Progress toward Web-Service APIs for Pre-Retrieval Data Operations

Reflecting the Status of ODSIP
An EarthCube Building Blocks Project
by Dave Fulkerson (OpenDAP, Inc) for the EarthCube Tech-Hands Meeting
Wednesday, 8 April, 2015
Advancing "data as a service" to embrace a richer set of server-side operations (beyond filtering/subsetting)

*ODSIP is led by:
OpenDAP, an open-source-focused nonprofit (opendap.org & GitHub)
+ U-NC, U-HI, U-RI & Unidata
UNDERLYING CONCEPTS
FROM DODS (NOW OPeNDAP) CIRCA 1994

URL ≈ dataset | URL w/ constraint ≈ subset

Retrieve dataset descriptions (metadata)
dataset content (typed/structured)

Retrieval mechanisms are built into libraries
(in multiple computer languages)

many, diverse clients
flexible data typing
arrays (~coverages) tables (~features)
Domain name often is an organization’s web server. Servers often have hierarchical collections. Each URL references a distinct DAP “dataset.” Suffixes specify return types. Depending on the suffix, DAP returns (textual) metadata or (binary) content, with options for human- and machine-readable forms (XML, NetCDF4...). Suffix “dmr” yields metadata only.

http://laboratory.edu/device/experiment/granule.dmr

Under DAP4 (as with DAP2)

URL ≈ Dataset (Granule)
Dataset identifier as above, except return-type is NetCDF4 (= HDF)

DAP4 "constraint expressions" yield sub-arrays & other proper subsets

DAP4 "function expressions" enable server-specific extensions

Constraints specify subsets by array indices, variable names, and (in some cases) content. Function expressions allow new services (statistics, etc.) to be offered as part of data retrieval.

The expression &dap4.func=... enables DAP extensions, such as server functions
Well-Established Server-Side Ops: Subset Selection from Arrays or Tables —

- Select variables by name
- Select rows of a table (records of a sequence) by value constraints
- Select subarrays by limiting the ranges of (named) indices
- But — for example, array subsets generally cannot be selected using value constraints
DATA-ACQUISITION NEEDS
motivating ODSIP Web-services extensions —

Much richer subsetting
- Non-rectangular structures
  - Triangular/polygonal meshes
- Value constraints on arrays
  - What shape result?
- Other constraints requiring complex calculations...

Additional operations*
(pre-retrieval!)
- Statistical summarization, binning...
- Remapping, regridding...
- Feature extractions
- Etc.

*Note: As a rule, such ops must decrease data-transfer volumes
In Brief, the ODSIP Project Intends

- Develop a service-invocation protocol that
  - Extends the well-used OPeNDAP protocol
  - Supports rich pre-retrieval processing
- Prototype its use in 3 demanding geo contexts
  - Climate-model downscaling for native-Hawaiian use
  - Storm surge prediction for coastal NC
  - SST front analysis/synthesis from satellite imagery
- Engage the EarthCube community...
Advances Available Now

- Tabular Output—as CSV, e.g.—of Selected (n-Dim) Array Elements
  - Starting to bridge feature-coverage gap...
- Requester-specified aggregations of (similar) datasets (granules)
  - Long lists (via Post) of URLs + DAP ops
  - Results packaged via zip or aggregated table
- Sub-Mesh Selection for Non-Rectangular Grids (UGRID)
- Encodings (next slide)
**ENCODINGS**  
*OPENDAP as MIDDLEWARE*

Data **ingest** via encoding-specific adapters

A growing set of open-source modules map source formats and databases -> (Coverage-like) DAP data model

**Response** encodings (suffix-spec'd)

- netCDF-2\4 (=> HDF5), XML (=> HTTP), DAP-2\4
- Less well known: W10n (JSON), WMS, WCS (beta)

Objects via libraries: Java (script), Python, C++, Fortran...