Data Systems Working Group Report

Paul Wagner

October 2, 2011

Aura Science Meeting - Data System Working Group

Pasadena,CA

Jet Propulsion Laboratory

California Institute of Technology

Copyright 2012 California Institute of Technology. Government sponsorship acknowledged.

Topics

- Instrument Team Ground Data System Reports
  - TES (Doug Shepherd)
  - MLS (Elmain Martinez)
  - OMI (Jacques Claas)
- Aura HDF-EOS Guidelines
- GES DISC Status and Data Preservation (James Johnson)
- ESDIS Report (Rama)
- Archiving and Preservation Overview (Rama and John Moses)
- HIRDLS Experience with Data Archiving (Joanne Loh)

TES

- Prototype developed for joint TES-MLS CO product
  - Incorporated MLS callable forward model (cfm)
  - Naming conventions and joint product metadata developed
- Prototype developed and being exercized for joint TES-OMI Ozone product
  - VLIDORT radiative transfer package
  - Targeted for December 2012 release
- Release 12.4 will read GEOS-5.9 data in netcdf4/hdf5 format
  - Targeted for mid-2013 release
- Release 13 will
  - Create joint TES-MLS CO and TES-OMI products
  - New methanol and formic acid products
  - Update the TES ground data systems to 64 bits for executables and libraries

MLS

- New near-real time version v3.40-nrt-06
  - Improved ozone with higher resolution
  - *New Products: CO, H2O, HNO3, N2O, SO2*
- Standard processing continues with 2 versions
  - V2.24, upgraded to for compatibility with "B"-side scan table
  - V3.33, the newer algorithm, also upgraded for "B"-side compatibility
- V3.41 and v2.31 should replace these before GMAO 5.20 ends
  - Will use GMAO 5.9
  - Incorporates level 1 fixes for THz module declining strength
- Over 10 million scans performed since launch
- V4.00 will be next major release, more than a year away
  - Eliminate problems with V3 products

OMI

- Standard products currently at Data Collection 3
- Some data also available on the OMI very fast delivery (VFD)
  - Less than 15 minute delay
  - Including level 2 volcanic and ozone
- Near-real time also available on LANCE, TEMIS
- A Direct Broadcast Facility is being developed in Alaska Ground Station
- Row anomaly not currently corrected during processing
  - Affected ground pixels identified using flags
  - If requested daily correction parameters will be supplied
  - Changes to affected ground pixels will be noted—the dates processed with outdated flags will be post-processed in level 1, and reprocessed in level 2 and above
- No plans to implement correction parameters

Aura HDF-EOS Guidelines

- No changes since last year
- Draft version with already submitted changes to be made available by summer 2013
- Next version will include sections regarding DOI and archiving metadata

GES DISC Status and Data Preservation

- Receiving new MLS near-real time products
- Working with HIRDLS to preserve data, documentation, and software
- Will store final mission documents in parallel with science data
- Level of support TBD
- Preservation architecture to be built around Fedora Commons open source application
- GES DISC will work with science teams to develop list of items to be preserved

ESDIS report

- Delivers 312Aura products to 8185users at an average daily rate of 215 GB
- Maintains LANCE for near-real time availability of products from MODIS, OMI, AIRS, and MLS
- Reverb replaced WIST January 2012
- Moving to support the metadata format ISO 19115 in ECHO and its Reverb client – seminar November 2012
- Developed Preservation Content Specification

Archiving and Preservation Overview

- Four requirements
  - No loss as bits move across systems or over time
  - Readability over time
  - Long-term understandability
  - Repeatability of previously obtained results
- Motivated by experience with mission data and archives from early missions
- NASA has an Earth Science Data Preservation Content Specification
- Two kinds of information needed for understanding content of data
  - Provenance
  - Content
- Nine categories of content to be preserved
  - Last one is a checklist

HIRDLS Experience with Data Archiving

- Long Term Archiving requirement added in April 2012
- Now must go back and fill holes
  - Some information electronic, some hand-written only
  - Some information simply files stored on individuals' computers
  - Some information is in the form of many, many files stored hierarchically
  - Another example is a document composed of many documents, scanned, and stored as a pdf—handwritten notes in it are not legible
- A major difficuly is deciding what or much is sufficient versus when more becomes excess
- Checklist began as simply a way of tallying which requirements had been met; now seen as necessary to link preserved data name, location, and an "abstract"
- HIRDLS first of Aura teams to go through this; others will learn from its experience