Advancing the Power & Utility of Server-Side Aggregation

ESIP Summer-2016 OPeNDAP Workshop
Wednesday, July 20th, 2016, 13:00-17:00
(excerpted from an earlier presentation)

Dave Fulker & James Gallagher, President & Vice President of OPeNDAP, Inc.
subcontractor to Raytheon for NASA/ESDIS

Workshop supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C
Data systems often contain files or images (i.e., granules) that may be accessed only independently, even when kept in collections of highly similar entities.

Such granularity typically reflects how data are collected, unrelated or contrary to data utility.

Panel members are experts on needs-driven aggregations of related granules.

EOSDIS recently invested in enhancements to OPeNDAP’s aggregation features...
OPeNDAP Concepts
from Distributed Ocean Data System (DODS) circa 1994

- URL ≈ dataset*
- URL with constraint ≈ subset
- Retrieve
- Retrieval protocol built in to multiple libraries
- flexible data typing
- many, diverse clients
- dataset descriptions (metadata)
- dataset content (typed/structured)
- arrays (~coverages)
- tables (~features)

*dataset ≈ granule or virtual aggregation
motivation to enhance Multi-Granule Aggregation

Many servers allow DAP providers to form virtual aggregations of similar granules (files).

But until now, users generally could not choose

- Granules to be aggregated
- Forms of aggregation

Furthermore, array- & table-style subsetting could not be mixed (with or without aggregation)
recent OPeNDAP work on Multi-Granule Aggregation

Outcomes

- Acquire data from 1,000s of files with one request
  
  *N.B. Necessitates use of HTTP POST (to avoid huge URLs)*

- Two forms of aggregation response
  
  - Zipped netCDF files
  
  - Concatenated tables (CSV)

  *N.B. Arrays may be aggregated as concatenated tables!*
This panel session and some of the work to be discussed were supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C