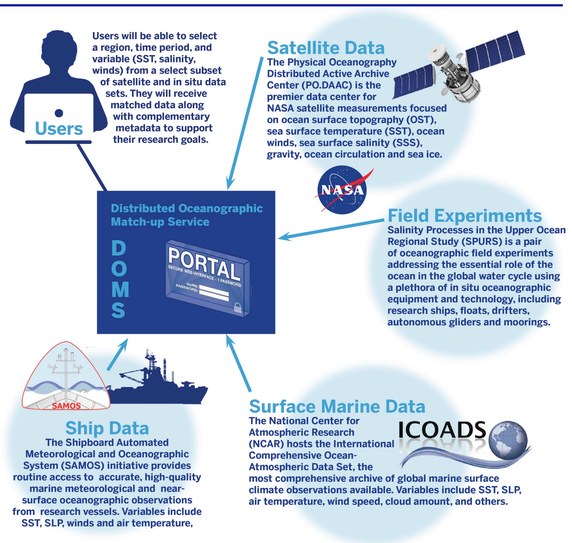
Developing a Distributed Oceanographic Match-up Service

Shawn R. Smith[1], Jocelyn Elya[1], Thomas Huang[2], Steve Worley[3], Vardis Tsontos[2], Benjamin Holt[2], Zaihua Ji[3], Adam Stallard[1], and Mark A. Bourassa[1,4]

[1]Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA
[2]Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA, USA
[3]National Center for Atmospheric Research, Boulder, CO, USA
[4]Earth, Ocean, and Atmospheric Science Department, Florida State University, Tallahassee, FL, USA
Contact: smith@coaps.fsu.edu

Overview

The Distributed Oceanographic Match-up Service (DOMS) is a collaborative effort between COAPS, NCAR, and NASA JPL. The service is currently in the prototype phase and scheduled to go online in late 2016. DOMS will reconcile satellite and in situ datasets in support of NASA's Earth Science mission. The service will provide a mechanism for users to input a series of geospatial references for satellite observations and receive the in situ observations that are matched to the satellite data within a selectable temporal and spatial domain. DOMS includes several characteristic in situ and satellite observation datasets. It will be used by the marine and satellite research communities to support a range of activities. The service is designed to provide a community-accessible tool that dynamically delivers matched data and allows the scientist to only work with the subset of data where the matches exist.

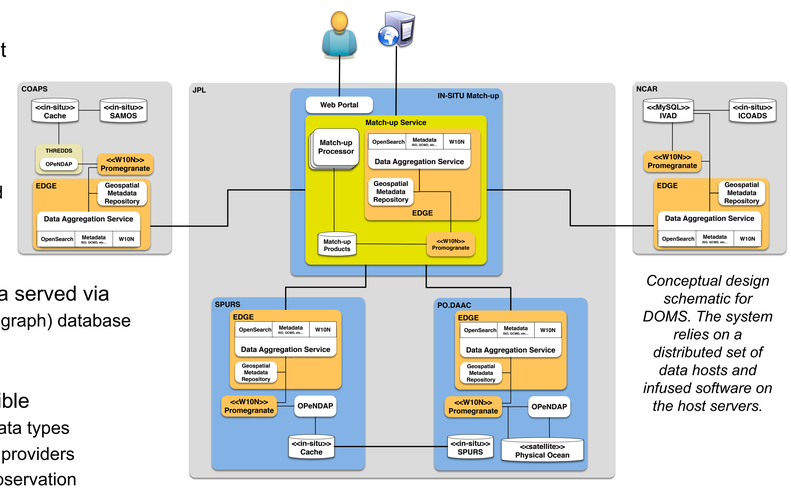


Why DOMS is Needed

- A wide user community seeks to match satellite to in situ observations to meet goals that include:
  - Satellite algorithm calibration, validation, and/or development
  - Decision support for planning future field campaigns
  - Investigations to support process studies, data synthesis, etc.
- The DOMS prototype will focus on algorithm cal/val activities.
- Presently, matched datasets are created using one-off programs that require satellite and in situ data to be housed on one's local computer.

DOMS Architecture

- The DOMS team is infusing common data access services at FSU, NCAR, and JPL.
  - Data indexing using Apache Solr
  - Extensible Data Gateway Environment (EDGE) – a data aggregation service that supports OpenSearch, metadata export, and W10N protocol
  - Pomegranate – an implementation of the W10N specification
- Hosts populating index from data served via
  - FSU – THREDDS and triple-store (graph) database
  - NCAR – MySQL database
  - JPL – No SQL database
- DOMS is designed to be extensible
  - Incorporate other oceanographic data types
  - Integrate data from additional data providers
  - Support match ups for terrestrial observation
  - Future matching between satellites and/or model datasets



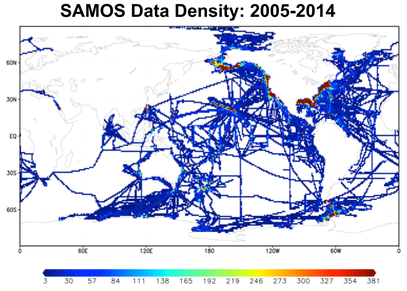
Conceptual design schematic for DOMS. The system relies on a distributed set of data hosts and infused software on the host servers.

In Situ Data Hosts

FSU: SAMOS

- Shipboard Automated Meteorological and Oceanographic System (SAMOS) initiative provides high-quality underway data from research vessels.
- Hosted at COAPS/FSU
- ~30 vessels participating in 2015
  - Vessels operated by WHOI, SIO, U. Hawaii, U, Washington, U. Alaska, BIOS, NOAA, USCG, USAP, IMOS, SOI, LUMCON
  - ~30-40K one-minute observations per month, per vessel

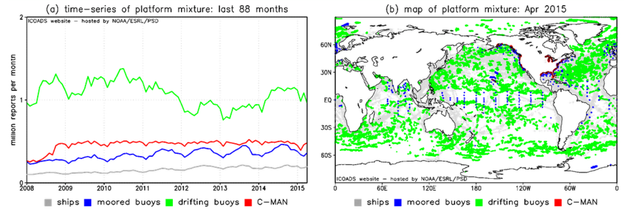
SAMOS Data Density: 2005-2014



Number of Observations per 2˚ bin

- Data include routine navigation (position, course, heading, speed), meteorology (wind, air temperature, humidity, pressure, rainfall, radiation), and oceanography (sea temperature and salinity)
- All data undergo scientific quality control

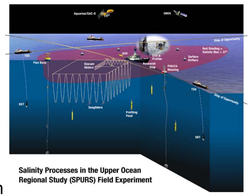
NCAR: ICOADS

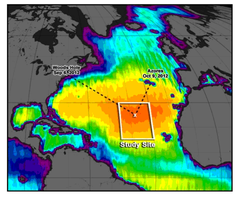


- Global coverage from ocean observing systems (~3M records/month)
  - VOS and research ships
  - Moored buoys: GTMBA and national systems
  - Drifting buoys: surface and ARGO
- Percent of ocean coverage varies by parameter
- Common parameters: SST, sea level pressure, air temperature, winds, humidity, clouds, evaporation
- Updated monthly with NCEP + NCDC GTS data streams
- Each record has UID and observing system tracking metadata

JPL: SPURS

- NASA-funded oceanographic field campaigns/science salinity process studies:
  - SPURS-1: N. Atlantic (2012-2013): salinity max region
  - SPURS-2: Eastern Equatorial Pacific (2016-2017): high precipitation/low evaporation region
- DOMS will select data from SPURS-1 campaign
- Advanced sampling technologies deployed in a nested design within a 900 x 800-mile^2 study area centered at 25˚N, 38˚W
- SPURS-DMP converted 15 natively heterogeneous formats to NCEI NetCDF standard
  - Archived at the PO.DAAC, http://podaac.jpl.nasa.gov/spurs

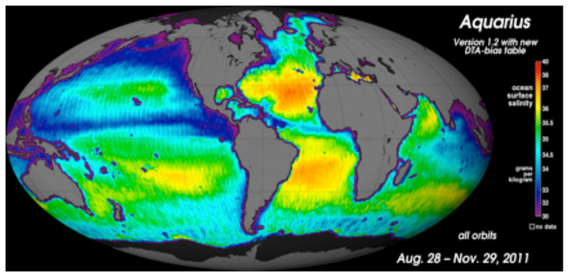




Top – SPURS-1 "sensor web".
Bottom – SPURS-1 sampling area

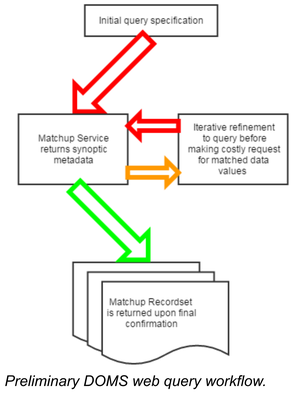
Satellite Data Host

- Satellite data will be hosted by the Physical Oceanography Distributed Active Archive Center (PO.DAAC).
- DOMS prototype will use:
  - Aquarius L2 v3.0 100 km – Sea surface salinity
  - ASCAT L2 25 km – Wind speed and direction
  - MODIS L2 P 1 km + MUR SST 1 km daily – Sea surface temperature
- Prototype will explore match ups to both swath and gridded datasets.



Planned User Interface

- DOMS will provide a web portal interface for users to browse and to submit match-up requests interactively.
  - To be hosted at JPL
  - Interface will allow users to "test/evaluate" searches by returning metadata only, creating visualizations, and then follow with a full matched dataset.
  - Will use flexible filtering and query specification based on indexed search criteria
- Additionally, DOMS will provide an underlying web service interface for machine-to-machine match-up operations to enable scalable data processing by external applications and services.
  - Tools will be provided to aid users in developing proper syntax for web service queries.



Preliminary DOMS web query workflow.

Indexed Search Criteria

- User queries will be facilitated by indexing (via Solr) the following
  - Parameter to match – salinity, sea temperature, or winds
  - Date and time range – ISO 8601 UTC
  - Horizontal domain – latitude and longitude box
  - Vertical domain above/below sea level (200 and 20 m limits, respectively)
  - Data source
  - Platform (ship, mooring, satellite, glider, etc.)
  - Device (CTD, bathythermograph, radiometer, scatterometer, etc.)
  - Mission (Aquarius, ASCAT, MODIS, SAMOS, etc.)
  - Data quality flag for each matching parameter (method TBD)
- Users will also specify spatial and temporal tolerances for locating a match (e.g., within 3 hours and 50 km)

Other Design Considerations

- Technical Challenges
  - Ensuring that the match-up algorithms perform with sufficient speed to return desired information to the user
  - Performing data matches using datasets that are distributed on the network
  - Returning actual observations for the matches [e.g., salinity] with sufficient metadata so the value difference can be properly interpreted.
- Interoperability
  - Mapping indexed values to community controlled vocabularies
    - CF standard parameter names
    - SeaVox platforms and SeaDataNet devices

Acknowledgments

The DOMS project is supported via NASA's Earth Science Technology Office from the Advanced Information Systems Technology program. The project is funded at FSU (lead institution, NNX15AE29G), NCAR (NNX15AG22G), and JPL via individual grants to each partner.