

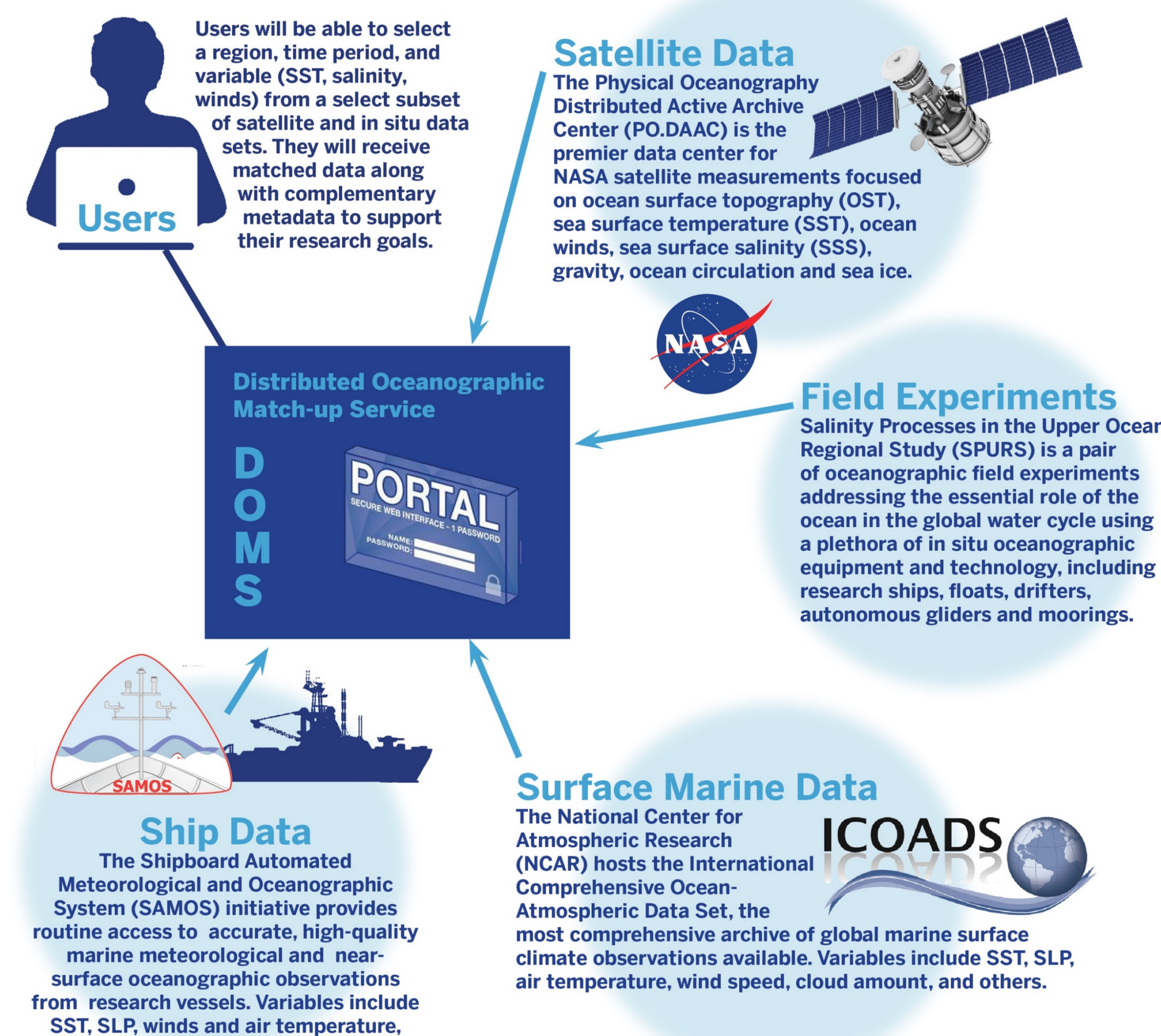
DOMS: A Prototype Service to Match Satellite and In Situ Datasets

Shawn R. Smith¹, Jocelyn Elya¹, Adam Stallard¹, Mark A. Bourassa^{1,4}, Thomas Huang², Vardis Tsontos², Benjamin Holt²,
Nga Quach², Kevin Gill², Frank Greguska², Steven Worley³, and Zaihua Ji³

¹Center for Ocean-Atmospheric Prediction Studies, Florida State University, Tallahassee, FL, USA, ²Jet Propulsion Laboratory/California Institute of Technology, Pasadena, CA, USA,
³National Center for Atmospheric Research, Boulder, CO, USA, ⁴Earth, Ocean, and Atmospheric Science Department, Florida State University, Tallahassee, FL, USA
Contact: srsmith@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 went online in August, 2016. DOMS reconciles satellite and in situ datasets in support of NASA's Earth Science mission. The service provides 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 search 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 scientists to only work with the subset of data where the matches exist.

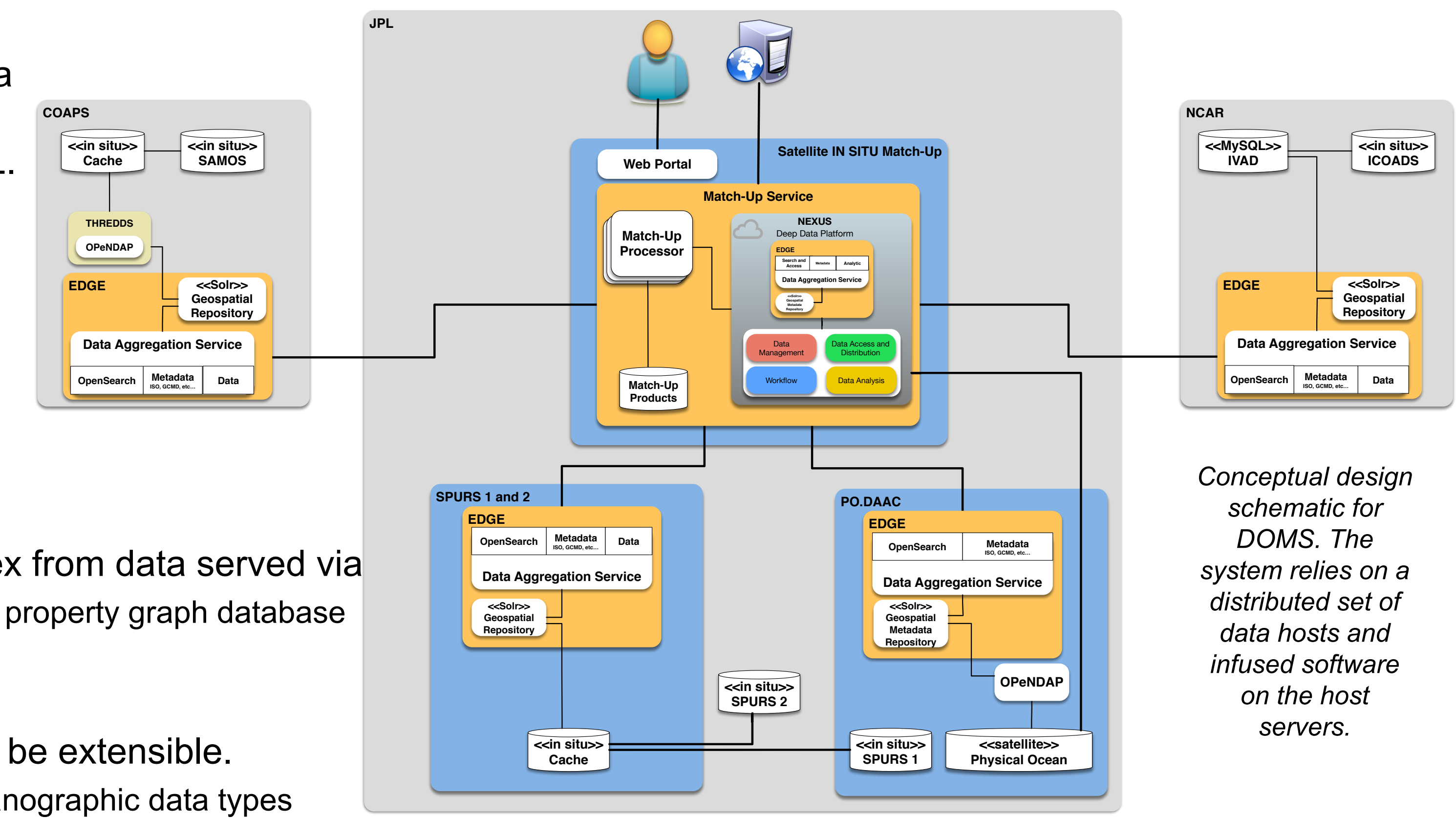


Why DOMS is Needed

- 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.
- DOMS eliminates the need for one-off match-up programs that require satellite and in situ data to be housed on one's local computer

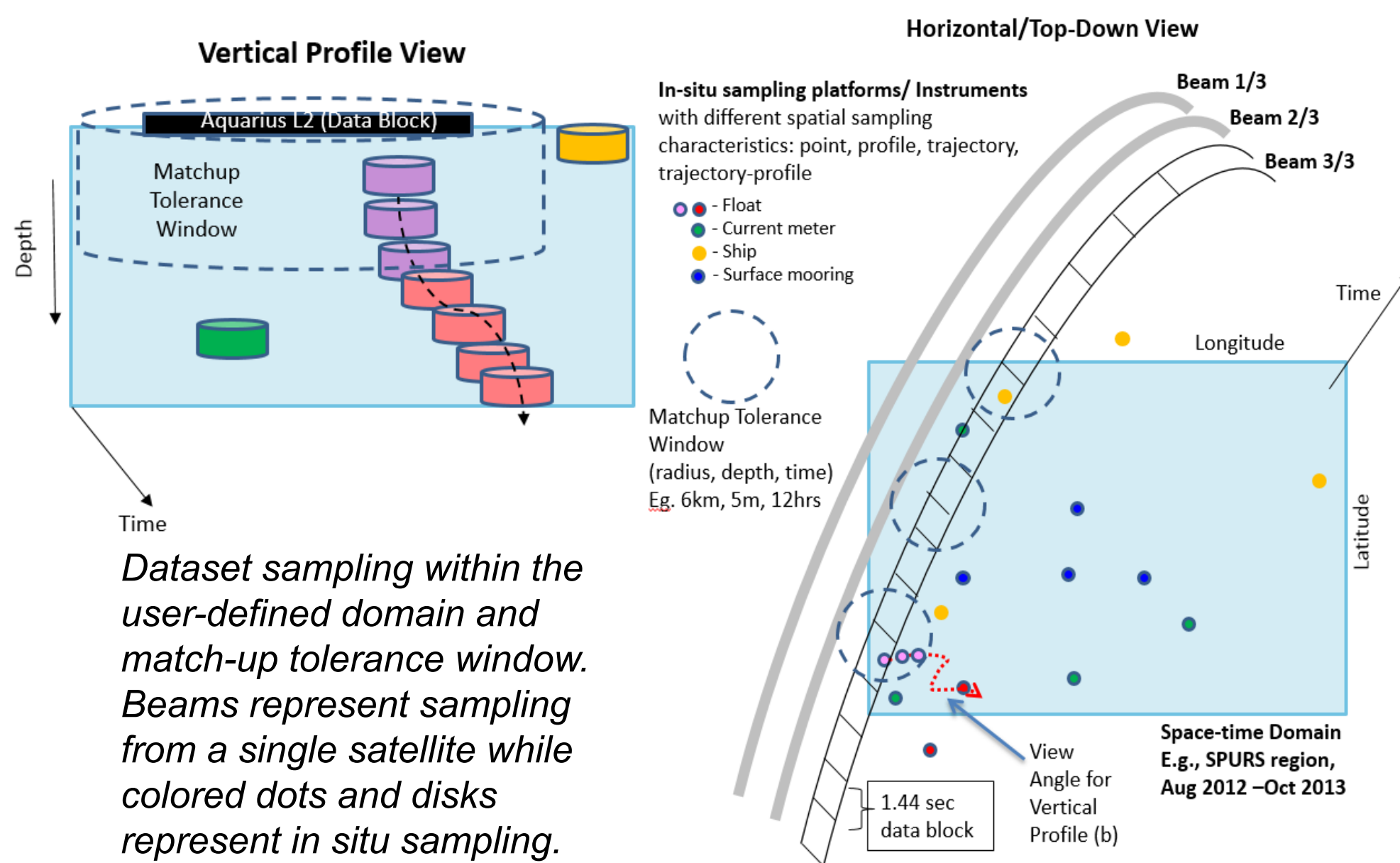
DOMS Architecture

- 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 and metadata export
- Hosts populating index from data served via
 - FSU – THREDDS and property graph database
 - NCAR – MySQL
 - JPL – NoSQL
- DOMS is designed to be extensible.
 - Incorporate other oceanographic data types
 - Integrate data from additional data providers
 - Support match ups for terrestrial observations
 - 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.

Search Domain & Match-Up Tolerances



- Queries facilitated by indexing the following in Solr
 - Parameter – salinity, sea temperature, or winds
 - Temporal search domain – ISO 8601 UTC
 - Horizontal search domain – latitude/longitude box
 - Vertical search domain above/below sea level
 - Data source
 - Satellite: JPL SMAP L2B v2.0 salinity, ASCAT L2 Coastal 12.5 km winds, MUR L4 1 km daily sea surface temperature
 - In situ: ICOADS Release 3.0, SAMOS, SPURS-1, 2
 - Platform type (ship, orbiting satellite, etc.)
 - Device type (CTD, current profiler, radiometer, etc.)
 - Mission (Aquarius, ASCAT, MODIS, SAMOS, etc.)
 - Data quality flag (simplified mapping)
- Users will also specify spatial and temporal match-up tolerances for locating a match (e.g., within 1 hours and 30 km)

Prototype User Interface

- Web portal for users to browse and to submit match-up requests interactively
 - 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
- Web service interface for machine-to-machine match-up operations to enable scalable data processing by external applications and services.

Prototype Steps

- Select Satellite Source
- Select Date Range
- Select Depth Range
- Select Match-Up Tolerances (Space + Time)
- Draw Bounding Box

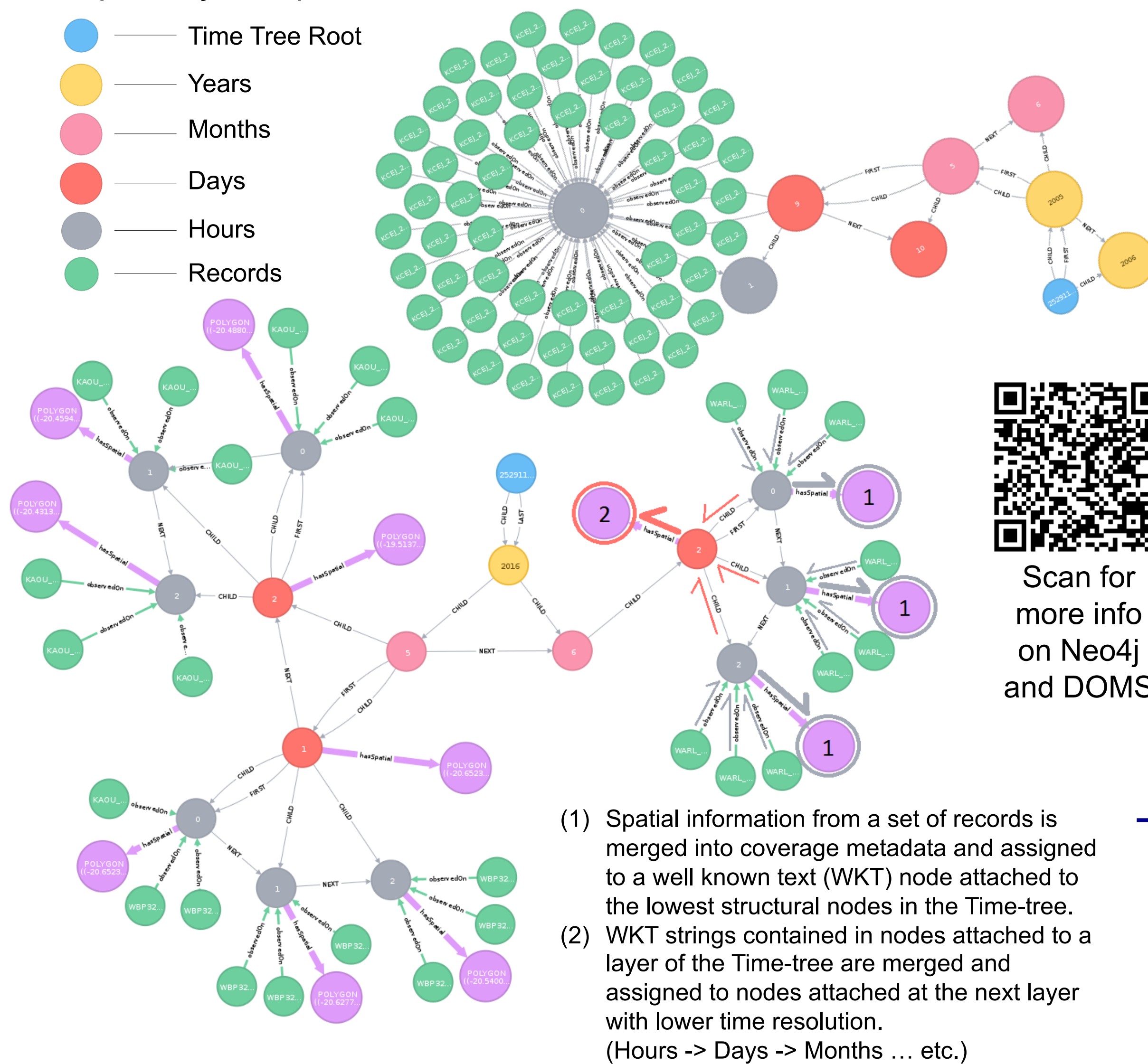
Step 5 triggers search of in situ data in box and displays counts

Once in situ counts are displayed and mapped (as shown above) user triggers data matching algorithm with **Submit** button. Resulting matched data are provided in tabular and graphical displays (below). Data can be downloaded as CSV or netCDF-CF files.

Source	Time	Lat	Lon	Depth (m)	SST	SSS	Wind Speed	Wind Direction
SMAP_L2B_SSS	2015-07-04 08:04:51	28.220	-92.604	0.000	0.000	32.160	0.000	0.000
SMAP_L2B_SSS	2015-07-02 08:29:17	28.445	-92.090	0.000				
samos	2015-07-02 07:30:00	28.510	-92.010	0.000				
samos	2015-07-02 07:31:00	28.510	-92.010	0.000				
samos	2015-07-02 07:32:00	28.500	-92.010	0.000				
samos	2015-07-02 07:33:00	28.500	-92.010	0.000				
samos	2015-07-02 07:34:00	28.500	-92.010	0.000				
samos	2015-07-02 07:35:00	28.500	-92.010	0.000				

Exploring Graph Databases

- The goal of this exploration is to compare query performance between a graph database and Solr for DOMS.
- Since graph technologies are well suited for the semantic web, we aim to describe the data in a way that is well suited for discovering matched data, in hopes of semantically linking SAMOS to the other data sets in the DOMS network.
- The Neo4j Time Tree plugin by GraphAware is used to model a time based hierarchy directly into the graph. The graph supports discovery of temporally bound metadata via an agglomerative clustering-like process that can be used to facilitate search, especially for spatial information.



Future Plans

- At the end of a 2-year development cycle, DOMS will achieve a technology readiness level (TRL) of 4.
- To achieve an infusion-ready TRL6, the team plans to
 - Stress test DOMS to support large-scale data queries (e.g., whole globe or entire satellite mission),
 - Develop delayed-mode data delivery protocols,
 - Implement filtering using data quality information, and
 - Package and document DOMS for infusion.
- Further enhancement of DOMS capabilities are also being considered, including the following:
 - Supporting satellite-to-satellite and in situ-to-in situ data matching,
 - Supporting satellite/in situ to numerical model matching, and
 - Including additional high-priority science datasets.
 - Exploring alternate uses of graphs in DOMS such as a storage mechanism for matched data, a global match table (possibly for raw data or metadata, or even both), or as an output graph.

Acknowledgements

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), JPL, and NCAR (NNX15AG22G) via individual grants to each partner.