# GLOBAL DATA ASSEMBLY CENTER (GDAC) REPORT TO THE GHRSST SCIENCE TEAM

**Edward Armstrong[(1)], Jorge Vazquez[(1)], Rob Toaz[(1)], Yibo Jiang[(1)], Thomas Huang[(1)], Cynthia Chen[(1)], Chris Finch[(1)], Vardis Tsontos[1)]**

*(1) Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA*
*Email: edward.m.armstrong@jpl.nasa.gov*

## ABSTRACT

In 2015-2016 the Global Data Assembly Center (GDAC) at NASA's Physical Oceanography Distributed Active Archive Center (PO.DAAC) continued its role as the primary clearinghouse and access node for operational GHRSST data streams, as well as its collaborative role with the NOAA Long Term Stewardship and Reanalysis Facility (LTSRF) for archiving.

## 1. Introduction

The summary accomplishments and milestones performed by the GDAC are noted below:

- Management of GHRSST data
  - Nearly all GHRSST datasets are now GDS2
  - Consistent monthly distribution near 30TBs
  - Supported operational datastreams for L2P/L3/L4 data from 15 RDACs
  - Maintained linkages to data providers and LTSRF archive
  - Coordinated with NASA ESDIS components on Sentinal-3A data
- Continual development and improvement of tools and services for data usage Web services, Subsetting, Visualization, Data Aggregation, Metadata services
- User community engagement
  - Responded to GHRSST user queries
  - Worked with applications users
  - Populated PO.DAAC forum with data recipes and tutorials
- Coordination activity on new Regional Global Task Sharing (R/G TS) architecture proposals
  - Goal: Decentralize the ingest and distribution locations
  - Focus on specific datasets and RDACs

## 2. Distribution metrics

The following figures show distribution metrics from the GDAC since 2005. Users, data volumes and number of files are all steady or have slightly increased. Users are leveraging interfaces and services such as OPeNDAP, THREDDS and LAS more so than in the past.

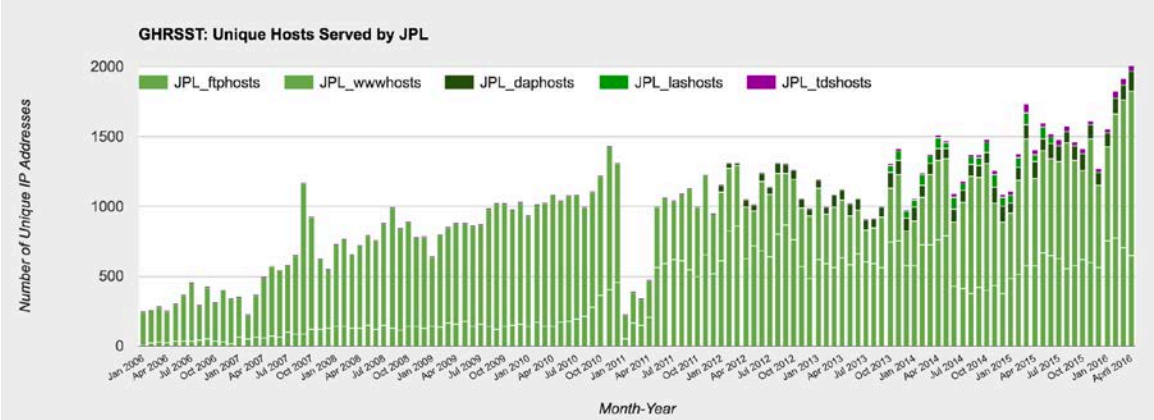


*Figure 1. Number of unique monthly users via FTP, OPeNDAP, LAS and THREDDS*

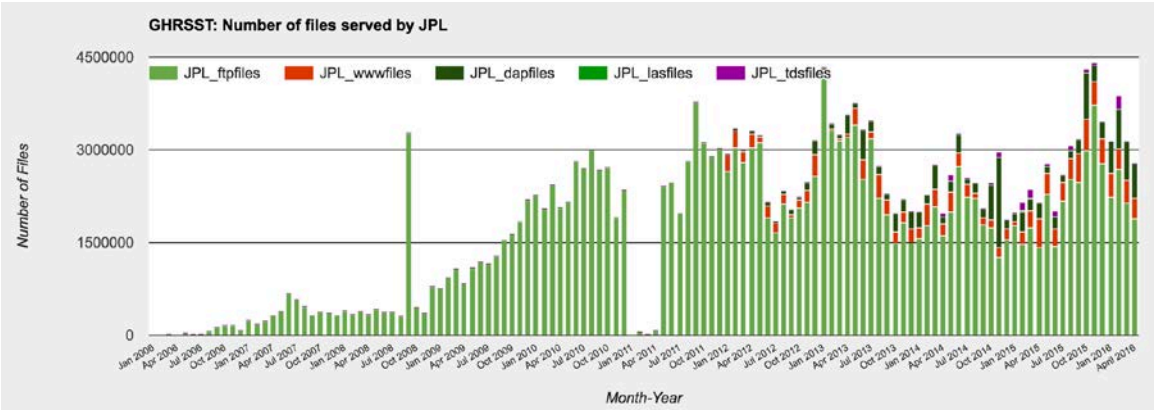


*Figure 2. Number of monthly files distributed*

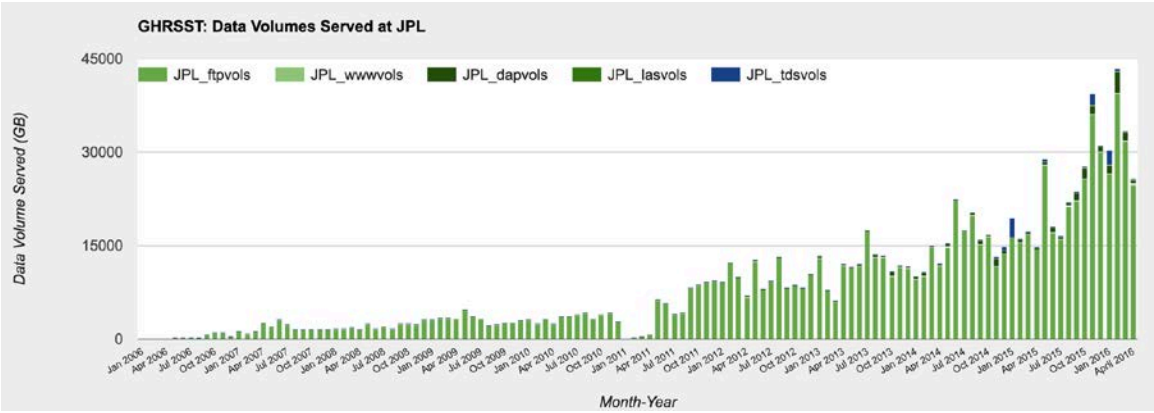


*Figure 3. Volume of monthly files distributed*

## 3. New and Emerging Technologies

- New version of HiTide L2 subsetter to be released Summer 2016
- PO.DAAC “Drive” will replace FTP access
- Virtual Quality Screening Service (VQSS)
    - Seamlessly applying GDS2 quality information (quality_level, l2p_flags, etc.) to granule data extraction and subsetting requests
- OceanXtremes
    - Climatology generation, 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)
    - Text and relevance mining of science literature
    - Coordinating with NASA Earth Science Data Systems Working Group (ESDSWG) on Search Relevance
        - Chairs: Ed Armstrong and Lewis Mcgibbney

## 4. Summary

- GHRSST GDS2 “catalog” near complete
    - datasets online, discoverable, available via tools and services
- PO.DAAC continues to improve tools and services implemented for subsetting, discovery, dataset and granule web services.
    - New interface “PO.DAAC Drive” for data download
    - L2 subsetting service (L2SS) and revised HiTide coming
    - Further JPL technology development has implications for GHRSST data and users (Armstrong et al. poster)
- Issues for consideration:
    - Regional/Global Task Sharing re architecture proposal (in DAS-TAG)
    - Improving access to quality information
    - Improving search relevance

**Acknowledgements**

This work was carried out at the NASA Jet Propulsion Laboratory, California Institute of Technology. Government sponsorship acknowledged. Copyright 2016 California Institute of Technology. Government sponsorship acknowledged.