



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Intro & Recent Advances

Remote Data Access via OpenDAP Web Services

ESDIS Presentation & Demonstration for the
CEOS Working Group on Information Systems & Services (WGISS)

May, 2015

Dave Fulker, President of OpenDAP,
subcontractor to Raytheon for NASA/ESDIS

Presentation supported by NASA/GSFC under
Raytheon Co. contract number NNG10HP02C

Part I





Introduction to OpenDAP* Web Services

*OpenDAP is an organization and an acronym:

“Open-source Project for a Network Data Access Protocol”

OpenDAP Concepts

originally from Distributed Ocean Data System (DODS) circa 1994

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

*dataset \approx granule

URL \approx Granule*

per OpenDap's Data Access protocol (DAP)

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

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 suffix, DAP returns metadata or content, with options for human- or machine-readable forms (XML, NetCDF4...). Suffix "dmr" \rightarrow metadata only.

*dataset \approx granule

OpenDAP

Datatype Philosophy

- ★ Internal data model has few data types
 - ★ For simplicity...
- ★ Types are domain-neutral but flexible
 - ★ Structures & attributes → rich syntax & semantics
- ★ These types support many domain-specific needs
 - ★ A recent crawl* (23,000 domains in .gov, .edu, .org) found >1400 collections with DAP servers

*Performed by the National Snow & Ice Data Center
in an NSF/EarthCube project

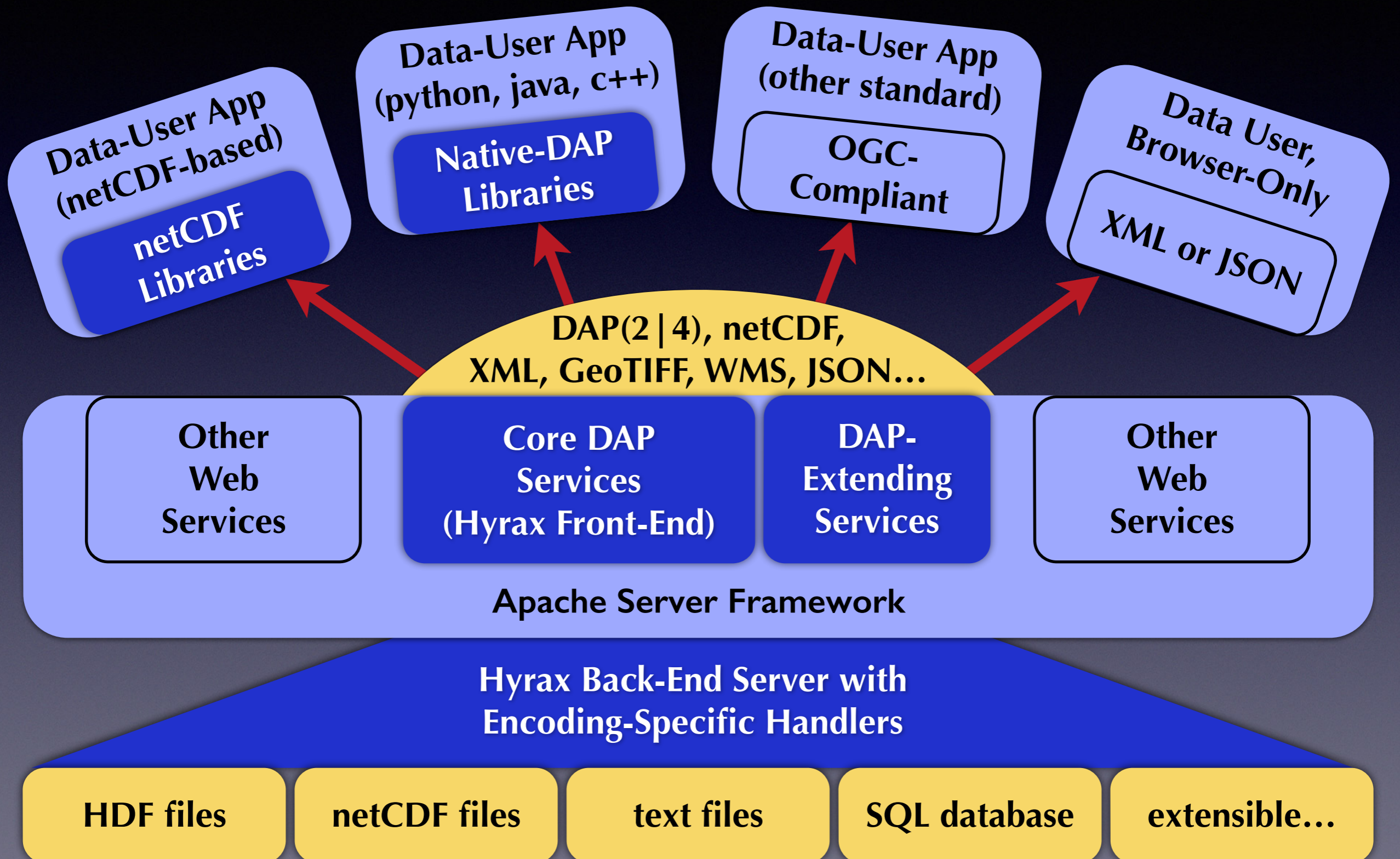
OpenDAP services

Function as Middleware

- ★ Data ingest via encoding-specific adapters
 - ★ Handlers for a growing set of source-data types
- ★ Multiple response encodings
 - ★ Native DAP—useful in Python, Java, C++ ,Fortran...
 - ★ netCDF (also GeoTIFF where possible)
 - ★ XML (\Rightarrow HTTP via style sheets)
 - ★ Recently added: WMS, W10n (JSON), WCS (beta)

Architectural Overview of Hyrax

a widely-used DAP server



URL + Query → Subset
& (future) results from other server functions

http://.../granule.nc4?dap4.ce=constraints&dap4.func=functions

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

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

DAP4 "function expressions"
enable extensions

Constraints specify subsets by variable names, by array indices & (for tables) by content. Likely extensions include statistics, UGRID subsetting, feature extraction...

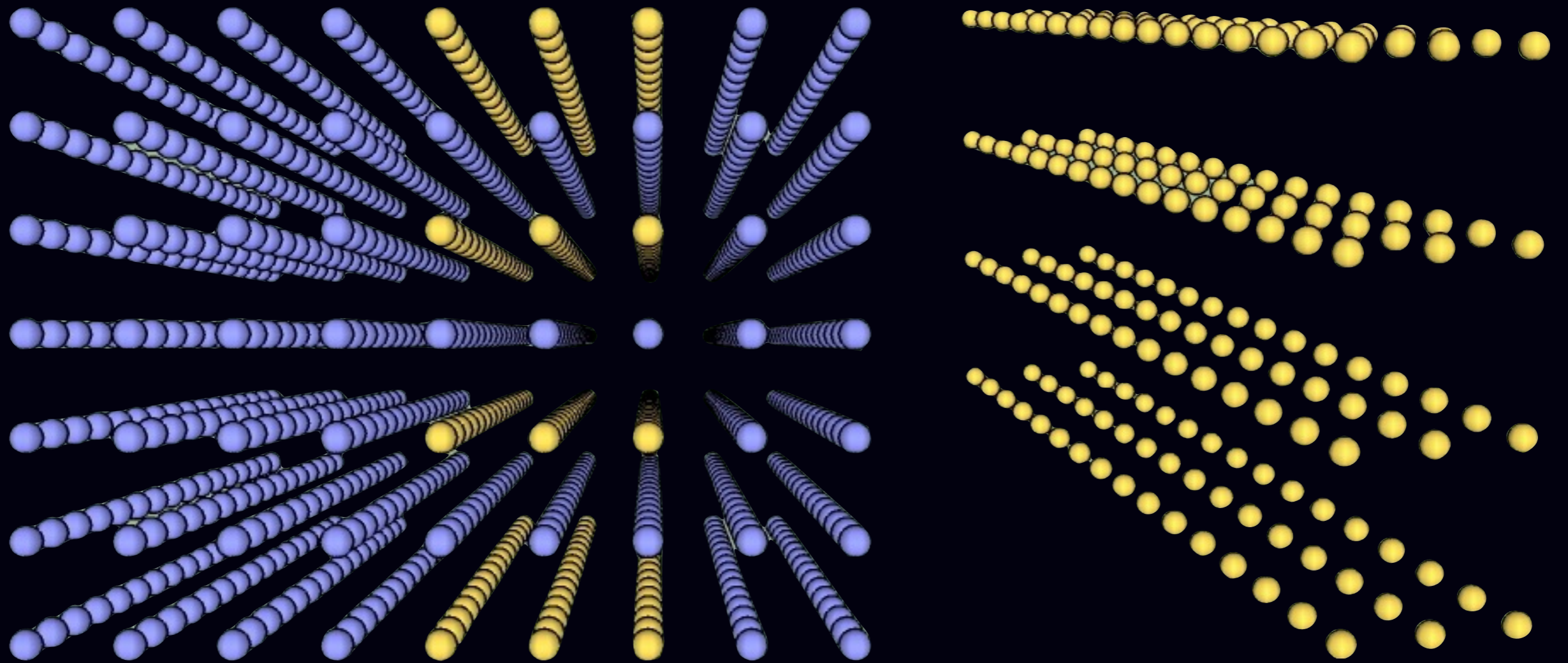
The query form **&dap4.func=...** enables
DAP extensions ⇒ new server functions

DAP-based Subset Selection (from arrays | tables)

- ★ Select variables by name
 - ★ For tabular data, this means selecting columns
- ★ Select rows of a table via column-specific value constraints
 - ★ Allows both domain-based & range-based subsetting
- ★ Select sub-arrays by constraining their indices

(array-style)

Index-Constrained Subsetting



Source Array



Sub-Array (response)

caveat —

Index-Based Subsetting

- ★ Excellent if desired subset is a bounding box parallel to source array (indices \leftrightarrow coordinates)
- ★ Less useful when
 - ★ Subset selection not based on domain coordinates
 - ★ Source is not organized as coordinate-mapped arrays
 - ★ Desired subset is polygonal or is skewed (relative to source-array orientation)



Part II

Recent Enhancements of OpenDAP Web Services

With Demonstrations



EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

*This part of the presentation is drawn
primarily from a project report on:*

NASA Data Interoperability

An EOSDIS Presentation & Demo

Originally given March 27, 2015


Original Presenters: James Gallagher & Nathan Potter (OpenDAP)

*This work was supported by NASA/GSFC under
Raytheon Co. contract number NNG10HP02C*



main NASA motivations for OpenDAP Enhancements

- ★ Easier software builds & better documentation
- ★ Authentication of data users
- ★ More response encodings
 - ★ Open Geospatial Consortium (OGC) Web Services (WMS, WCS...)
 - ★ JavaScript Object Notation (JSON) for Webification (w10n)
- ★ Requesting DAP ops on many granules at once
 - ★ Response = concatenated CSV (arrays → tables) or
 - ★ Response = zipped files



progress on simplifying

OpenDAP Server Installation






Context

-  Hyrax-install complexity was once a barrier to use



Key Accomplishments

-  Adding modules does not increase the package count
-  Source build: now just 3 distinct packages
 - Previously 18 packages
-  Binary install: now just 2 RPMs + 1 WAR
 - Previously 15 RPMs + 1 WAR



progress enhancing OpenDAP's
Website & Documentation



Key Accomplishments



Various Website repairs



760 fixed links (from automated before/after crawls)



Five documents added

- Client configuration for authorization
- Server configuration for authorization
- Source-code build how-to
- Summary of Winter-2015 ESIP-panel on Web-services performance
- Server configuration for WMS provision

progress on

User Authentication

(via EarthData login at NASA EOSDIS)

★ Context/Things to Notice

★ Fine-grained access control for individual directories

★ Demo is Web-only, but cURL tests work as well

- cURL—like most client applications—is built around libcurl, thus serving as a lowest common denominator
- EarthData credentials are simply stored in a user's .netrc file

★ Live Demo...



prior context for enhancing
Multi-Granule Aggregation

- ★ Many servers have allowed DAP *providers* to form (virtual) aggregations of (similar) granules
- ★ But until now, end users could not choose
 - ★ Granules to be aggregated
 - ★ Forms of aggregation
- ★ Furthermore, array- & table-style subsetting could not be mixed (with or without aggregation)



progress on requester-specified
Multi-Granule Aggregation



Context/Things to Notice

- 
- Request data from 1,000s of files with one operation**

N.B. Necessitates use of HTTP POST (to avoid huge URLs)

- 
- Two forms of aggregation response**

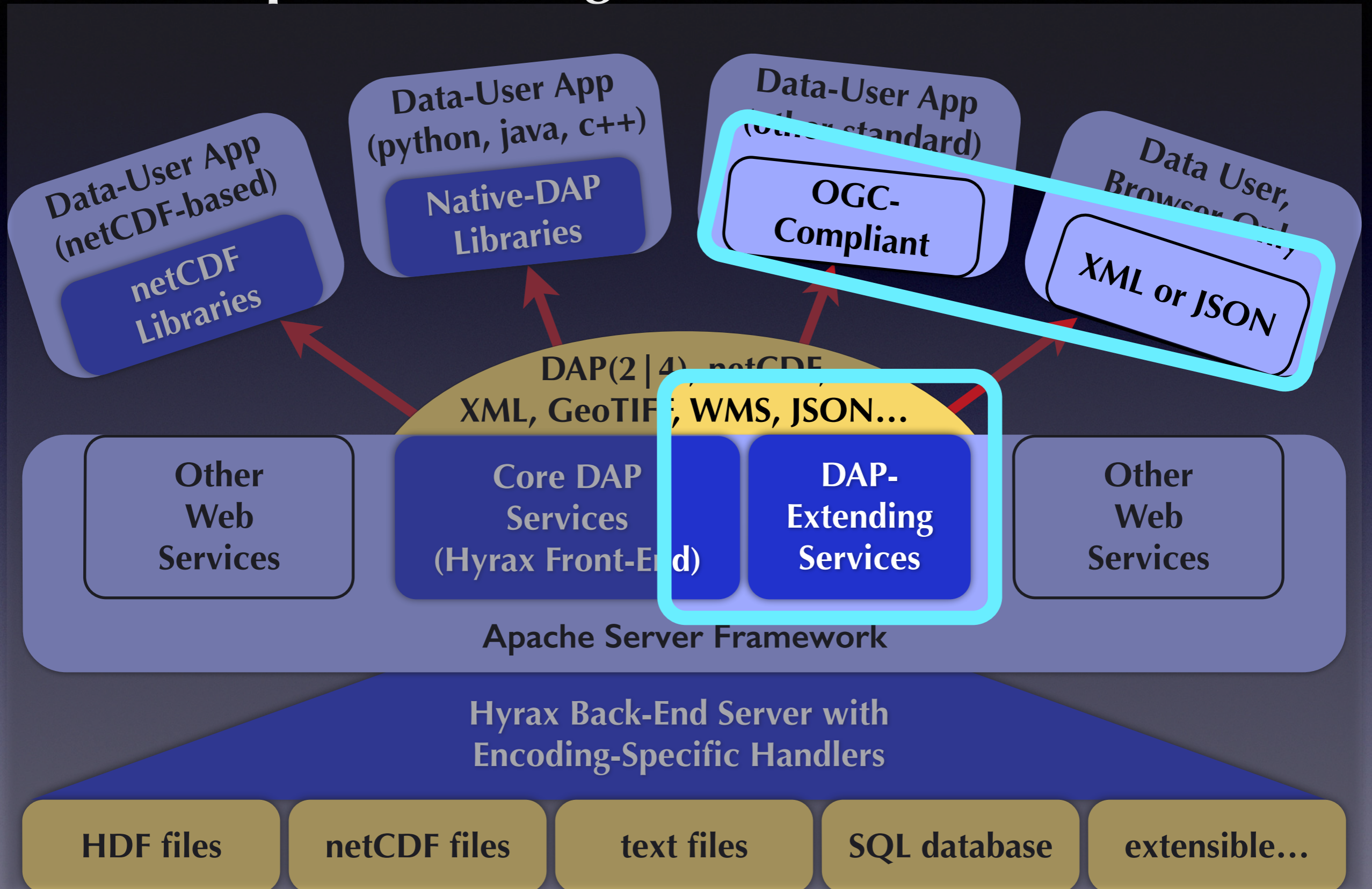
- Zipped netCDF files
- Concatenated tables (CSV)

N.B. Arrays may be aggregated as concatenated tables!



Live Demo...

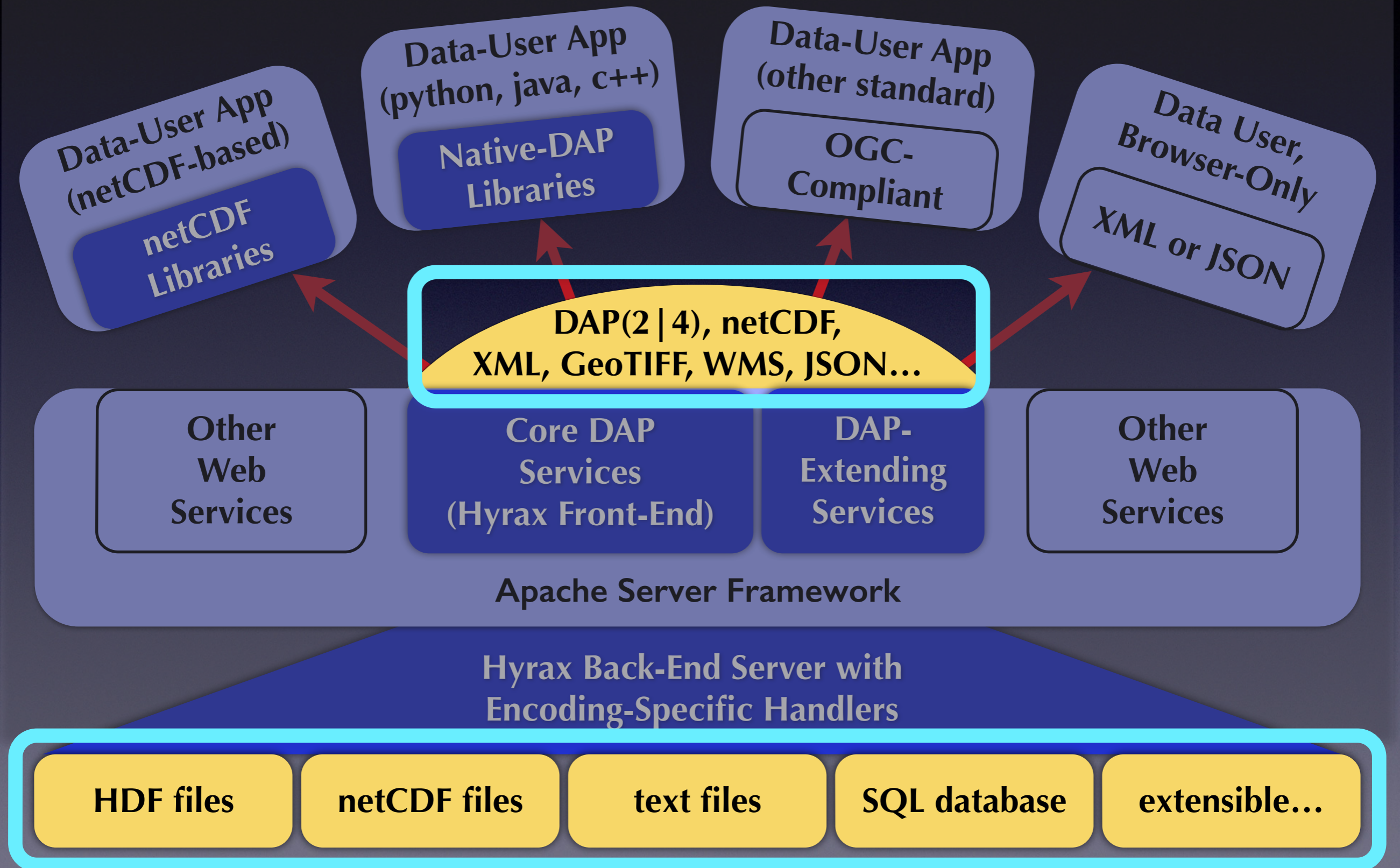
DAP Output-Encoding Extensions



OGC Protocol: WMS Web Mapping Service

- ★ **WMS (Web Mapping Service)**
 - ★ Great for 2-dim geospatial data on 'maps' (but not for higher-dimensional data types)
 - ★ A bridge to display tools, notably, Google Earth
- ★ **Live Demo...**

DAP Interoperability Leverage



relevance:

OpenDAP & Interoperability

★ We demonstrated

★ NASA (HDF5) files → OpenDAP → WMS → Google Earth

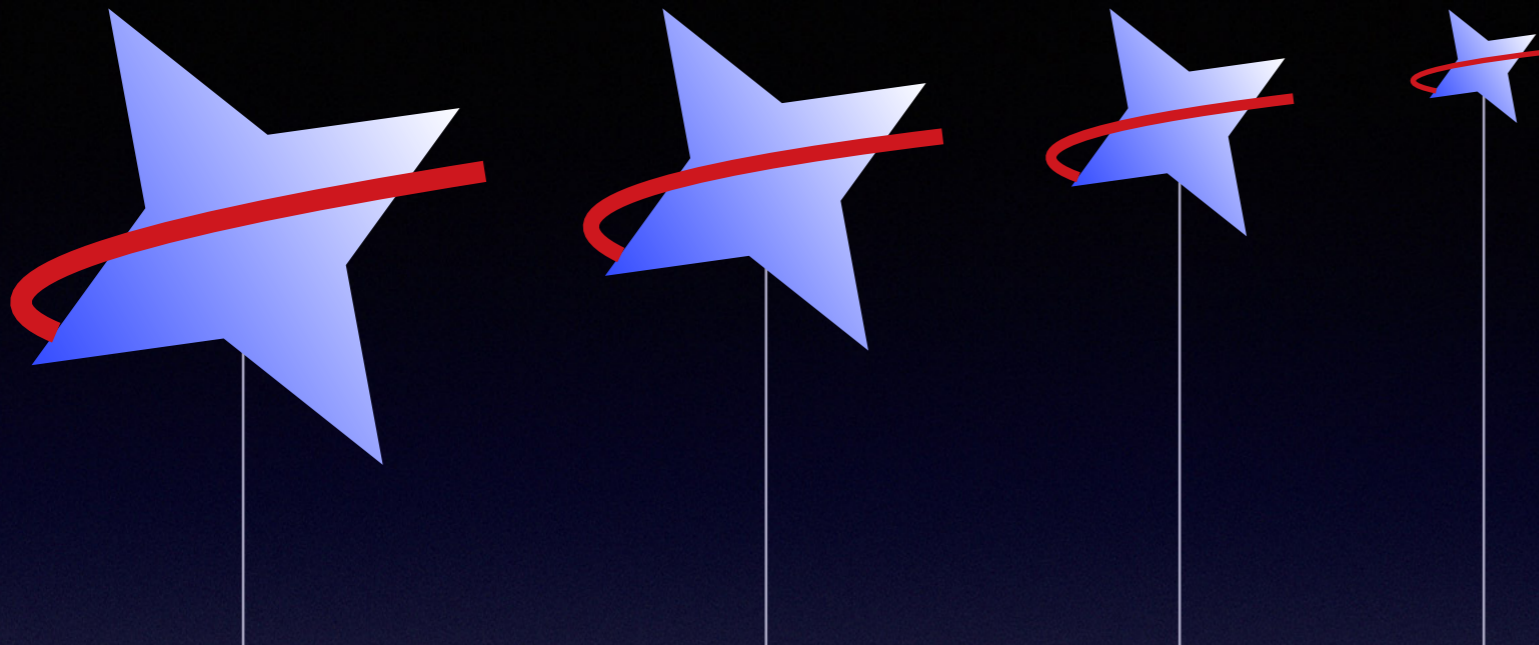
★ Notably, it seems unlikely that either

★ Google Earth engineers anticipated reading HDF5 or

★ NASA engineers planned to display data on Google Earth!

★ This suggests* a definition for interoperability:
“supporting unanticipated uses”

**Paraphrasing John Orcutt*



This presentation, and
the recent work described,
were supported by NASA/GSFC under
Raytheon Co. contract number NNG10HP02C

Raytheon