

# CISL's Data Support Section to collaborate on "Big Data" project

BY BRIAN BEVIRT  
03/05/2015

As the volume of data being produced by research instruments increases, new methods are required to validate, manage, and use it. CISL's Data Support Section is part of a new multi-institution collaboration that will produce scientifically meaningful ocean data products for current and future research efforts. This project involves contributors from Florida State University (FSU), NASA's Jet Propulsion Laboratory (JPL), and NCAR. The project team leaders include Shawn R. Smith, lead investigator, and Mark A. Bourassa from FSU's Center for Ocean-Atmospheric Prediction Studies, Thomas Huang (co-PI), Benjamin Holt, and Vardis Tsontos from JPL, and Steven Worley (co-PI) from NCAR CISL.

The project is titled "A Service to Match Satellite and *In-situ* Marine Observations to Support Platform Inter-comparisons, Cross-calibration, Validation, and Quality Control." Its goal is to address the problem of matching large volumes of marine observations made by Earth-orbiting satellites to those from instruments on ocean-based observing platforms. Funded by the NASA Science Mission Directorate under the Advanced Information Systems Technology (AIST) program, this project will develop software tools and implement a big-data management infrastructure to improve the accessibility of ocean data.

The project will support NASA's Earth science mission by providing a web-based portal for users to input geospatial references (date, time, location) for satellite observations and receive the corresponding observations from ships, buoys, floats, and other observing platforms. Users will also be able to input positions of ocean-based platforms and receive the corresponding satellite observations. The first version of this service will provide match-up services for three types of data: sea surface temperature, sea surface salinity, and winds over the ocean. Data matching is necessary because the various instruments collect different types of data in different spatial patterns at different times.

We all rely on weather forecasts to plan for our daily activities and to be prepared for severe weather events, and most weather forecasts are developed using information from computer models. A primary data source used by these models – especially over the oceans – comes from space-based satellites. Since satellites are in orbit and not in physical contact with the ocean, they make measurements "remotely," and these remote measurements must be evaluated using data from instruments that are in contact with the ocean. The better the correlation between observations from satellites and *in situ* platforms, the more confidence we have in the data used by the models and in the weather forecasts provided to the public.

The two-year \$1.2M NASA award supports the FSU-JPL-NCAR collaborators in developing software tools to match observations collected by all these instruments and make the service available via a web portal. Other researchers will then use these tools to calibrate and validate the satellite data and perform quality control on the observations. The project will also create a two-function user interface (a web portal for human interactions and web services for machine-to-machine interoperability) that allows data subsets to be selected, matched as needed for each request, then delivered along with the metadata needed for scientific interpretation.

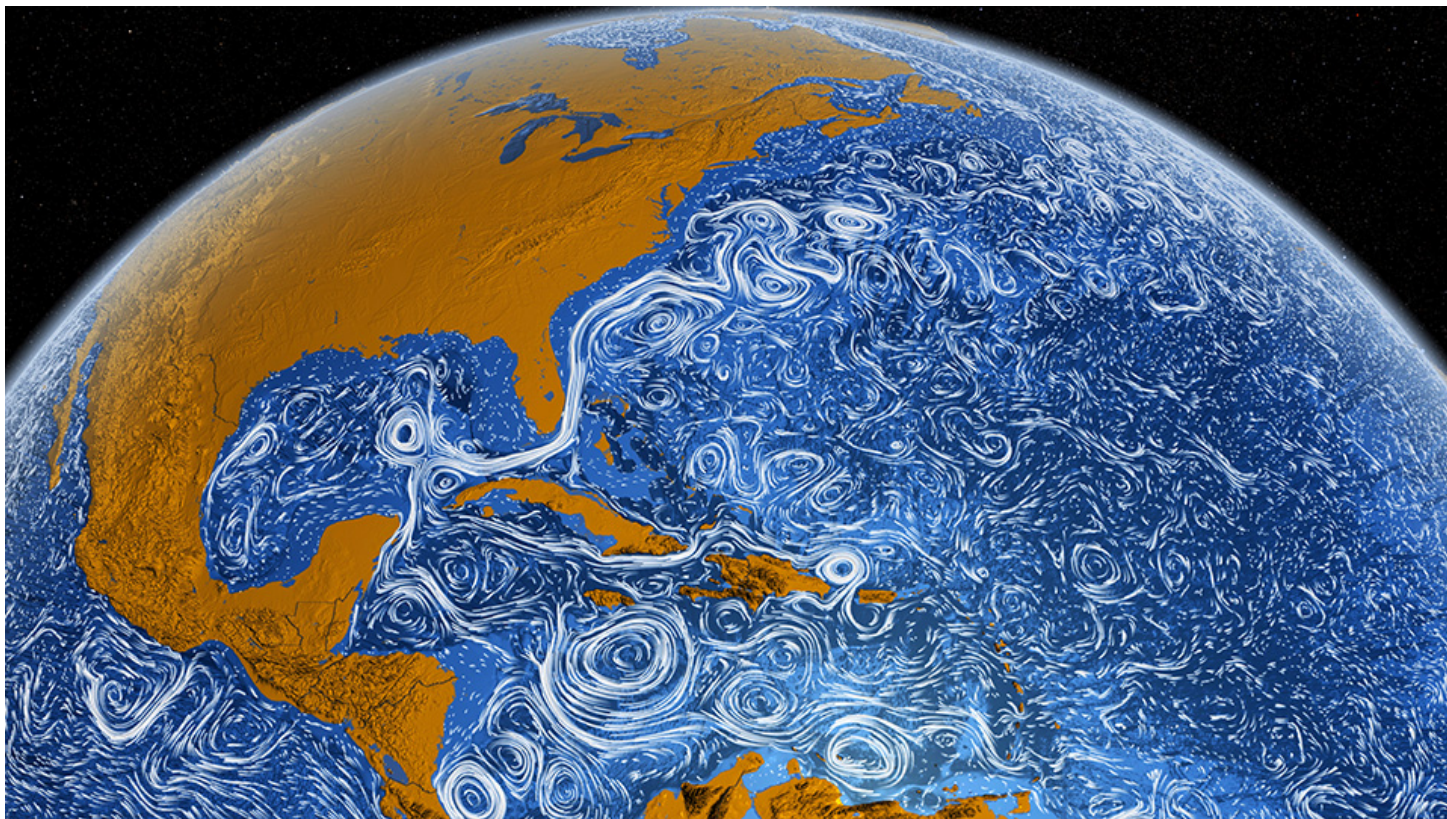
FSU, as lead institution, will manage the project, provide research ship-based observations for the match-up service, and report progress to the NASA funding agency. The JPL team will develop and host the user interface and provide both satellite and experimental data for the match-up service. NCAR will provide curated historical weather and ocean datasets for the match-up service. Participants from all three institutions will develop middleware to link the user interface to the datasets and conduct the spatial, temporal, and parameter data matching. Finally, all collaborators will also design and test the user interface and test the match-up service.

Although this service will be developed and initially tested on only three types of data typical of oceanographic field campaigns, the collaborators will develop a generic match-up service that can be ported and used widely by the Earth Science community. The service will provide a web portal interface for web users to browse and submit match-up requests interactively. It will also have an underlying web service interface for machine-to-machine match-up operations, thus enabling external applications and services. Results provided by the service will include both reconciled metadata (addressing the veracity of the data) and geophysical data based on flexibly applied spatial, temporal, and parameter criteria.

Typically, match-up utilities developed for specific field campaigns are proprietary, non-scalable, highly customized solutions employing a mixture of technologies. Most research efforts match observations by downloading large volumes of data to local computers using software they develop for their particular mix of instrument data. In general, these customized solutions employ technologies that are not easily sharable or scalable. Repetition of this process consumes substantial human and financial resources.

This new collaboration addresses these limitations with a generalized match-up service architecture and infrastructure that can support NASA missions and Earth science applications. A comprehensive, standards-based, open, scalable, and sustainable match-up service for remotely sensed and *in situ*

data will ultimately save time and costs for many agencies and research teams. For data provided through this service, researchers will no longer need to develop customized data-matching software, freeing up their personnel resources to focus on the scientific questions that can be answered using matched datasets. Further, providing file-level data access via the web will lower the technological barriers to entry for future contributors.




Big data makes detailed visualizations like this possible. This image is one frame from a **2011 NASA animation** that shows global ocean surface currents from June 2005 to December 2007. The animation was created using data from NASA satellites, direct ocean measurements, and a numerical model jointly developed by NASA JPL and the Massachusetts Institute of Technology. To enable realistic descriptions of how ocean circulation evolves over time, data was synthesized at resolutions that can resolve ocean eddies and other current systems that transport heat and carbon throughout the oceans. This work is helping researchers quantify the ocean's role in the global carbon cycle, understand the recent evolution of the polar oceans, and monitor time-evolving heat, water, and chemical exchanges between different components of the Earth System. The work being undertaken by the new FSU–JPL–NCAR collaboration will improve data management and usage for ocean research at increasingly higher resolutions. —Image courtesy of NASA/Goddard Space Flight Center Scientific Visualization Studio

[Contact Us](#)

[CISL Organization](#)

[For Staff](#)

© 2015 UCAR | [Privacy Policy](#) | [Terms of Use](#) | [Copyright Issues](#) | [Sponsored by NSF](#) | [Managed by UCAR](#) | [Webmaster/Feedback](#)  
Postal Address: P.O. Box 3000, Boulder, CO 80307-3000 • Shipping Address: 3090 Center Green Drive, Boulder, CO 80301

 The National Center for Atmospheric Research is sponsored by the National Science Foundation. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[Login](#)