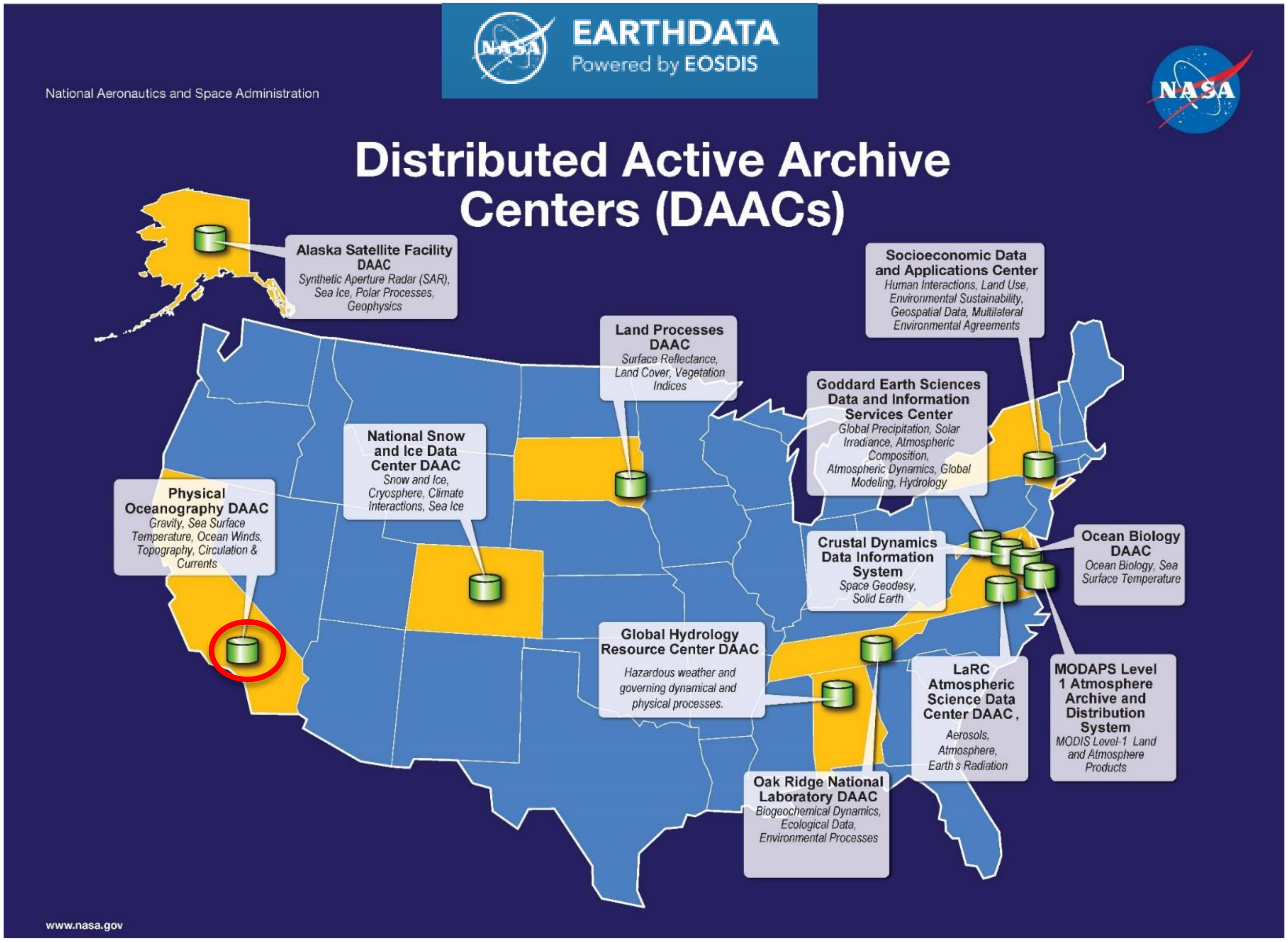
ESA Living Planet Symposium 2016

URS CL#16-1745

ESA-LPS May 9-14, 2016
Prague, Czech Republic

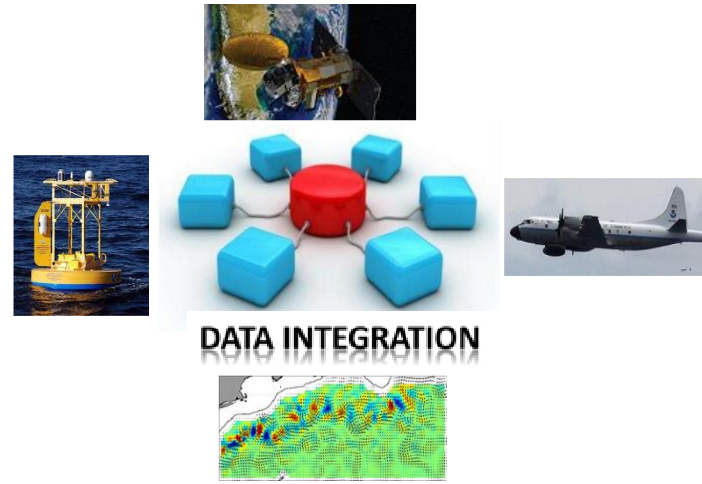
PO.DAAC https://podaac.jpl.nasa.gov/

- 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



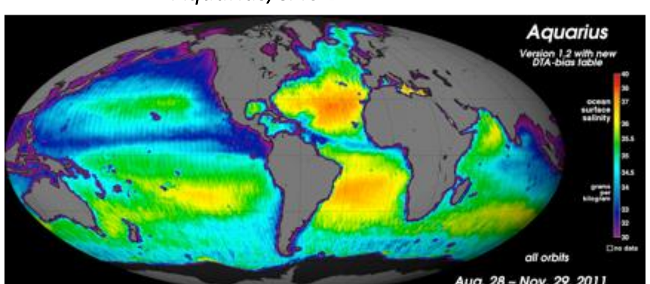
- Increased availability of oceanographic data online and improved access protocols & tools for access
- Outstanding Issues from the user perspective:
  - Inherent complexity/heterogeneity of data
  - Disparate, distributed systems that serve them
  - Need more unified data access
- Problem Origin:
  - 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

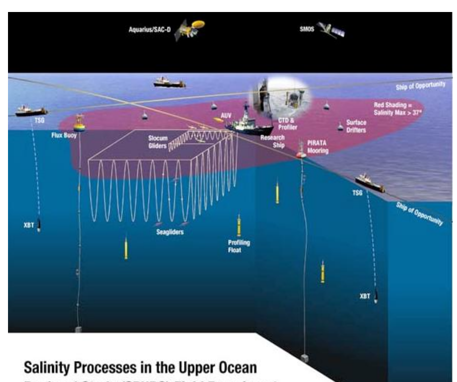
Aquarius and the SPURS Field Program

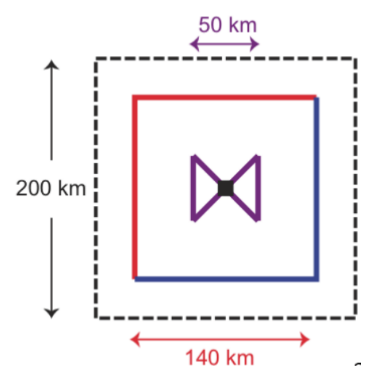
- **Aquarius-SAC/D NASA salinity mission**: http://aquarius.nasa.gov
  - Global surface salinity observations: improved understanding of water cycle
  - 3 years, 9 months of data from Aug. 2011 archived & distributed by PO.DAAC

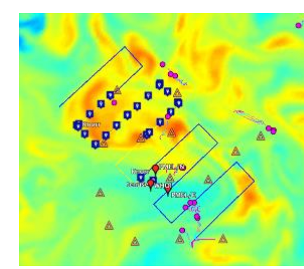
- **SPURS**: NASA-funded salinity oceanographic field campaigns http://spurs.jpl.nasa.gov
  - Salinity process study and synergies with Aquarius:
    sub-footprint SSS variability column structure & surface salinity field characterization
  - Inter-agency & International effort/support
  - SPURS I: salinity 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



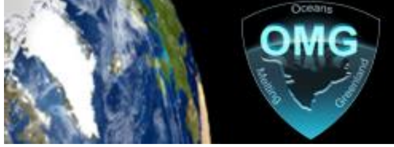






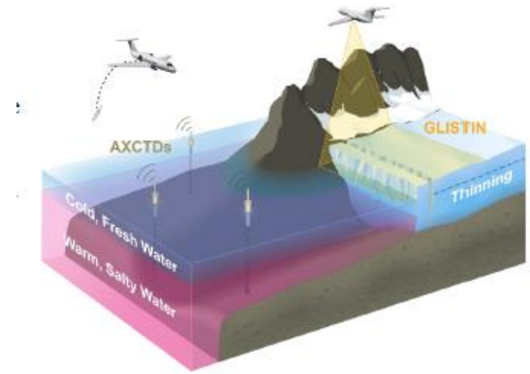
Ocean Melting Greenland (OMG)

https://omg.jpl.nasa.gov/portal/



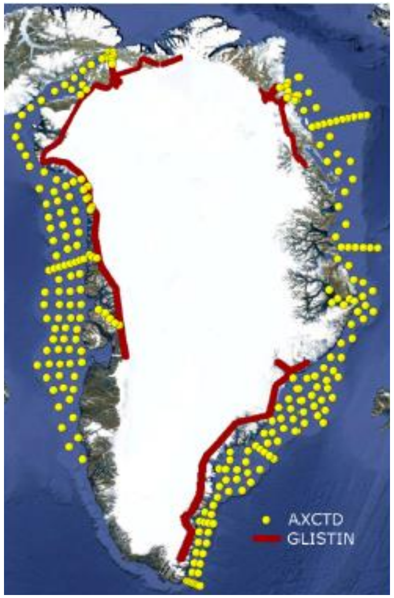
- NASA 5-year airborne & oceanographic field campaign (initiated July 2015)

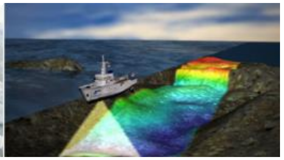
- Aim: understand role of ocean-ice interactions in contributing to loss of ice from the Greenland Ice Sheet
  - observe changing water temperatures on the continental shelf surrounding Greenland
  - glaciers response to the presence of warm, salty Atlantic Water
  - role of complex shoreline & bottom topography on flows/melt rates



- Observations & Data: Archival/Distribution via PO.DAAC

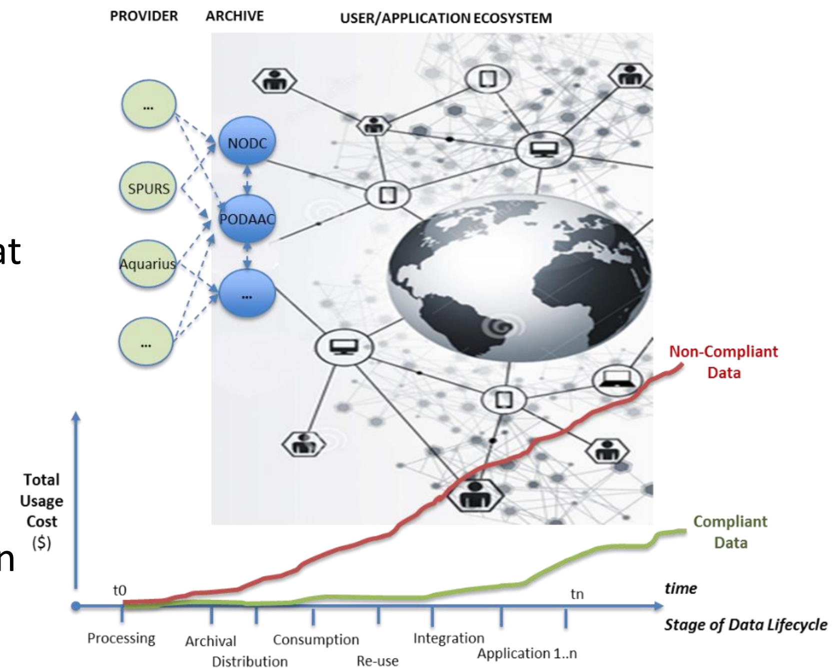
Venue	Instrument	Level	Data Product
Ocean	AXCTD	0	AXCTD raw data
		1	T & S Profiles with GPS
		2	Iso- T & S Profile Maps
ICE	GLISTIN-A	0	GLISTIN raw data
		1	SLC data
		2	DEM of Ice Surface
Bathymetry	AIRGrav	0	AIRGrav raw data
		2	Gravity maps by SGL
		3	Bathymetry maps (Gravity based)
	Ship Survey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

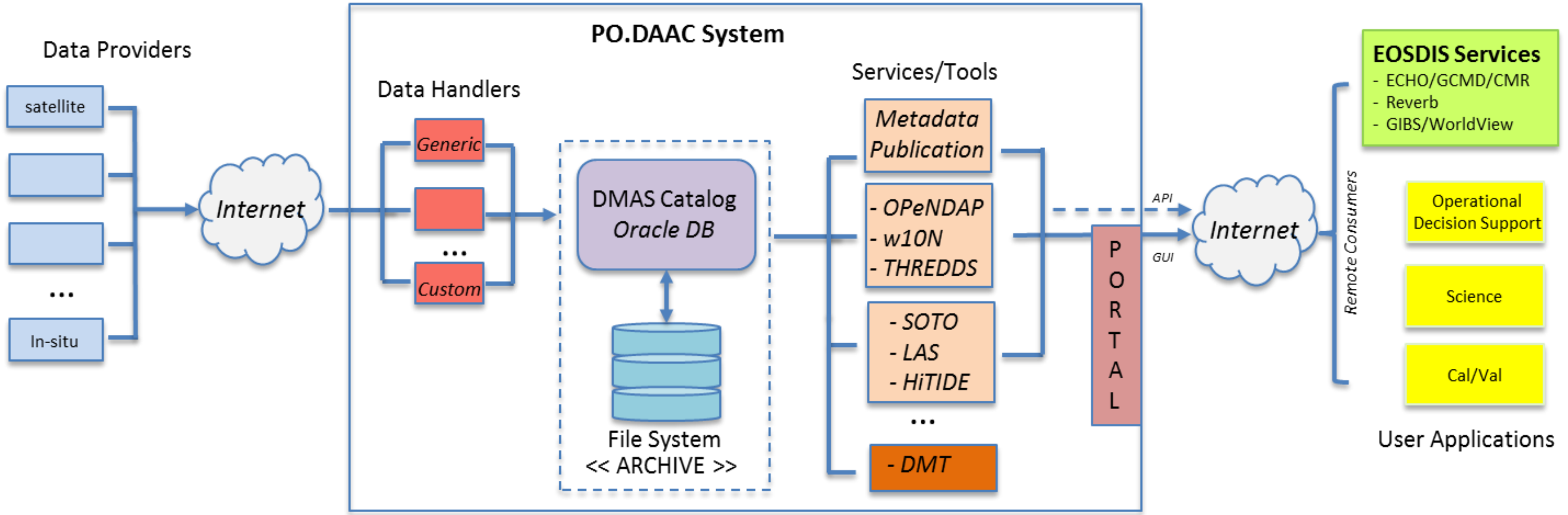




INTEROPERABILITY: IMPLICATIONS

- 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 <u>existing</u> 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

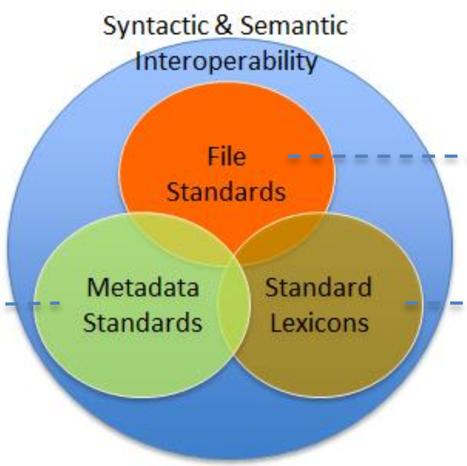
EARTHDATA Powered by EOSDIS

NATIONAL OCEANOGRAPHIC DATA CENTER (**NODC**)

Interoperability

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

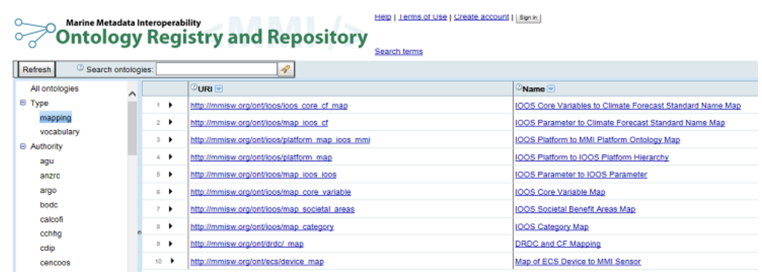
- Communicate and exchange data via common formats & protocols
- Meaningfully interpret and reproducibly act on exchanged data

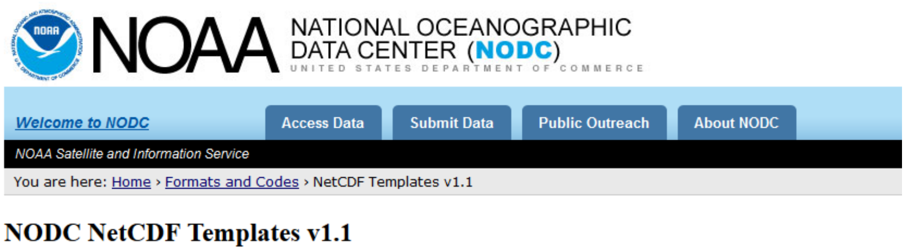


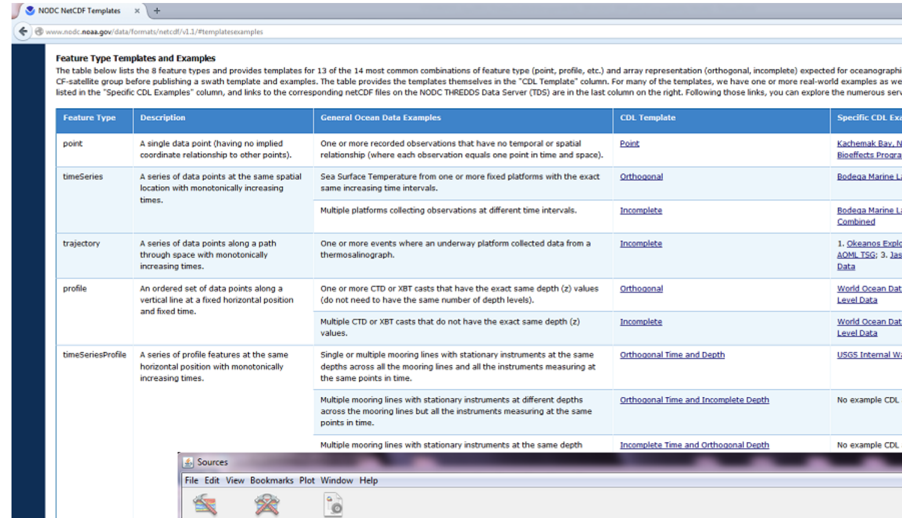
- Self-describing file formats
  eg. netcDF, HDF
  NODC Feature Class .nc templates

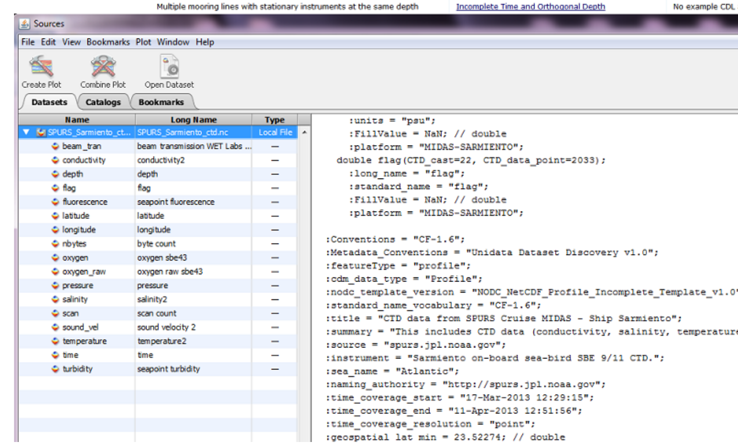
- Metadata models
  eg. Climate Forecast (CF)
  ACDD
  ISO19115

- Common vocabulary & semantic relationships
  eg. CF Standard Names/Units
  GCMD keywords
  NERC/BODC Controlled Vocabulary Server
  (SeaVoX Platform & SeaDataNet device categories)









- 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

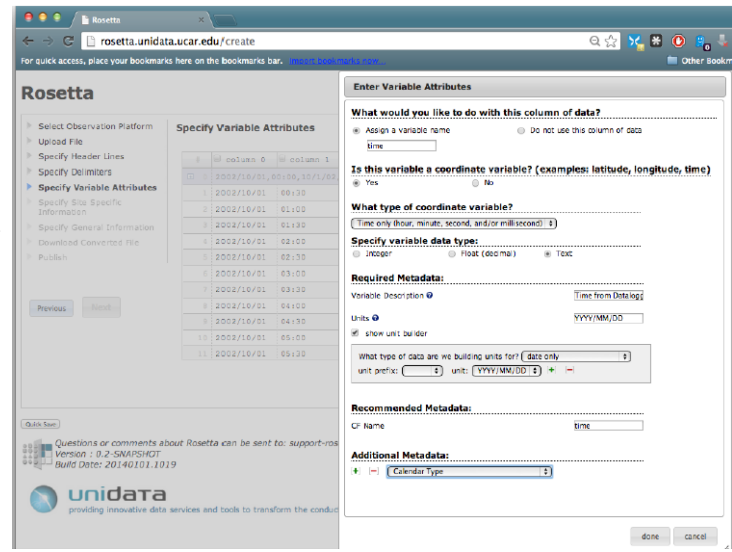
- PODAAC’s assessment:
  - Important standardization framework for oceanographic field data
  - NODC Templates central to *in-situ* data archival solution at PODAAC
  - Recommended to SPURS & OMG: adopted
  - Provided guidance to SPURS for implementation
  - Caveat & Lessons Learned from SPURS:

    Widespread adoption constrained by lack of an associated conversion tool

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
- 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

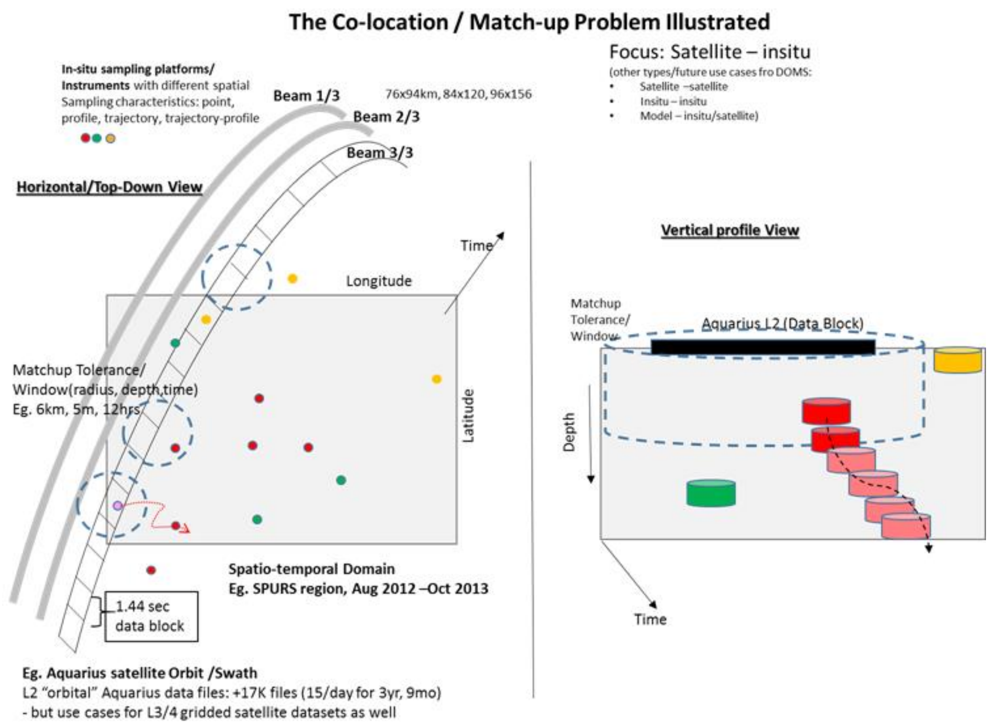
Developing a Generalized Satellite-In Situ Collocation Service

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

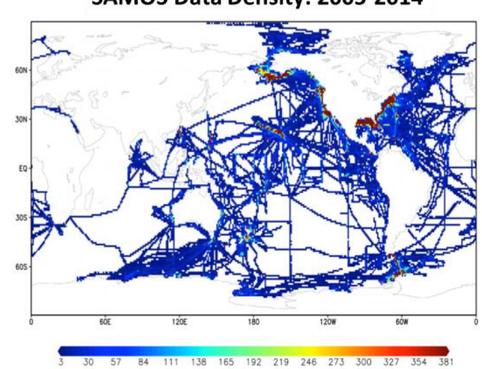
S. Smith (FSU), M. Bourassa (FSU), T. Huang (JPL), V. Tsontos (JPL), B. Holt (JPL), S. Worley (NCAR)

- 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

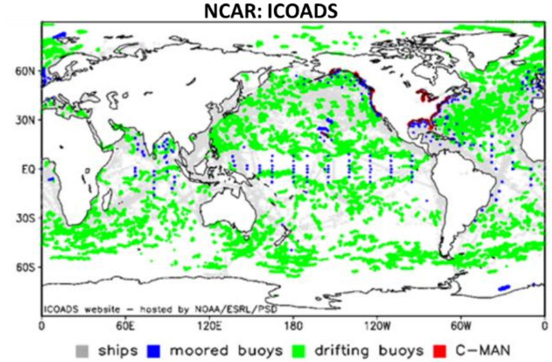
  Satellite SST, SSS, Winds (L2/3/4)

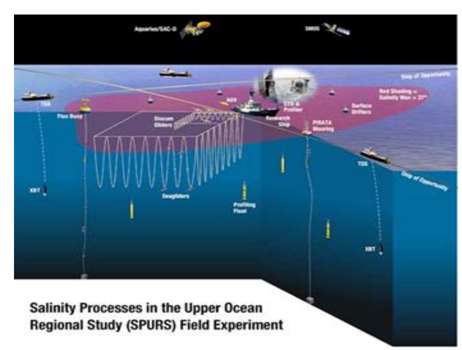


SAMOS Data Density: 2005-2014



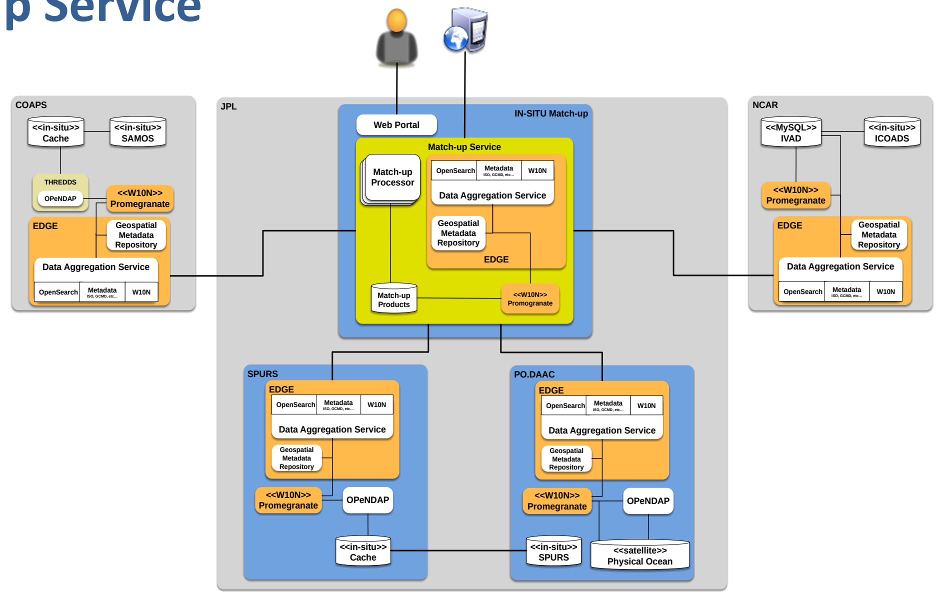
NCAR: ICOADS





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



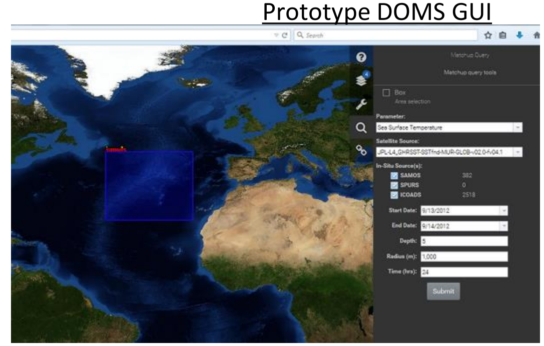
- DOMS interfaces: Web-service & Form-based querying with flexible filtering
  - by instrument, sensor, parameter
  - spatio-temporal domain & matchup window criteria (x,y,z,t)

Example DOMS metadata query URI

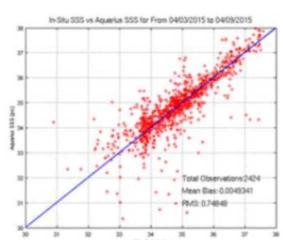
SPURS Data Service
- Sample query: `curl "https://deepdata.jpl.nasa.gov/ws/search/spurs?startTime=2012-08-01T00:00:00Z&endTime=2013-10-31T23:59:59Z&bbox=-45,15,-30,30"`
- Processed datasets
  - SPURS1_CTD
  - SPURS1_ADCP
  - SPURS1_ARGO
  - SPURS1_SEAGLIDER
  - SPURS1_MOORING_WHOI
- 13,817,392 documents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

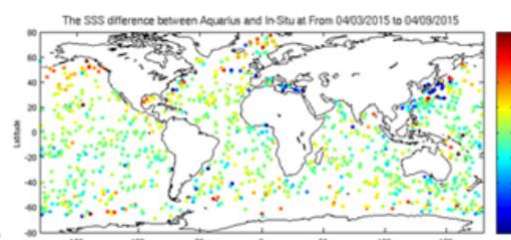
Satellite Data Service
- Sample query: `curl -X GET 'https://deepdata.jpl.nasa.gov/nexus/doms/datainbounds?ds=JPL-L4_GHRSST-SSTfnd-MUR-GLOB-v02.0-fv04.1&env=DEV&minLat23.93&maxLat=25.2&minLon=-38.74&maxLon=-37.33&startTime=1345280400000&endTime=1345366800000'`
- Dataset: GHRSST Multi-scale Ultrahigh 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