Integrating the Distributed Oceanographic Match-Up Service into OceanWorks

**Overview**

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 focuses on algorithm calibration/validation.

- 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.

**Architecture of the DOMS Prototype**

- The prototype infuses common data access services at FSU, NCAR, and JPL.
- Data indexed using Apache-SOLR
- Extensible Data Gateway Environment (EDGE) - a data aggregation and query service that supports OpenSearch and metadata export
- In-situ data are indexed from data served via FSU - THREDDS
- JPL - NoSQL
- Satellite data are tiled and indexed using the NEXUS deep data platform.

**Integrating DOMS into OceanWorks**

- First development cycle for DOMS completed in spring 2017 with 2nd prototype at a technology readiness level (TRL) of 4.

- Through the OceanWorks project (started in September 2017), the team plans to achieve a TRL-6 by
  - Building new capability to support large-scale data queries enabling scalable data processing by external applications and services.
  - Implementing filters using data quality information, and
  - Building new capability to support large-scale data queries (e.g., satellite to in situ data matching, presently supporting DOMS at FSU)

- Integrating DOMS into OceanWorks via deployment to the AISt Managed Cloud Environment.

- Ongoing architectural changes to DOMS include
  - Enhancements to the data match-up capability
  - Harvesting the SOLR-indexed in situ data from the remote data hosts using the OpenSearch interface.

- These changes should eliminate the on-the-fly data movement and network stall overheads.

- They also will make the code ready for implementation in a cloud environment.

**Vision for the Future**

- Under OceanWorks, DOMS can address big data approaches to finding (1) problems with data used in calibration (satellite-to-in situ and satellite-to-satellite), and (2) physical relationships in the observed variables. For example,
  - Algorithm development depends highly on low-observations from moorings; however, in high winds moorings are riding very large waves and are suspected to report non-representative winds when the mooring is in the wave trough. Quantifying this problem requires collocating large numbers of satellite to in situ data.
  - Diurnal warming links SSTs, winds, and chlorophyll among other variables. To understand the physical relationships requires matching multiple data parameters across a range of observing platforms.

- Further DOMS enhancements being considered include the following:
  - Supporting satellite-to-satellite and in situ-to-in situ data matching,
  - Supporting satellite/in situ to numerical model matching,
  - Including additional high-priority science datasets.

**Acknowledgements**

- The DOMS project is supported via NASA's Earth Science Technology Office from the Advanced Information Systems Technology program. Separate awards presently supporting DOMS at FSU (NNX15AD09G), NCAR (NNX15AG22G), and JPL.

**Search Domain, Match-Up Tolerances**

- A GUI supports browsing and submitting match-up requests interactively
- Allows users to “test/evaluate” searches by returning metadata only, creating visualizations, and then follow-up with a full matched dataset
- Uses flexible filtering and query specification based on indexed search criteria

**Prototype User Interface**

- 1st Prototype Steps
  1. Select Satellite Source
  2. Select Date Range
  3. Select Depth Range
  4. Select Match-Up Tolerances (Space & Time)

- 2nd GUI provides separate screens to select primary and secondary data along with filter criteria. Changes recommended by ESP testbed external review.

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