Data Access Services that Make Remote Sensing Data Easier to Use

Christopher Lynnes

Goddard Earth Sciences Data and Information Center
GES DISC began as the Goddard Distributed Active Archive Center (DAAC)

Ingest, process, store and distribute Earth science data (mostly remote sensing)

In the last decade, services have been added

Discovery
Access-related
The Data Usage Cycle

Search → Visualize → Prepare → Acquire → Select → Analyze

Prepare
Preparation Steps

- Subsetting
  - Variable
  - Space
  - Time
- Gridding / (re)projection
- Reformatting to work in the analysis tools
- Quality Filtering

How much of the Preparation process can we build into the Access step?
On-the-Fly Web Services: executed on acquisition
On-the-Fly Web Services

- REST-like: acquire as URLs
  - Limits error return possibilities
  - Requires an HTTP trick (shhh...) for long-running processes

- Accommodates any executable that...
  - Takes one file as input
  - Produces one file as output

- On-the-fly execution means minimal disk buffer requirements
  - No need to stage the whole request for pickup
On the Fly Subsetting

Subsetted Data Download Instructions:

- **Multiple** file download:
  - To use `wget`, you will need version 1.11.1 or later. To find the version, run `wget --version`.
  - If necessary, obtain the latest version of `wget`.
  - Download the list of URLs
  - Run: `wget --content-disposition -i wget_khElxSN`

- **Single** file download: click on the link for each file to be downloaded:
  - MERRA100.prod.assim.inst3_3d_asm_Cp.19790131.SUB.nc
  - MERRA100.prod.assim.inst3_3d_asm_Cp.19790130.SUB.nc
  - MERRA100.prod.assim.inst3_3d_asm_Cp.19790129.SUB.nc
  - MERRA100.prod.assim.inst3_3d_asm_Cp.19790128.SUB.nc
  - MERRA100.prod.assim.inst3_3d_asm_Cp.19790127.SUB.nc

*Modern Era Retrospective-Analysis for Research and Applications*
On-the-fly Conversion to netCDF (network Common Data Form)

-most Earth Observing System datasets are in Hierarchical Data Format (HDF)

-but, many visualization tools understand netCDF “better”

TRMM Monthly Rainfall rate for Oct 2011 in Panoply
http://www.giss.nasa.gov/tools/panoply/
Data Quality Screening Service

- Level 2 Satellite data often comes with quality control flags.
- Until now, each user typically had to write his/her own software to filter bad quality data—or ignore them.

![Image of data quality screening with maps and color scales for kg/m² and quality flags for Hurricane Ike, 9/10/2008.]
The Data Quality Screening Service for AIRS Level 2 swath data

Original data array

Mask based on user criteria
Quality flag < 2

Good quality data pixels retained

Total column precipitable water

Output file has the same format and structure as the input file, with fill values replacing the low-quality data
OPeNDAP*:
a protocol standard for remote access

*Open-source Project for a Network Data Access Protocol
OPeNDAP: Subsetting and more

- Subsetting
  - individual variables
  - slices of variables

- Reformatting: download as...
  - ASCII
  - netCDF
Varieties of OPeNDAP

- Hyrax
  - High performance
  - Reformat to netCDF
- GrADS Data Server
  - Multiple input formats
  - Server-side processing
- THREDDS Data Server
  - Aggregation
  - Web Coverage Service, netCDF Subsetter
- Others: ERDDAP, PyDAP, Dapper...
Giovanni: online analysis and visualization
Giovanni

- Analysis and visualization server
- Workflow paradigm
- Steps for:
  - Fetching
  - Subsetting
  - Quality filtering
  - Regridding
  - Averaging
  - Visualization
- Output can be downloaded
Example: Carbon Monoxide from 2010 Russian wildfires
The Data Usage Cycle Refactored

Find → Select (Giovanni) → Visualize → Acquire & Prepare

Prepare:
- OPeNDAP
- Subsetting
- Quality Filter
- Reformat
Frontier: Seamless interaction of steps

Select

Find

Acquire & Prepare

Analyze
Seamless Search and Analysis

Start: 2010-10-31 00:00:00
End: 2010-10-31 23:59:59
Area: -115 to -25
Area: -22.5
Measurement: Soil Moisture (SMAP)
Filter Quality?: X
Fetch