

# The Distributed Oceanographic Match-up Service

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



<sup>1</sup>Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA

<sup>2</sup>Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA, USA

<sup>3</sup>National Center for Atmospheric Research, Boulder, CO, USA

<sup>4</sup>Earth, Ocean, and Atmospheric Science Department, Florida State University, Tallahassee, FL, USA

Contact: jelya@coaps.fsu.edu



## Overview

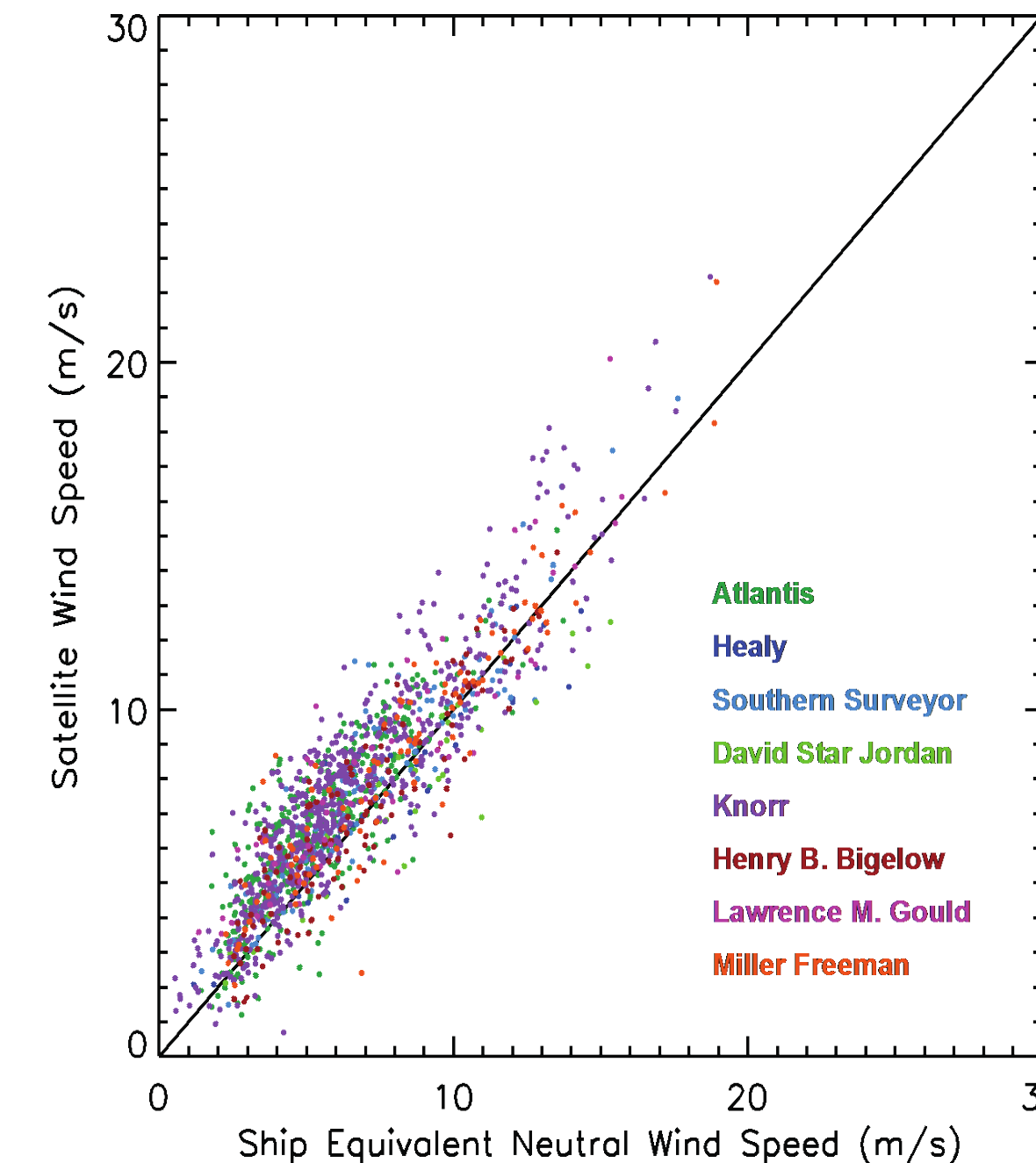
A two-year collaborative effort is underway to develop a distributed oceanographic data match-up service that will support web-based user queries to collocate in situ and satellite observations.

- Partners include:
  - Center for Ocean-Atmospheric Prediction Studies, Florida State University (COAPS, FSU)
  - Jet Propulsion Laboratory (JPL), California Institute of Technology
  - National Center for Atmospheric Research (NCAR)
- Distributed Oceanographic Match-up Service (DOMS)
  - Cloud computing match-up service
  - Host at JPL where it can be physically close to the PO.DAAC archive
  - Publically accessible web portal and web service endpoints
  - Generalized satellite-in situ data collocation web service
  - Distributed data sources: Initially 3 providers affiliated with the project and select in situ and satellite datasets

Provider/Collaborator	In situ Data Collection	Satellite Data Collection (swath & gridded)
NCAR	ICOADS/ARGO	-
FSU COAPS	SAMOS	-
PO.DAAC	SPURS	Aquarius L2 SSS, ASCAT L2 winds, L4 MUR SST, MODIS L2 SST

## Why DOMS is Needed

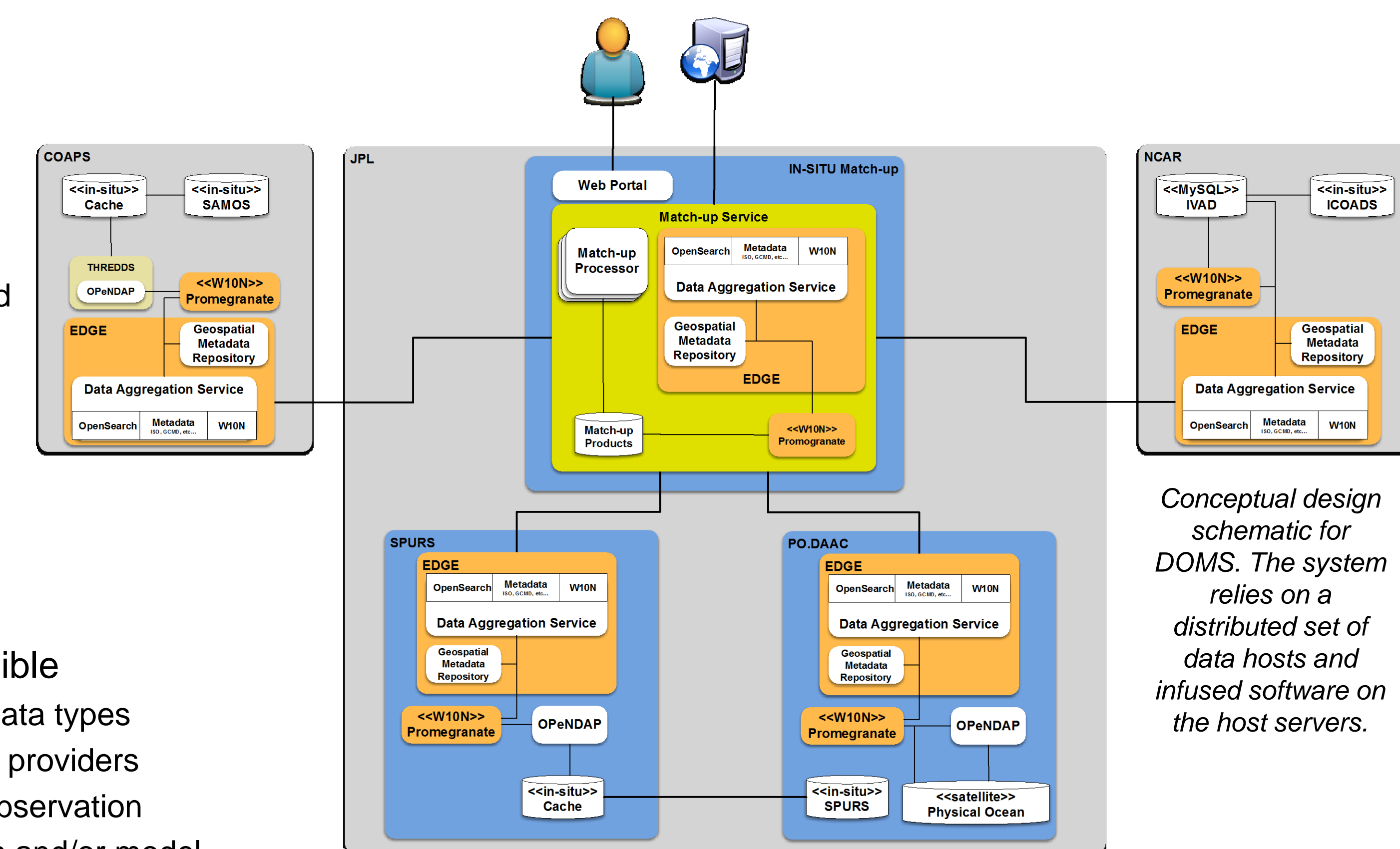
- A wide user community seeks to match satellite to in situ observations to meet goals that include:
  - Satellite algorithm calibration, validation, and/or development
  - Decision support for planning future field campaigns
  - Investigations to support process studies, data synthesis, etc.
- The DOMS 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.



Example of satellite product validation comparing winds from research vessels to winds from scatterometers. Ships provide necessary auxiliary parameters (air and sea temperature, pressure, humidity) and instrumental metadata (sensor height) to adjust ship observations to standard 10 m equivalent neutral wind.

## DOMS Architecture

- DOMS will infuse common data access services at FSU, NCAR, and JPL.
  - Extensible Data Gateway Environment (EDGE) – a data aggregation service that supports OpenSearch, metadata export, and W10N protocol
  - Pomegranate – an implementation of the W10N specification
- Prototype will test searches across data stored using THREDDS and SQL, NoSQL, and graph databases.
- 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

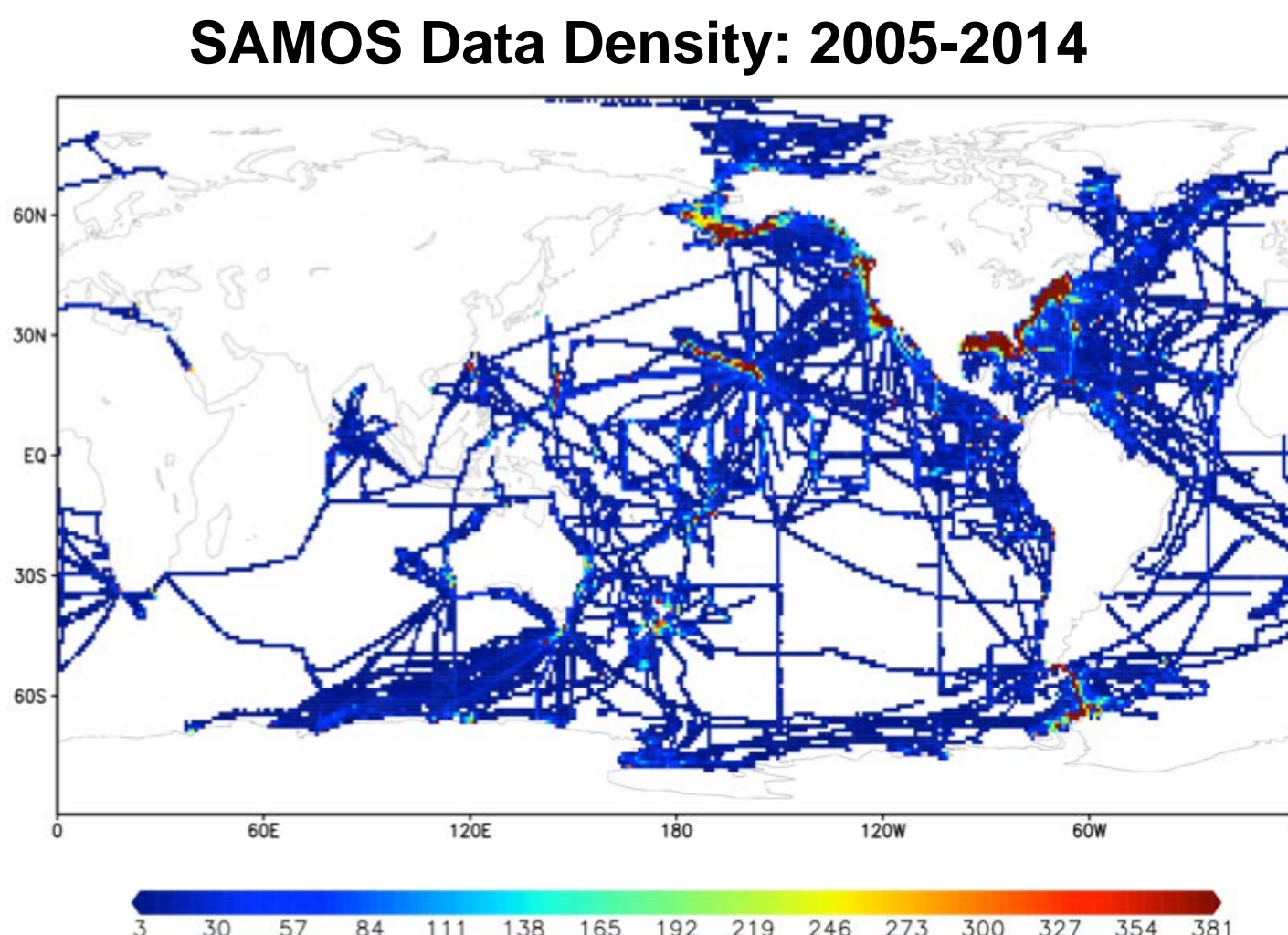


Conceptual design schematic for DOMS. The system relies on a distributed set of data hosts and infused software on the host servers.

## In Situ Data Hosts

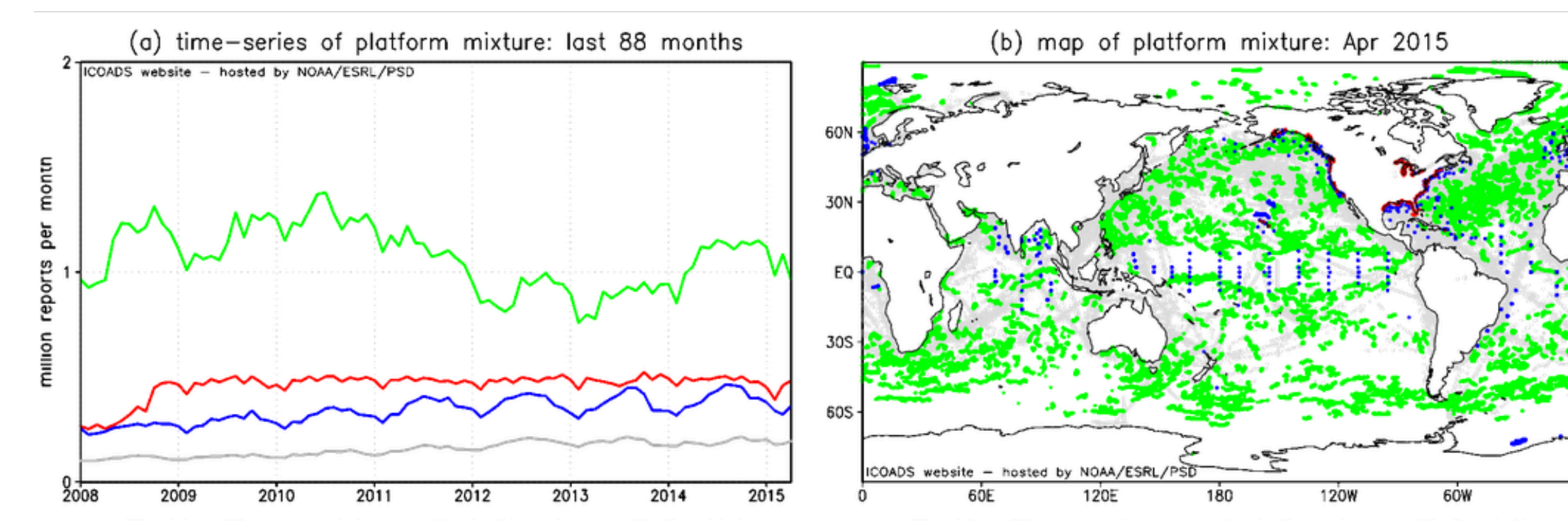
### FSU: SAMOS

- 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
- 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



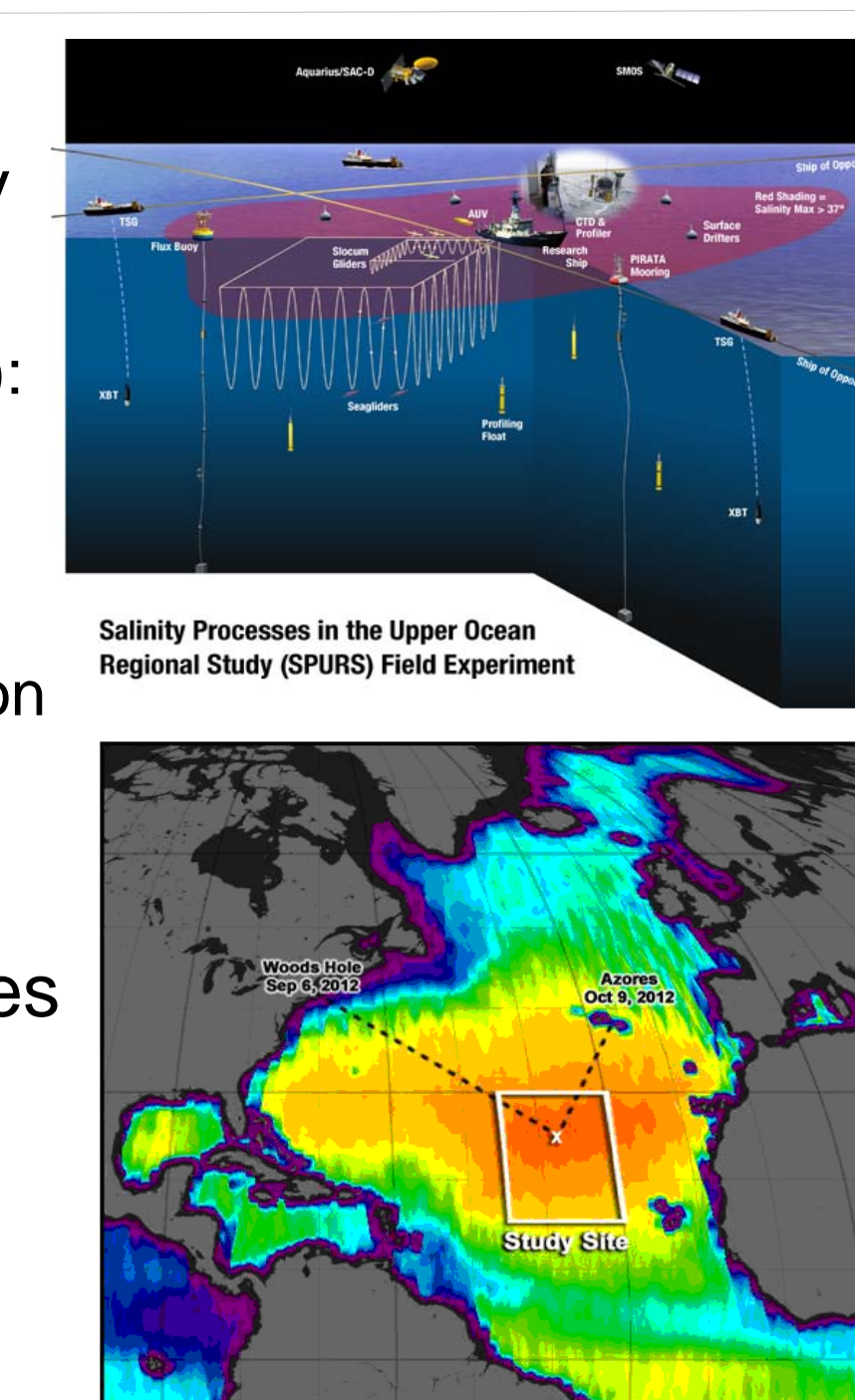
### NCAR: ICOADS

- 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

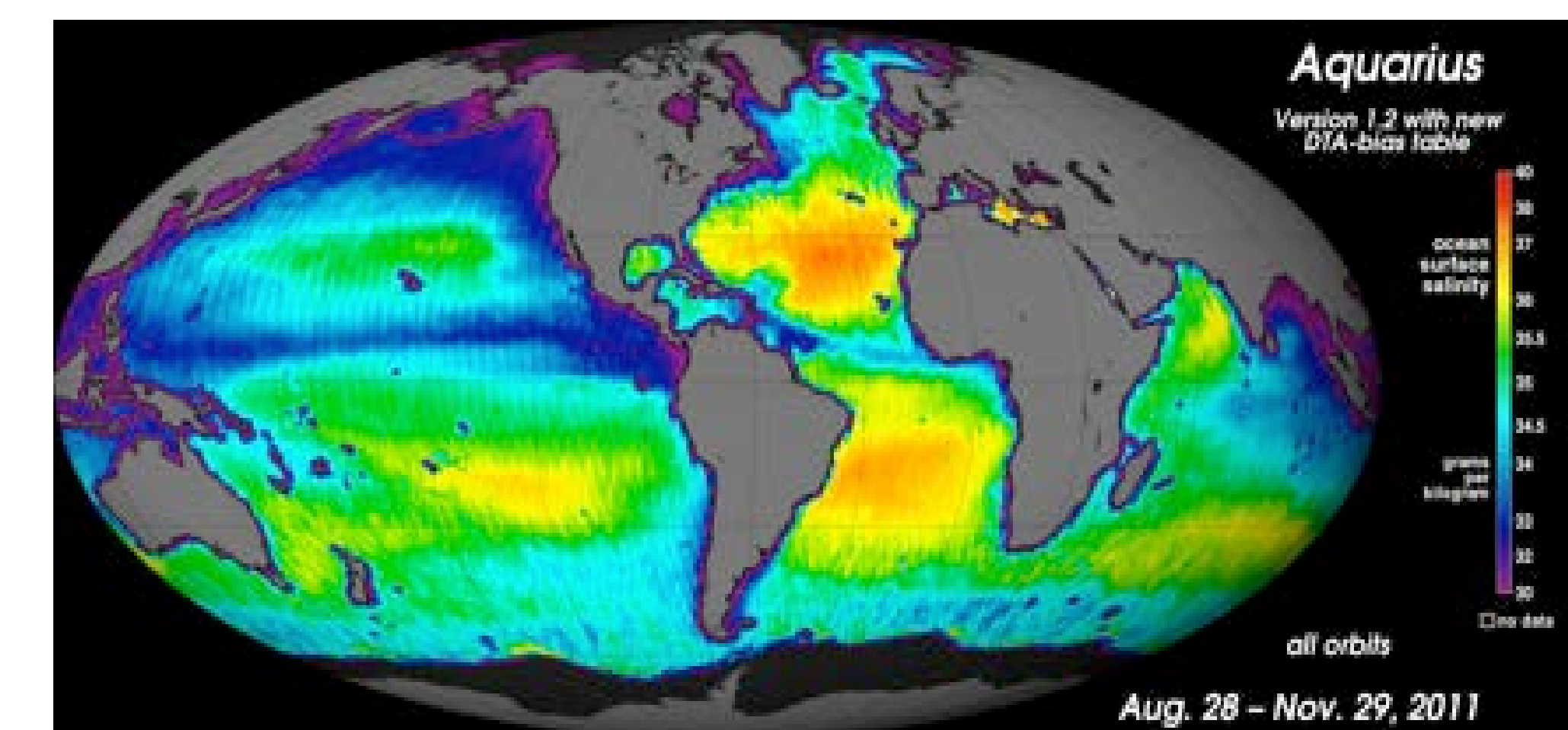
- NASA-funded oceanographic field campaigns/science salinity process studies:
  - SPURS-1: N. Atlantic (2012-2013): salinity max region
  - SPURS-2: Eastern Equatorial Pacific (2016-2017): high precipitation/low evaporation region
- DOMS will select data from SPURS-1 campaign
- Advanced sampling technologies deployed in a nested design within a 900 x 800-mile<sup>2</sup> 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

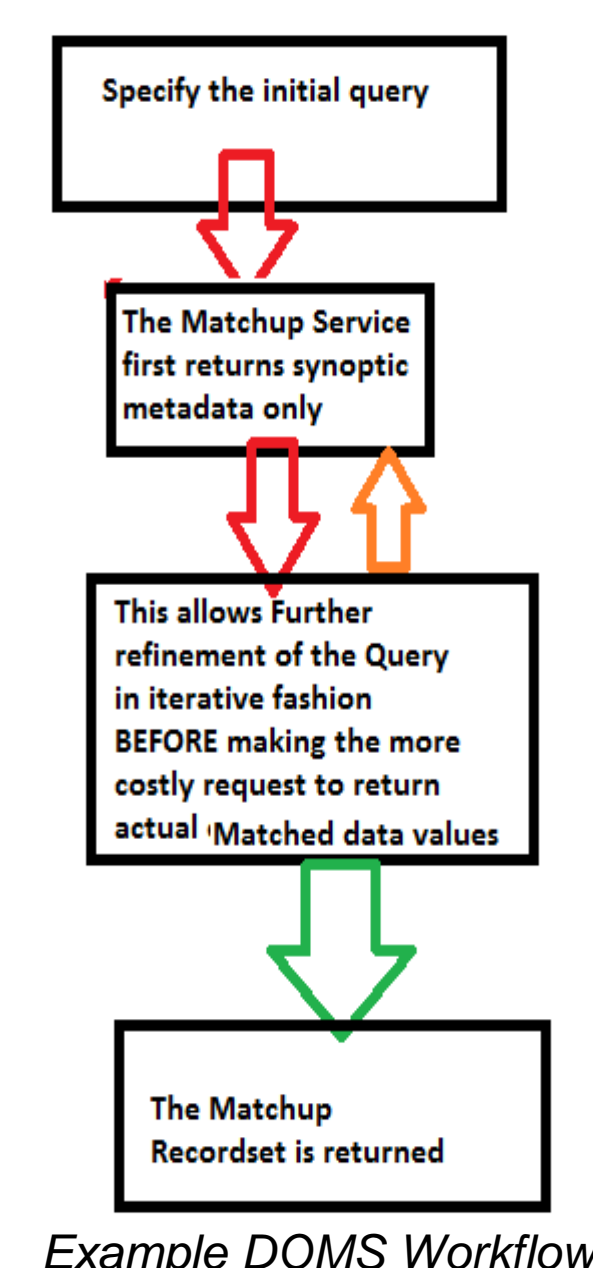
## Satellite Data Host

- 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



## User Interface

- DOMS will provide a web portal interface for users to browse and to submit match-up requests interactively.
  - Planned 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.
  - DOMS will provide flexible filtering and query specification by:
    - Instrument, sensor, parameter, provider
    - Match-up criteria: spatio-temporal domain (in x, y, z, t) and search radii/tolerances
- 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.



## Search Criteria

- Via the user interface or web service, the following options will exist to refine one's query:
  - 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 (constrained in prototype to ~ +/- 20 m)
  - Data source (e.g., which satellite vs. which in situ datasets)
  - Spatial and temporal tolerance for locating a match (e.g., within 3 hours and 50 km)
- Since most datasets used by DOMS will also have quality control flags, the system is being designed to:
  - Provide data filtered by the host using documented analysis of QC flags as a default
  - Allow the user to the option to receive all data, regardless of QC flags

## 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.

## 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, grant number NNX15AE29G), JPL, and NCAR via individual grants to each partner.

