

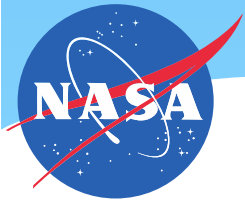
# Global Data Assembly Center (GDAC) Report to the GHR SST Science Team

Edward Armstrong, Jorge Vazquez, Rob Toaz, Thomas Huang,  
Cynthia Chen, Chris Finch

16<sup>th</sup> GHR SST Science Team Meeting

ESA/ESTEC, The Netherlands

20 July 2015



National Aeronautics and  
Space Administration

**Jet Propulsion Laboratory**  
California Institute of Technology  
Pasadena, California

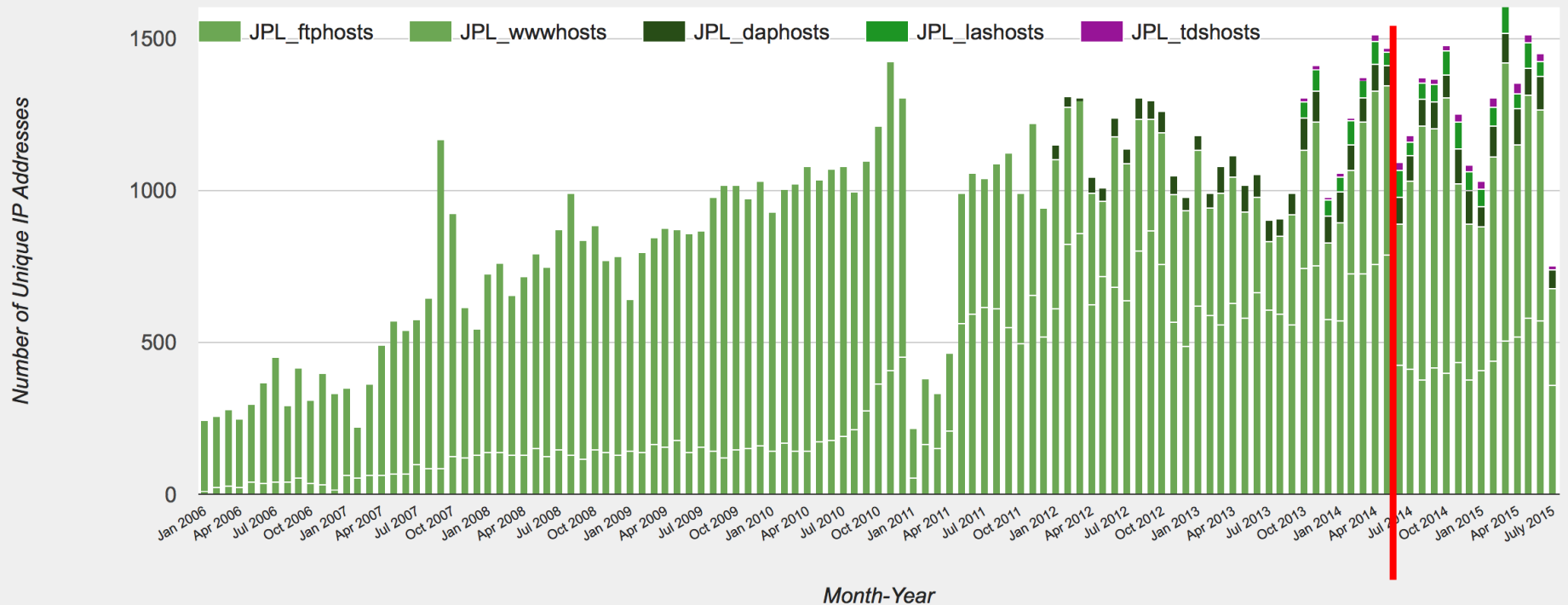
© 2015 California Institute of Technology.  
Government sponsorship acknowledged.

# GDAC Highlights

- \* 2014-2015 focused on ingesting and distributing more GDS2 datasets
  - \* 28 new GHRSSL2P/L3/L4 datasets ingested
    - \* New sensors: AMSR2, VIIRS, IASI
    - \* Data lifecycle policy implemented for new GDS2 datasets
  - \* Maintaining existing GDS1 datastreams
    - \* New NASA version of ocean MODIS products since July 2015 (v2014.0)
    - \* MODIS L2P still in GDS1 format. Expected reprocessing to GDS2 in Fall 2015
  - \* Support operational datastreams for L2P/L3/L4 data from 14 RDACs
  - \* Maintain linkages to data providers and LTSRF archive
- \* Develop and improve tools and services for data usage presented last year
  - \* Web services, Subsetting, Visualization, Data Aggregation, Metadata services
- \* User community engagement: Responded to 100+ GHRSSL user queries
- \* Reports in DAS-TAG on new information technologies in development

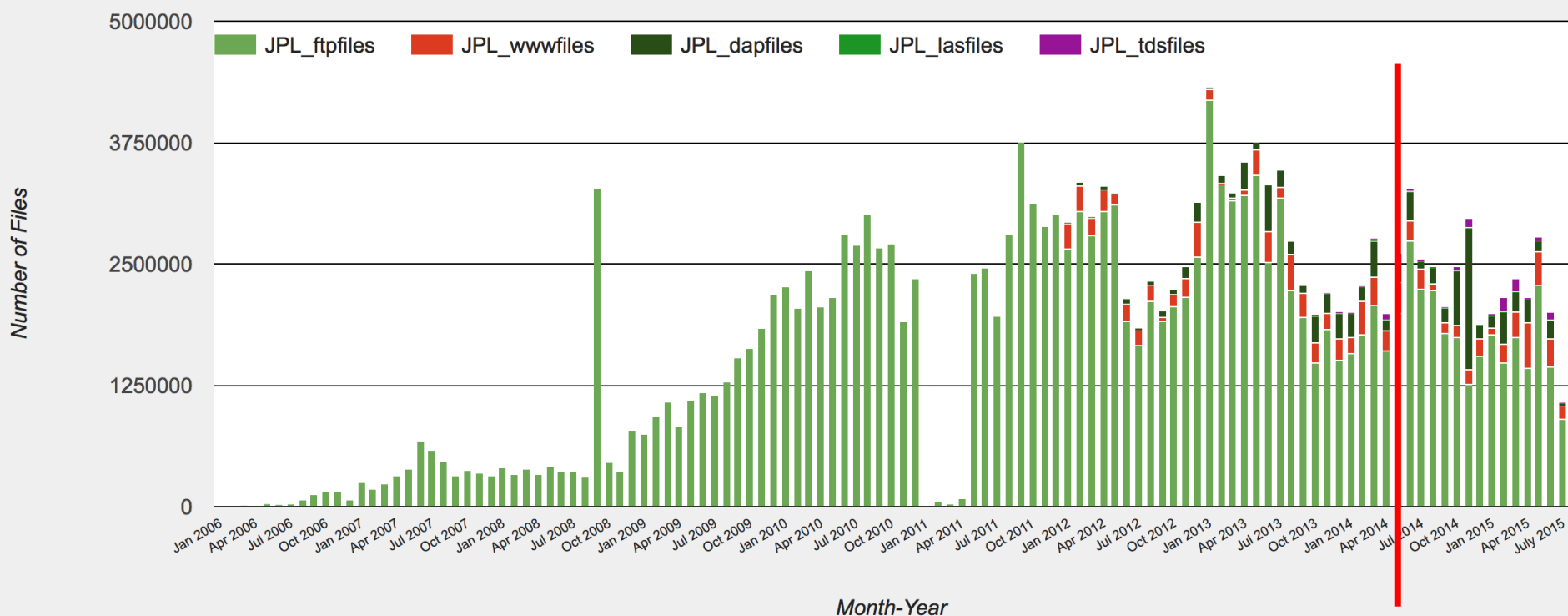
# PO.DAC Distribution metrics: Monthly Unique Users

**GHRST: Unique Hosts Served by JPL**



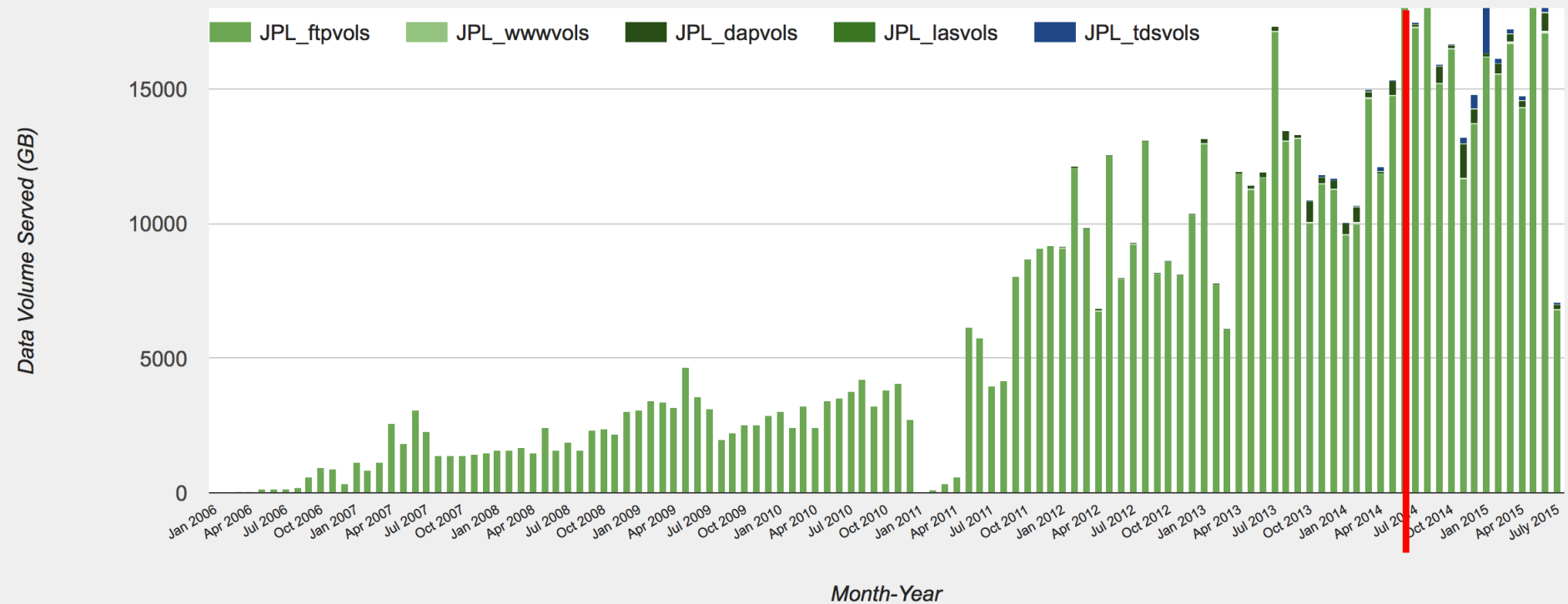
# Number of Monthly Files

**GHRSSST: Number of files served by JPL**



# Volume of Monthly Files

**GHR SST: Data Volumes Served at JPL**



# Tool summary

- \* **SOTO2D:** visualization including GHRSSST MODIS L2P, Windsat L3, G1SST L4
- \* **PO.DAAC Web Services:** search, discovery, metadata, extract as “chained” services
- \* **HiTIDE:** GUI based L2 subsetting
- \* **Webification:** Arbitrary data store exposed as URLs. Subsetting by value.
- \* **Coastal Marine Discovery Service:** GIS interface for satellite data
- \* **Datacasting:** RSS Informed earth science data availability
- \* **Live Access Server (LAS)** for L3/L4 subsetting
- \* **Metadata Compliance Checker:** Granule level metadata checks

# Metadata Compliance Checker (MCC)

- \* Extends open source software developed by the NOAA IOOS program (python wrapped)
  - \* Improved and simplified for updating
  - \* Checks:
    - \* Climate and Forecast (CF) Metadata Conventions v1.6
    - \* Attribute Conventions for Dataset Discovery (ACDD) 1.1 & 1.3,
    - \* GHRST GDS2 data model
    - \* Local files or remote OPeNDAP URLs
  - \* Web interface
  - \* Retains command line and API interface
  - \* Quantitative output scoring in HTML, PDF and JSON
  - \* Available via <http://podaac.jpl.nasa.gov/uat/mcc>

# MCC web menu and output



## Compliance Checker

netCDF File 5.00 GB max

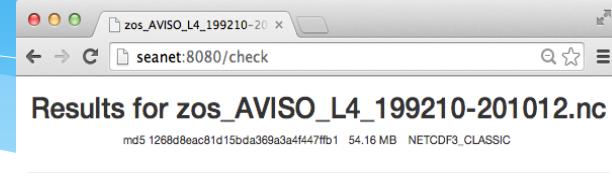
Local File [OPeNDAP URL](#)

No file chosen

### Checkers select one or more

- ☐ ACDD (Attribute Convention for Data Discovery)
  - Description:** These conventions identify and define a list of NetCDF global attributes recommended for describing a NetCDF dataset to discovery systems such as Digital Libraries. Software tools will use these attributes for extracting metadata from datasets, and exporting to Dublin Core, DIF, ADN, FGDC, ISO 19115 etc. metadata formats.
  - URL:** [http://wiki.esipfed.org/index.php?title=Attribute\\_Convention\\_for\\_Data\\_Discovery\\_1-1&oldid=45515](http://wiki.esipfed.org/index.php?title=Attribute_Convention_for_Data_Discovery_1-1&oldid=45515)
  - Version:** 1.1
- ☐ CF (netCDF Climate and Forecast Metadata Conventions)
  - Description:** The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, regridding, and display capabilities. The CF conventions generalize and extend the COARDS conventions.
  - URL:** <http://cfconventions.org/Data/cf-conventions/cf-conventions-1.6/build/cf-conventions.html>
  - Version:** 1.6
- ☐ GDS2 (Group for High Resolution Sea Surface Temperature Data Specification, Version 2)
  - Level:**
  - Description:** The GHRST Data Specification (GDS) Version 2.0 is a technical specification of GHRST products and services.
  - URL:** <https://www.ghrsst.org/files/download.php?m=documents&f=121009233443-GDS20r5.pdf>
  - Version:** Version 2, Revision 5

[API](#) [Code](#) [PODAAC](#)



## ACDD Check 25 out of 99 passed

### Global Attributes 5 out of 50 passed

#### Highly Recommended 1 out of 4 passed

##### keywords all 2 failed

✗ check for existence failed because "summary" does not exist

- Description:** A paragraph describing the dataset

✓ check for existence passed because "title" exists

- Description:** A short description of the dataset.
- Value:** Obs-AVISO model output prepared for obs4MIPs NASA-JPL observation

#### Recommended 4 out of 32 passed

#### Suggested all 14 failed

### Variable Attributes 20 out of 49 passed

#### Highly Recommended 20 out of 49 passed

## CF Check 78 out of 90 passed

### Compliance Checker 78 out of 90 passed

#### Variable names all 7 passed

#### all\_features\_are\_same\_type all 0 passed



# New Technologies in Development

- \* **Virtual Quality Screening Service (VQSS)**
  - \* Seamlessly applying GDS2 quality information (quality\_level, l2p\_flags, etc.) to granule data extraction and subsetting requests
  - \* See poster tonight
  - \* Participate in GHRSSST data user survey lead by Mike Chin
- \* **OceanXtremes**
  - \* SST anomaly detection and mining using cloud based databases
- \* **Distributed Oceanographic Matchup Service (DOMS)**
  - \* Satellite to *in situ* (ICOADS, SPURS, ARGO, SAMOS) matchup service
- \* **Mining and Utilizing Dataset Relevancy from Oceanographic Dataset (MUDROD)**  
Metadata, Usage Metrics, and User Feedback to Improve Data Discovery and Access
  - \* Improving data search relevancy (finding the right datasets)
- \* **Picasso**
  - \* Client for granule visualization with georeferencing and data interrogation
- \* Will be further discussed in DAS-TAG breakout and poster session

# GHRSSST GDS2 datasets

- \* Public release of 28 GDS2 datasets with dataset lifecycle adherence
  - \* Consistency with treatment of data insured. Distribution, tools, services
  - \* Still need to improve quality documentation with assistance of provider
- \* Working with providers to insure data and metadata compliance
- \* Making new datasets available as soon as possible
  - \* Access via FTP/OPeNDAP before rest of lifecycle policy implemented

# Status of GDS2 datasets

- \* L2P datasets for AMSR2, AMSRE, TMI, VIIRS, AVHRR (N-19, MetOP-A and B), GOES, MTSAT, SEVERI, IASI
- \* L3C/U datasets for AVHRR, VIIRS, GOES, SEVIRI
- \* L4 datasets for MUR, OSTIA, CMCo.2deg, DMI\_OI

# Summary

- \* Many new GDS2 datasets online, discoverable, available via tools and services
  - \* Small number of GDS1 datasets remaining (e.g., MODIS, NCDC)
- \* PO.DAAC continues to improve tools and services implemented for subsetting, discovery, dataset and granule web services.
  - \* New tool for metadata validation
- \* Further JPL technology development has implications for GHRSSST data and users (see DAS-TAG presentation and VQSS poster)
- \* Issues for consideration:
  - \* Improving dataset lifecycle quality descriptions.
  - \* Dataset relevance (for 93 GHRSSST datasets). What can be done to improve this?
  - \* Dataset retirement of GDS1 or underutilized datasets
  - \* Discuss the Regional/Global Task Sharing re architecture proposal by Ken Casey

# Backup

# MCC API

## GET Example

```
$ curl http://seanet:8080/check?checkers=acdd&url-upload=https://github.com/ioos/compliance-checker/raw/master/compliance_checker/tests/data/2dim-grid.nc
```

```
{
  "fn": "2dim-grid.nc",
  "md5": "e7b1b61e7233ebad0a7114b9441b05df",
  "model": "NETCDF4",
  "results": [
    {
      "description": "These conventions identify and define a list of NetCDF global attributes recommended for describing a NetCDF dataset to discovery systems such as Digital Libraries. Software tools will use these attributes for extracting metadata from datasets, and exporting to Dublin Core, DIF, ADN, FGDC, ISO 19115 etc. metadata formats.",
      "hash": "2524994885125638958",
      "name": "Attribute Convention for Data Discovery",
      "passed": 18,
      "results": [
        {
          "hash": "4800022785557752635",
          "name": "Global Attributes",
          "passed": 0,
          "results": [
            ...
          ]
        }
      ]
    }
  ]
}
```

# Redesigned Web portal



# Webification

## Open specification:

<http://w10n.org>

## Summary:

- Resource is viewed as a tree of nodes and leaves.
- They have semantic URLs, accessible through HTTP.
- Meta info exchange format is JSON, by default.
- Full ReSTful style request/response. Read/Write.

## Disciplines:

Earth science (NetCDF, HDF 4/5, GRIB)

Planetary Science (VICAR/PDS)

Astronomy (FITS) and more



# Use Case – Quality filtering the SST observations

- \* Subset a L2P granule (by value!)
  - \* [http://host:port/path/2013/123/20130503-MODIS\\_T-JPL-L2P-T2013123065500.L2\\_LAC\\_GHRSSST\\_N-v01.nc.bz2/sea\\_surface\\_temperature\[-130<lon<-120,35<lat<45\]?output=format](http://host:port/path/2013/123/20130503-MODIS_T-JPL-L2P-T2013123065500.L2_LAC_GHRSSST_N-v01.nc.bz2/sea_surface_temperature[-130<lon<-120,35<lat<45]?output=format)
- \* Apply quality filter
  - \* [http://host:port/path/2013/123/20130503-MODIS\\_T-JPL-L2P-T2013123065500.L2\\_LAC\\_GHRSSST\\_N-v01.nc.bz2/sea\\_surface\\_temperature\[quality\\_flag>=4\]?output=format](http://host:port/path/2013/123/20130503-MODIS_T-JPL-L2P-T2013123065500.L2_LAC_GHRSSST_N-v01.nc.bz2/sea_surface_temperature[quality_flag>=4]?output=format)
- \* Quality filter, wind screen, subset all in one step !
  - \* [http://host:port/path/2013/123/20130503-MODIS\\_T-JPL-L2P-T2013123065500.L2\\_LAC\\_GHRSSST\\_N-v01.nc.bz2/sea\\_surface\\_temperature\[quality\\_flag>=4,wind\\_speed>6,-130<lon<-120,35<lat<45\]?output=format](http://host:port/path/2013/123/20130503-MODIS_T-JPL-L2P-T2013123065500.L2_LAC_GHRSSST_N-v01.nc.bz2/sea_surface_temperature[quality_flag>=4,wind_speed>6,-130<lon<-120,35<lat<45]?output=format)
- \* Virtual Quality Screening Service
  - \* NASA funded technology project to implement quality screening web service for GHRSSST and SMAP data

# Top Datasets in 2014-2015

## **FTP**

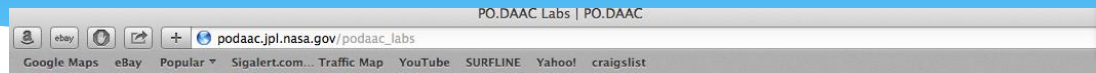
- \* GHR SST Level 4 MUR Global Foundation Sea Surface Temperature Analysis
- \* GHR SST Level 4 OSTIA Global Foundation Sea Surface Temperature Analysis
- \* MetOp-A ASCAT Level 2 12.5km Ocean Surface Wind Vectors

## **OPeNDAP**

- \* GHR SST Level 4 MUR Global Foundation Sea Surface Temperature Analysis
- \* OSCAR third degree resolution ocean surface currents
- \* Cross-Calibrated Multi-Platform Ocean Surface Wind Vector L3.0 First-Look Analyses

# Tools – PO.DAAC Labs

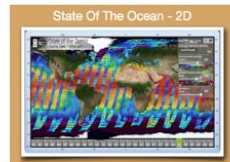
## [podaac.jpl.nasa.gov/podaac\\_labs](http://podaac.jpl.nasa.gov/podaac_labs)



### PO.DAAC Labs

**Explore New Ideas, Prototypes and Tools.**  
**Offer feedback directly to the engineers who developed them.**

You can reach us by email at: [podaac@podaac.jpl.nasa.gov](mailto:podaac@podaac.jpl.nasa.gov)



#### State of the Ocean 2D (SOTO 2D)

State of the Ocean 2D provides near real-time data layers (vector and image) that are visualized in an HTML5 interface utilizing open-source tools such as Leaflet.js. SOTO 2D data layers are annotated to give contextual descriptions of the ocean's features and events, and kml overlays (ice extent, hurricane tracks, clouds). Some layer implementations include Sea Surface Temperature (SST), wind vectors, and ocean current vectors.



#### PO.DAAC Web Services

PO.DAAC Web Services are application programming interfaces (APIs) that can be accessed through standard web protocols. The W3C defines a Web Service in part as, "A software system designed to support interoperable machine-to-machine interaction over a network," (for the full definition, see <http://www.w3.org/TR/ws-arch/#whatis>). The PO.DAAC Web Services use a Representational State Transfer (REST) model with calls issued over a Hypertext Transfer Protocol (HTTP) connection. On receipt of a request message, our services return the response in either an Extensible Markup Language (XML) structure or, optionally, a JavaScript Object Notation (JSON) format.



#### PO.DAAC HITIDE (Subsetter) v3.1.x

The High-level Tool for Interactive Data Extraction (HITIDE) is a web-based interface facilitating the search, imaging, and extraction of select Level 2 "swath" datasets from PO.DAAC's archive. HITIDE's v3.1.x user interface is powered by a set of Web Services which facilitate machine to machine interoperability.



#### Webification

Webification (W10n) is ReSTful Webservice technology providing simplified access to PODAAC data and metadata via HTTP/HTTPS protocols with URLs comprised of well-defined parameters. W10n supports major Earth science data formats like NetCDF and HDF 4/5, and abstracts an arbitrary data store as a hierarchical tree of nodes for each associated attributes which can be interrogated. Direct access to inner components of the tree is via HTTP requests from either a web browser, script or similar client. Results of W10n calls