



Approach to Managing MEaSUREs Data at the GSFC Earth Science Data and Information Services Center (GES DISC)

B. Vollmer, S. J. Kempler, H. K. Ramapriyan
NASA Goddard Space Flight Center, Greenbelt, MD, USA
Steven.J.Kempler@nasa.gov

Abstract

- A major need stated by the NASA Earth science research strategy is to "develop long-term, consistent, and calibrated data and products that are valid across multiple missions and satellite sensors." (NASA Solicitation for Making Earth System data records for Use in Research Environments (MEaSUREs) 2006-2010)
- Selected projects create long term records of a given parameter, called Earth Science Data Records (ESDRs), based on mature algorithms that bring together continuous multi-sensor data.
- ESDRs, associated algorithms, vetted by the appropriate community, are archived at a NASA affiliated data center for archive, stewardship, and distribution. See <http://measures-projects.gsfc.nasa.gov/> for more details.
- This presentation describes the NASA GSFC Earth Science Data and Information Services Center (GES DISC) approach to managing the MEaSUREs ESDR datasets assigned to GES DISC. (Energy/water cycle related and atmospheric composition ESDRs)
- GES DISC will utilize its experience to integrate existing and proven reusable data management components to accommodate the new ESDRs.
- Components include a data archive system (S4PA), a data discovery and access system (Mirador), and various web services for data access.
- In addition, if determined to be useful to the user community, the Giovanni data exploration tool will be made available to ESDRs.
- The GES DISC data integration methodology to be used for the MEaSUREs datasets is presented.
- The goals of this presentation are to share an approach to ESDR integration, and initiate discussions amongst the data centers, data managers and data providers for the purpose of gaining efficiencies in data management for MEaSUREs projects.

GES DISC Will Be Responsible for Archiving Approved Datasets Generated by 7 MEaSUREs PIs

PI	Start Date	End Date	Title
Chung-Lin Shie	5/08	5/11	Reprocessing of Goddard Satellite-based Surface Turbulent Fluxes (GSSTF) Data Set for Global Water and Energy Cycle Research
Richard McPeters	1/08	1/13	Creating a Long Term Multi-Sensor Ozone Data Record
Lucien Froidevaux	3/08	3/13	GOZCARDS: Global OZone Chemistry And Related trace gas Data records for the Stratosphere
Jay Herman	3/08	3/13	Earth Surface and Atmospheric Reflectivity Since 1979 from Multiple Satellites (TOMS, SBUV, SBUV-2, OMI, SeaWiFS, NPP, and NPOESS)
Christina Hsu	4/08	4/13	Consistent Long-term Aerosol Data Records over Land and Ocean from SeaWiFS
Eric Fetzer	5/08	5/13	A Multi-Sensor Water Vapor Climate Data Record Using Cloud Classification
Eric Wood	6/08	6/13	Developing consistent Earth System Data Records for the global terrestrial water cycle

Checklist for Supporting MEaSUREs Datasets and Services

- Preparations for ingesting, archiving and supporting MEaSUREs generated datasets (Highlights)
 - Identify points of contact from data center and MEaSUREs Project data provider (e.g., Principal Investigator or designee to handle data). Firstly, to discuss this checklist.
 - Establish meeting(s) of mutually agreed upon frequency.
 - Develop a mutually agreeable milestone schedule
 - Gather information on the data to be archived
 - Develop and put on-line documentation
 - Develop Working Agreement and possibly an Interface Control Document
 - Determine data transfer protocol
 - Is this a static or dynamic dataset?
 - Identify appropriate metrics that will be collected and reported (e.g., ingest, archive, user access)
 - Transfer data and associated services (if any) to data center
 - Integrate services at data center as necessary
- Preparations for making data/services known and accessible (Highlights)
 - Ensure data and documentation are in place
 - Test documentation and data services for public access
 - Setup and populate web pages
 - Establish 'help desk' points of contact.
 - Publish metadata to appropriate catalogs (Local portal(s), ECHO, GCMD, etc.)
 - Prepare (Coordinate) for public announcement with data provider
 - Identify appropriate user outreach and related conference attendance (paper/poster submission)

Special Situations

- Preliminary discussions with NSIDC on how to make Eric Wood's Hydrologic cycle data, slated for 2 DAACs, seamlessly accessible by hydrologists
- How do we handle early data deliveries?
- Is dataset closed (dataset delivered; no longer being produced) or open (Continuous ingest of data as long as it is being produced)
- What other special situations lurk?

Approach for Preparing MEaSUREs Data Sets for Archive, Distribution, and Stewardship

- Adhere to the goal to provide complete data management services for MEaSUREs datasets, as it does for all resident data holdings, in collaboration with MEaSUREs PIs.
- Maintain continuous communication to collect information on each data set, in order to maximizing familiarity with data and system capabilities
 - Data volume, parameters, services, user community
 - Identify requisite data system capabilities
- Establish implementation schedule
- Ingest data and information into the data system
- Advertise data availability, support user community

Timeline for Integrating MEaSUREs Datasets Into the GES DISC

- 2009 Establish contacts with projects and GES DISC
Start collecting information on project datasets
- 2010 Begin data integration planning and documentation
- 2011 Procure initial hardware for testing interfaces and archive
- 2012 Purchase hardware to host data sets and services
- 2013 Begin to transfer data sets ready for public distribution
- 2014 Finalize all data transfers from projects

MEaSUREs Data System (MDS) Architecture & Functional Diagram

