EOSDIS

NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Share Data with OPeNDAP Hyrax: New Features and Improvements

July 2016

James Gallagher (OPeNDAP)

This work is supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C and a Grant from the Australian National University.

SESIP-0716-JG

EOSDIS

Overview

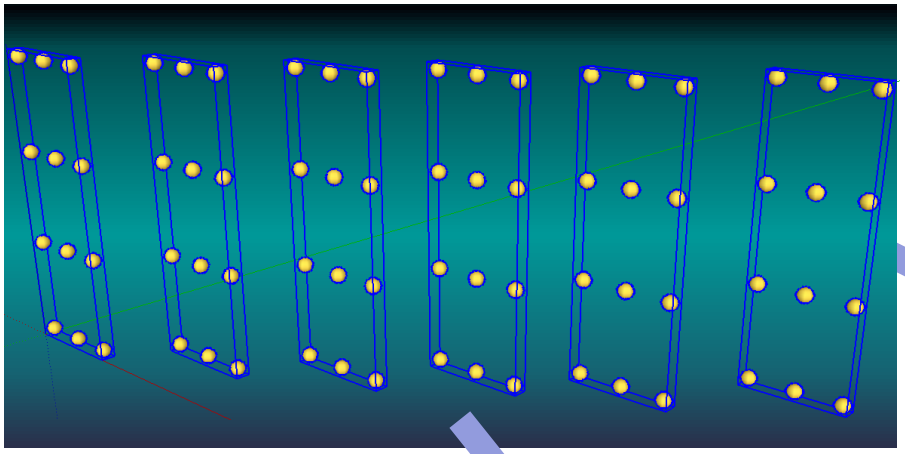
- Aggregation
  - Performance improvements
  - User-invoked aggregation
- Authentication
  - End-to-end (web & programmatic clients)
- Web-Service Protocols
  - W10n, WMS

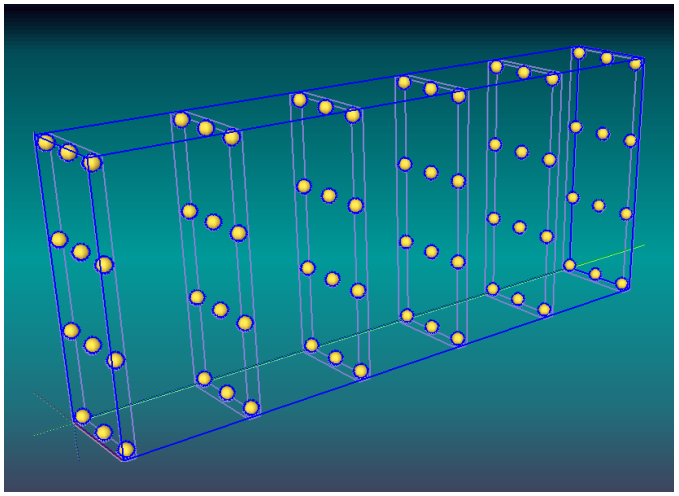
Brief review of Aggregation

- Combine discrete granules (e.g. files)
- They are discrete usually because of the mechanics of collection or processing
- Server aggregation frees users from having to understand the archive's exact structure
- Data type determines which aggregation techniques are appropriate
  - Regular data – e.g. level 3
  - Swath data – e.g. level 2

DAP Servers offer Virtual Aggregations

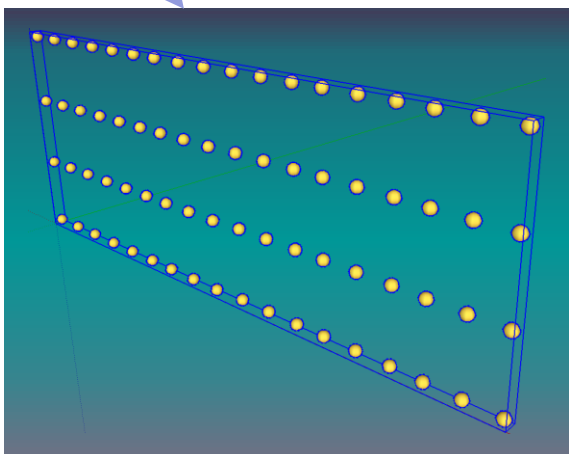
of Identically Shaped Variables in different granules





Or*

Lengthening
An existing
dimension



Adding Some
New Dimension

*as specified via NCML

EOSDIS

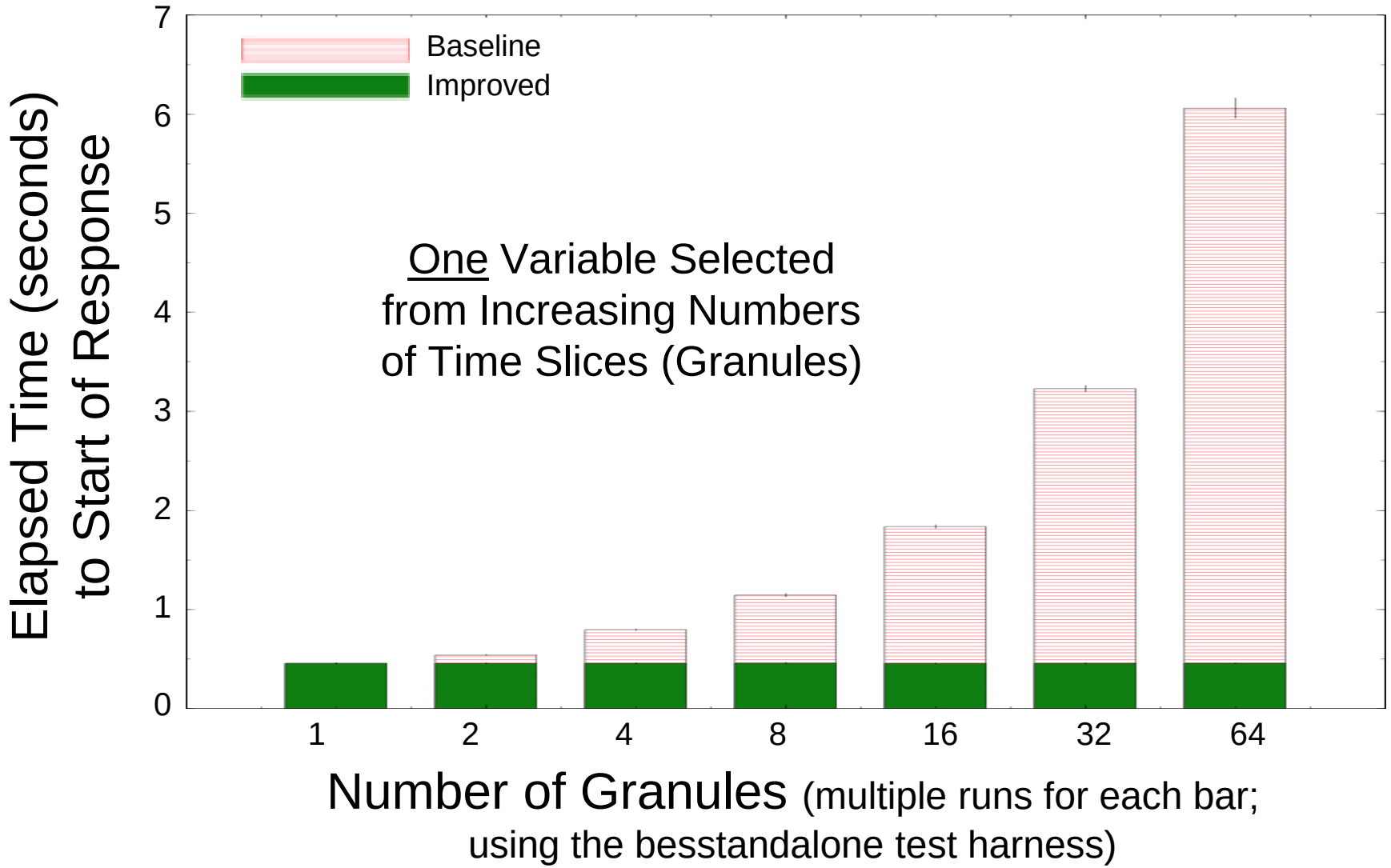
Aggregation: performance

- Test case*: Suppose a user were to ask for all of the data from a small aggregation – 120+ files of AIRX data:
  - The original server would leave the client hanging for ~26 minutes and then drop the connection… (because the network connection times out)
  - The new server will return all of the data (166GB) and begin to stream the results within fractions of seconds!
- What if the hardware were made large enough so that the original software could complete?
  - The original server has an effective transfer rate of 5MB per second…
  - The new server has an effective transfer rate of 33MB per second*, a factor of 6.8 increase!
- Server memory requirements were also reduced
  - The original server used memory (RAM) roughly equal to the response size…
  - The new server has a flat memory consumption that it equal to the smallest atomic unit accessible from the data store (e.g., array slab). For the test case above, the reduction is 4 orders of magnitude!
- Key Points:
  - The server can now handle requests it simply could not before
  - This scales to very large requests – because the implementation uses pipelining to minimize data held in memory at any given time

EOSDIS

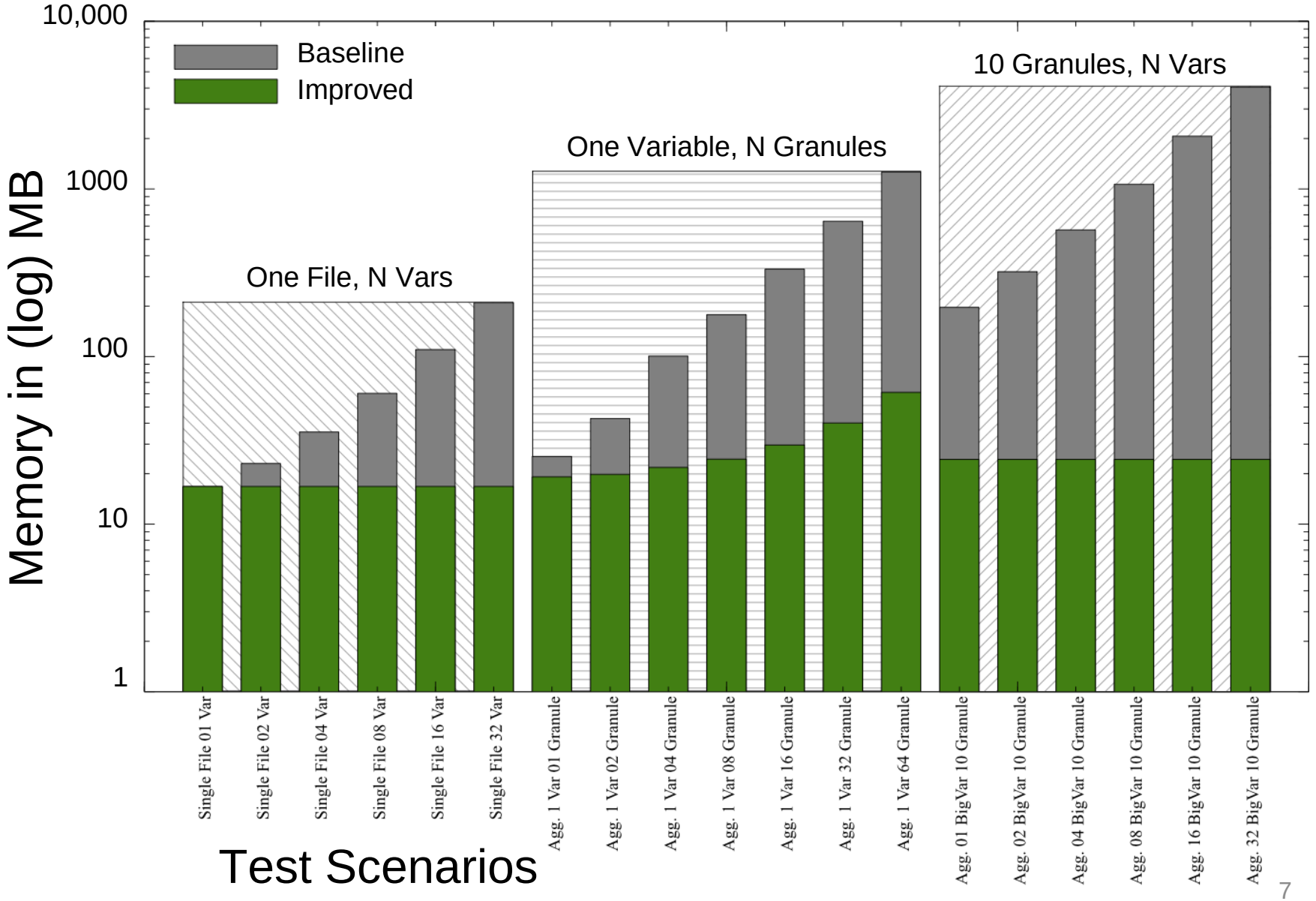
* Tests run between machines in the Amazon cloud, see notes

Quicker Initial Response



EOSDIS

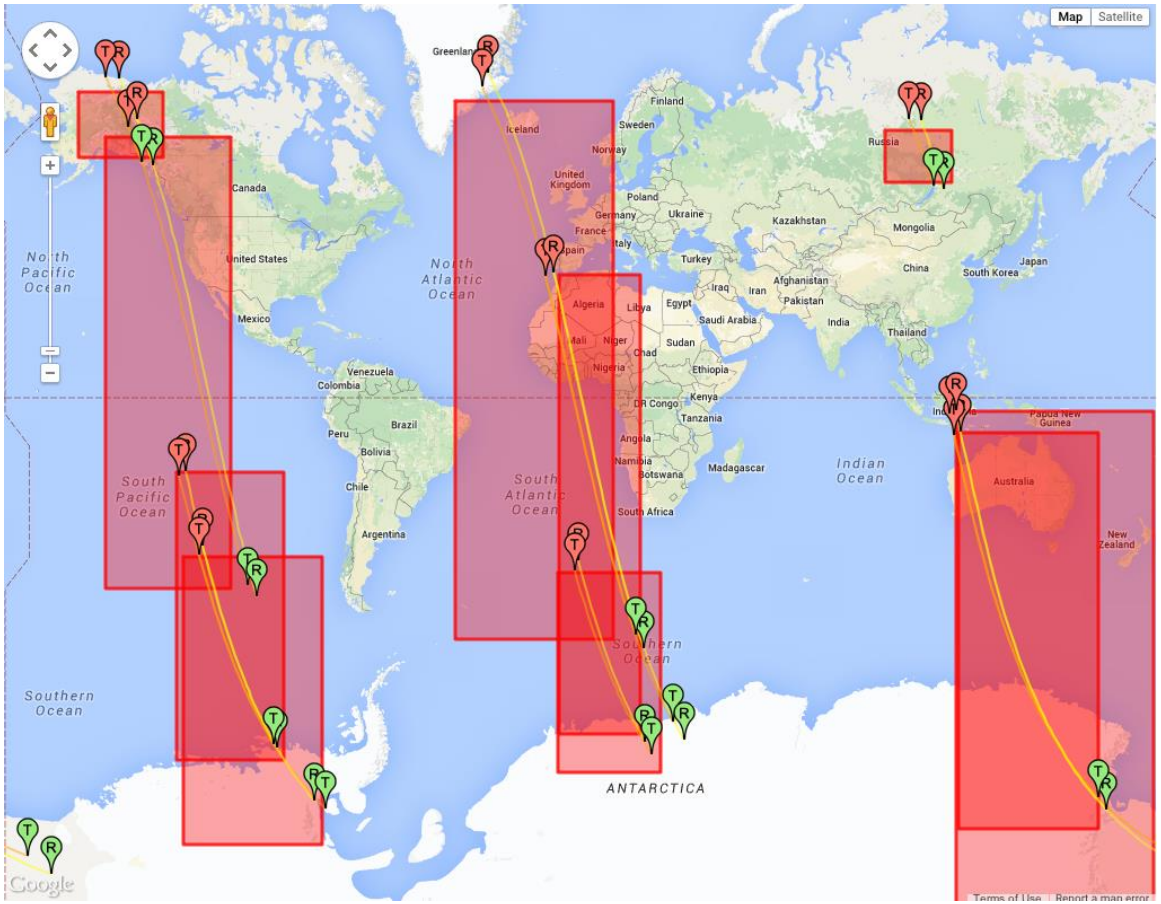
Less Server-Memory Use

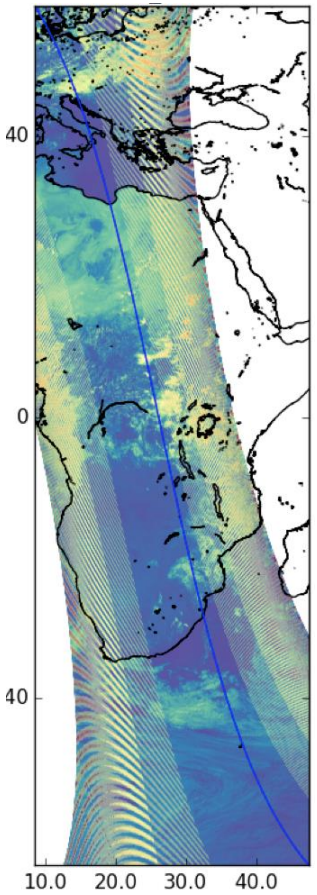


Aggregation of swath data

…is also supported, differently

Simple aggregation techniques do not apply to swath data





EOSDIS

Aggregation: user-driven

- User-driven: The client provides a list of datasets to aggregate
- Web service interface is easy to customize for differing clients
- While intended for ‘Swath data,’ this can be used with any collection of datasets
- Users may get data from *thousands* of files in a single operation (e.g., clicking a single link)
- Con: Clients must know the granules they want

User-driven Aggregation Mechanisms

- Two different response forms → different aggregation algorithms
  - Iterate over a set of granules applying a constraint to each, collecting discrete results in a single archive file
  - Iterate over a set of granules, transforming them to a table* and selecting a subset of those rows by value

*Actually DAP Sequence objects

User-driven Aggregation Interface Description

- Implemented using a new web service that relies on POST so it can accept larger inputs than HTTP GET will allow.
- It accepts a series of commands that describe the
  - Response format (e.g. files in a zip archive, CSV-encoded table)
  - Datasets, modified by their constraints, to aggregate

User-driven Aggregation Example

- Users of NASA's Earthdata Search Client* can receive single data responses from queries that involve many datasets.
- Example operations
  - Get the server's version
  - Get a simple aggregation, returned in an archive, values encoded in netCDF3 (other formats are supported).
  - Get an aggregation with array data transformed to a table, values encoded using CSV

EOSDIS

*This feature was initially developed for EDSC

…Server Version

- We can supply arguments using GET
- The simplest ‘operation’ is ‘version’

Request:
http://test.opendap.org/dap/aggregation/?&operation=version

Returns:

```
Aggregation Interface Version: 1.1
<?xml version="1.0" encoding="UTF-8"?>
<response xmlns="http://xml.opendap.org/ns/bes/1.0#" reqID="[ajp-bio-…">
  <showVersion>
    <Administrator>support@opendap.org</Administrator>
    <library name="bes">3.16.0</library>
    <module name="dap-server/ascii">4.1.5</module>
    <module name="csv_handler">1.1.2</module>
    <library name="libdap">3.16.0</library>
…
</response>
```

… Returning an Archive

- Use POST; multi-line input

Request:

```
&operation=netcdf3
&var=Latitude,Longitude,Optical_Depth_Land_And_Ocean
&file=/data/modis/MOD04_L2.A2015021.0020.051.NRT.hdf
&file=/data/modis/MOD04_L2.A2015021.0025.051.NRT.hdf
&file=/data/modis/MOD04_L2.A2015021.0030.051.NRT.hdf
```

Returns:

```
Archive:  d1.zip
    testing: MOD04_L2.A2015021.0020.051.NRT.hdf.nc   OK
    testing: MOD04_L2.A2015021.0025.051.NRT.hdf.nc   OK
    testing: MOD04_L2.A2015021.0030.051.NRT.hdf.nc   OK
No errors detected in compressed data of d1.zip.
```

…Returning a Table

- The request looks similar; the return type – CSV – requires different formatting

Request:

```
&operation=csv
&var=Latitude,Longitude,Image_Optical_Depth_Land_And_Ocean
&bbox="[49,Latitude,50][167,Longitude,170]"
&file=/data/modis/MOD04_L2.A2015021.0020.051.NRT.hdf
&file=/data/modis/MOD04_L2.A2015021.0025.051.NRT.hdf
&file=/data/modis/MOD04_L2.A2015021.0030.051.NRT.hdf
```

Returns:

```
Dataset: function_result_MOD04_L2.A2015021.0020.051.NRT.hdf
table.Latitude, table.Longitude, table.Image_Optical_Depth_Land_And_Ocean
49.98, 169.598, -9999
49.9312, 169.82, -9999
49.9878, 169.119, -9999
49.9423, 169.331, -9999
49.8952, 169.548, -9999
49.8464, 169.77, -9999 ...
```

Aggregation – Summary

- ‘Regular’ aggregation performance improved by orders of magnitude
- User-driven aggregation is a new feature
- Regular aggregation – defined by data provider
- User-driven aggregation – defined by user
- User-driven aggregations work on a wide variety of data types

Overview

- Aggregation
  - Performance improvements
  - User-invoked aggregation
- Authentication
  - End-to-end (web & programmatic clients)
- Web-Service Protocols
  - W10n, WMS

Authentication

- Hyrax works with NASA's Earthdata Login (OAuth2), LDAP and Shibboleth to provide user authentication
- Hyrax + Earthdata Login supports both web and programmatic access
  - EarthData Login access will work for data analysis tools too.
  - LDAP does as well!
  - Shibboleth does not (easily)
- Each of the three are 'Single Sign On' services:
  - One database of credentials → many data servers

Authentication - Configuration

- Apache modules provide the actual authentication – Advantage: robust code used by many sites
- Hyrax + Authentication software stack:
  - Apache → Tomcat → Hyrax
  - Configure Apache httpd and Tomcat to work together
  - Configure Apache httpd to authenticate
- Configuration information on the web
  - docs.opendap.org*

Authentication – Programmatic access

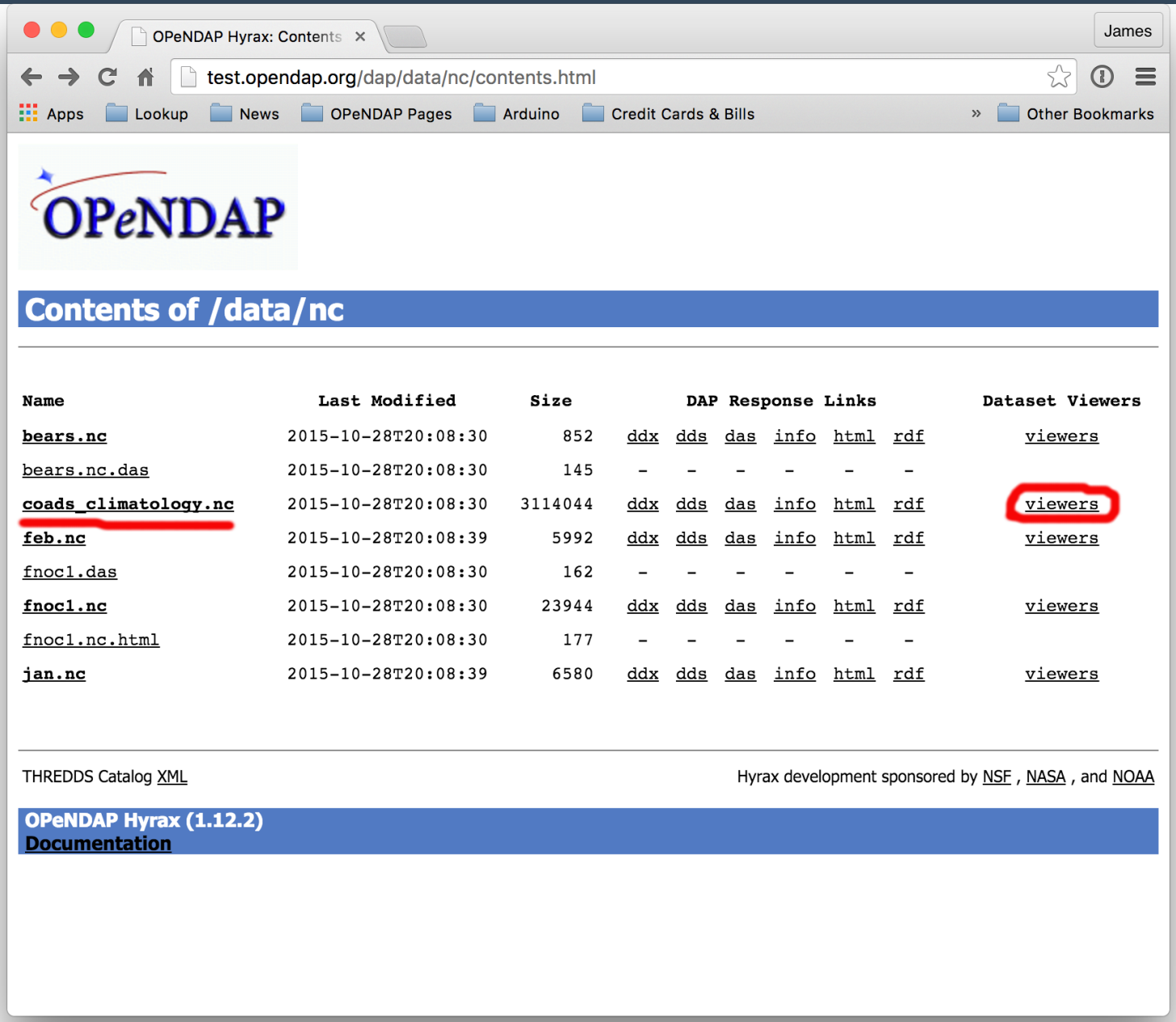
- Programmatic clients: put username and password info in a `.netrc` file
- This is better than asking users to install short-lived certificates (a process that they'd have to repeat often)
- This enables automatic access to data
  - Processing done periodically
  - Batch jobs with many accesses

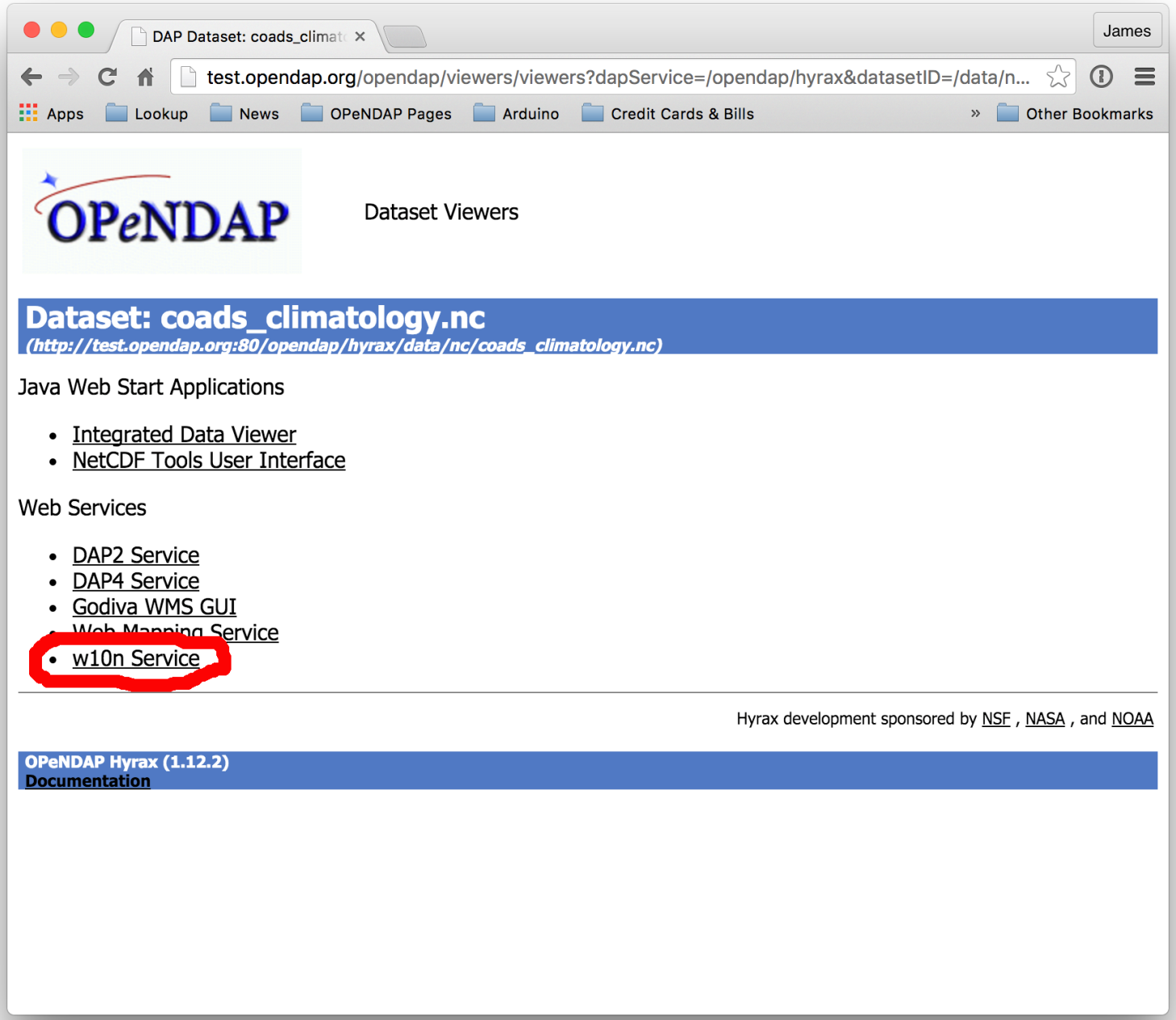
Overview

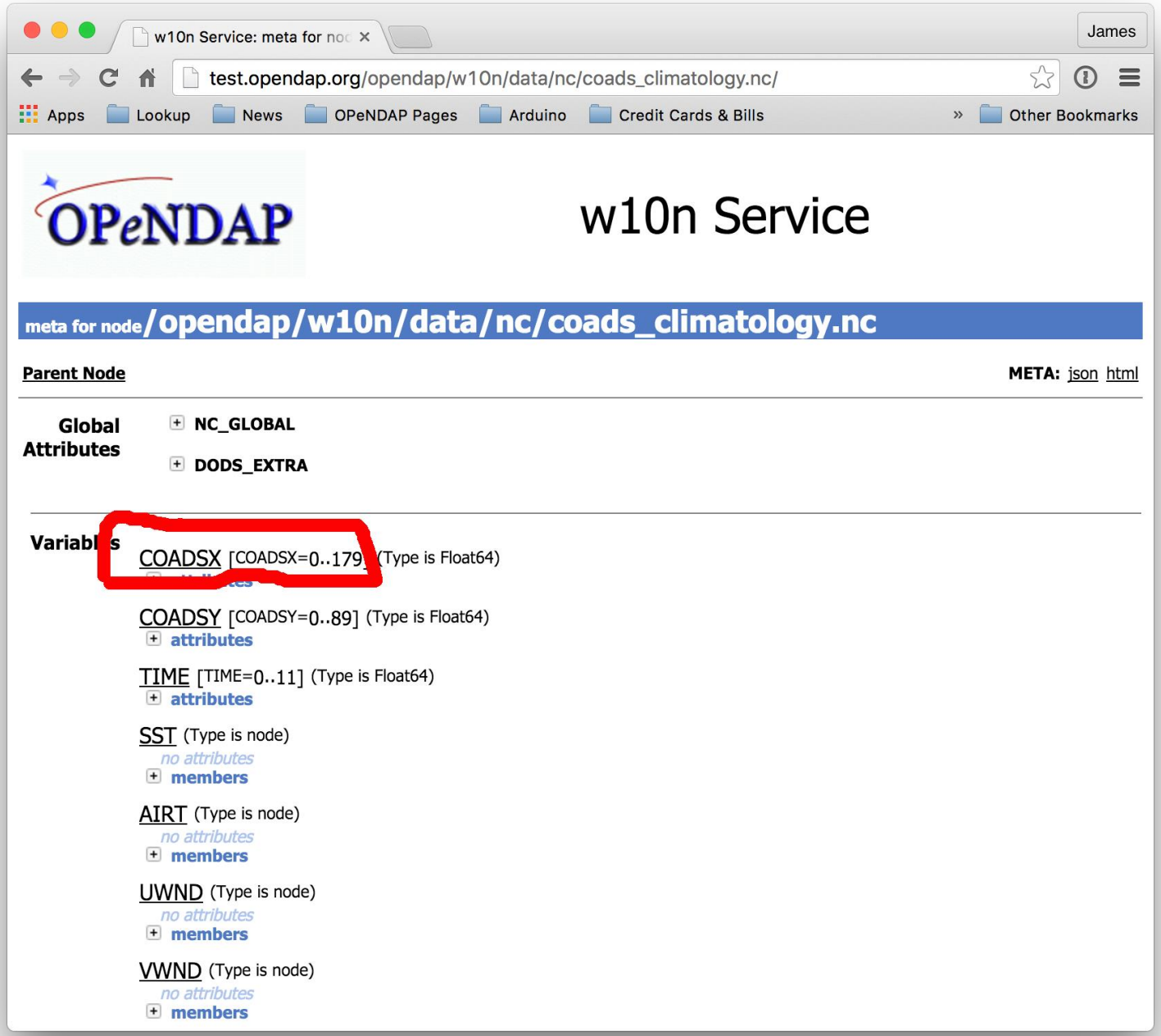
- Aggregation
  - Performance improvements
  - User-invoked aggregation
- Authentication
  - End-to-end (web & programmatic clients)
- Web-Service Protocols
  - w10n, WMS

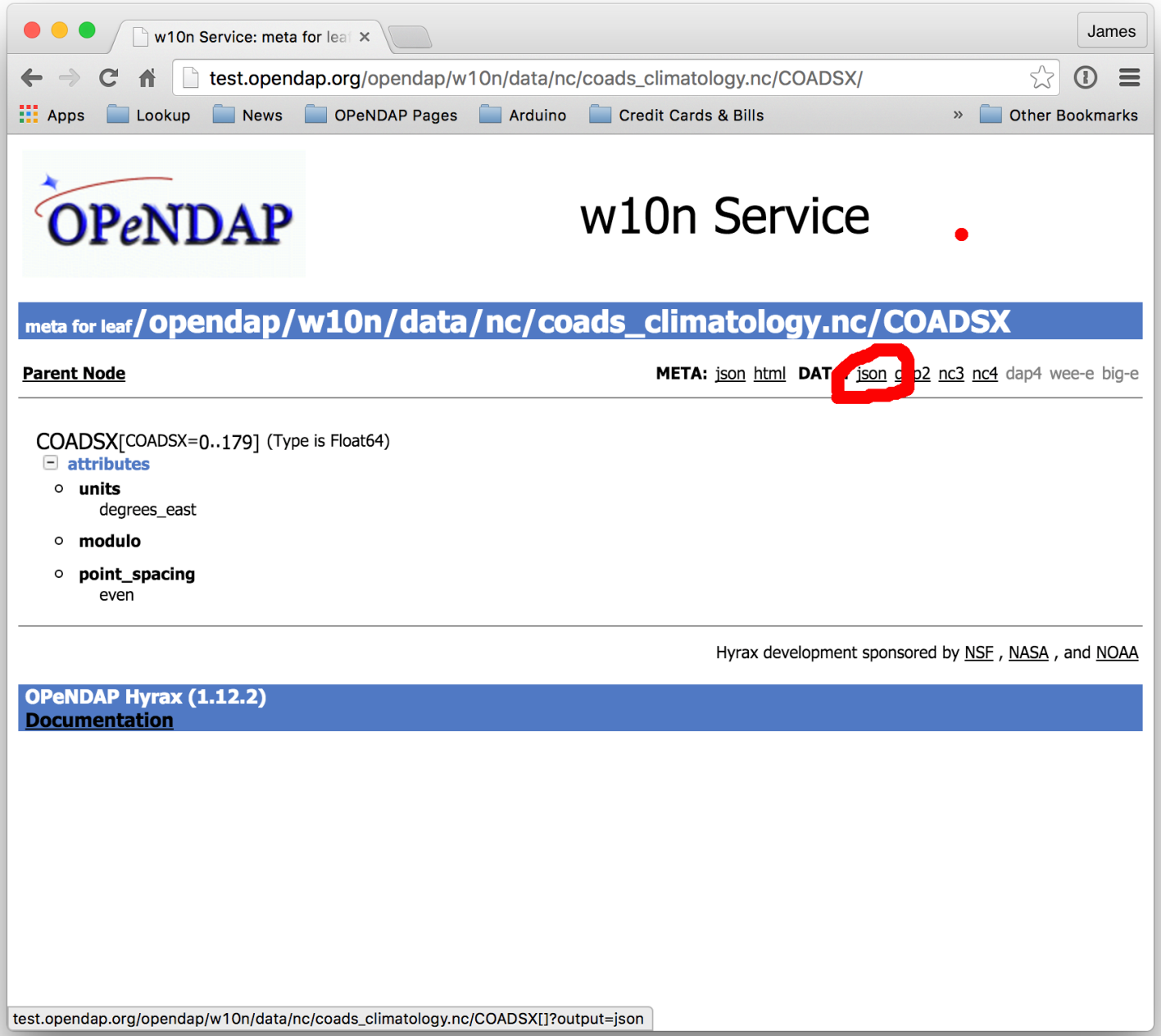
Web Services

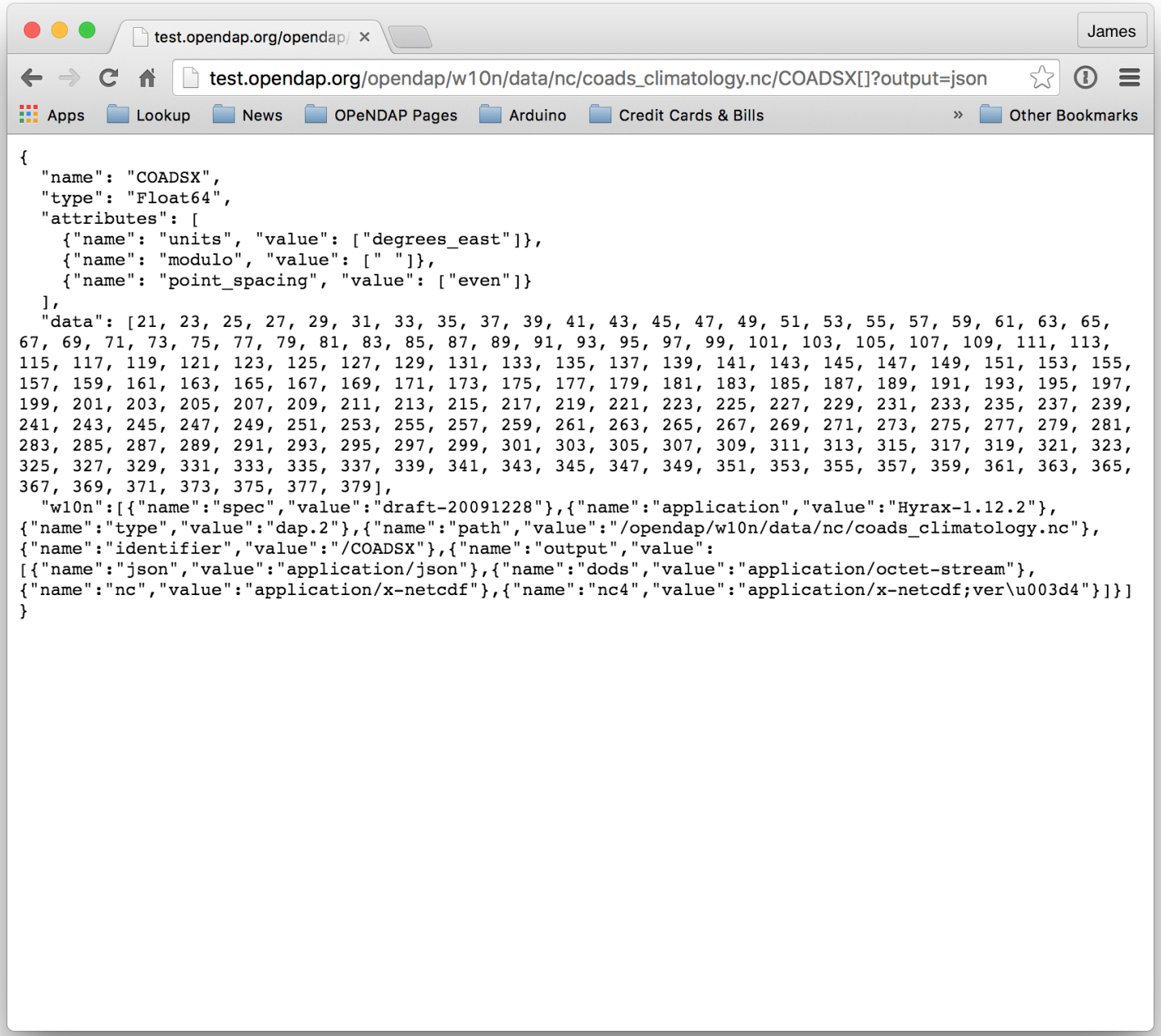
- Hyrax ships with support for w10n and WMS
- w10n - Webification
  - Use its JSON responses to build/control user interfaces
  - w10n supports navigating collections to get data
  - Its tree model extends **into** the granules, simplifying UI design & harmonizing data-storage schemes
- WMS
  - Maps: WMS works well with geospatial data that can be shown as a ‘map’ (but not other data types)
  - Google Earth: WMS offers a bridge to other tools…

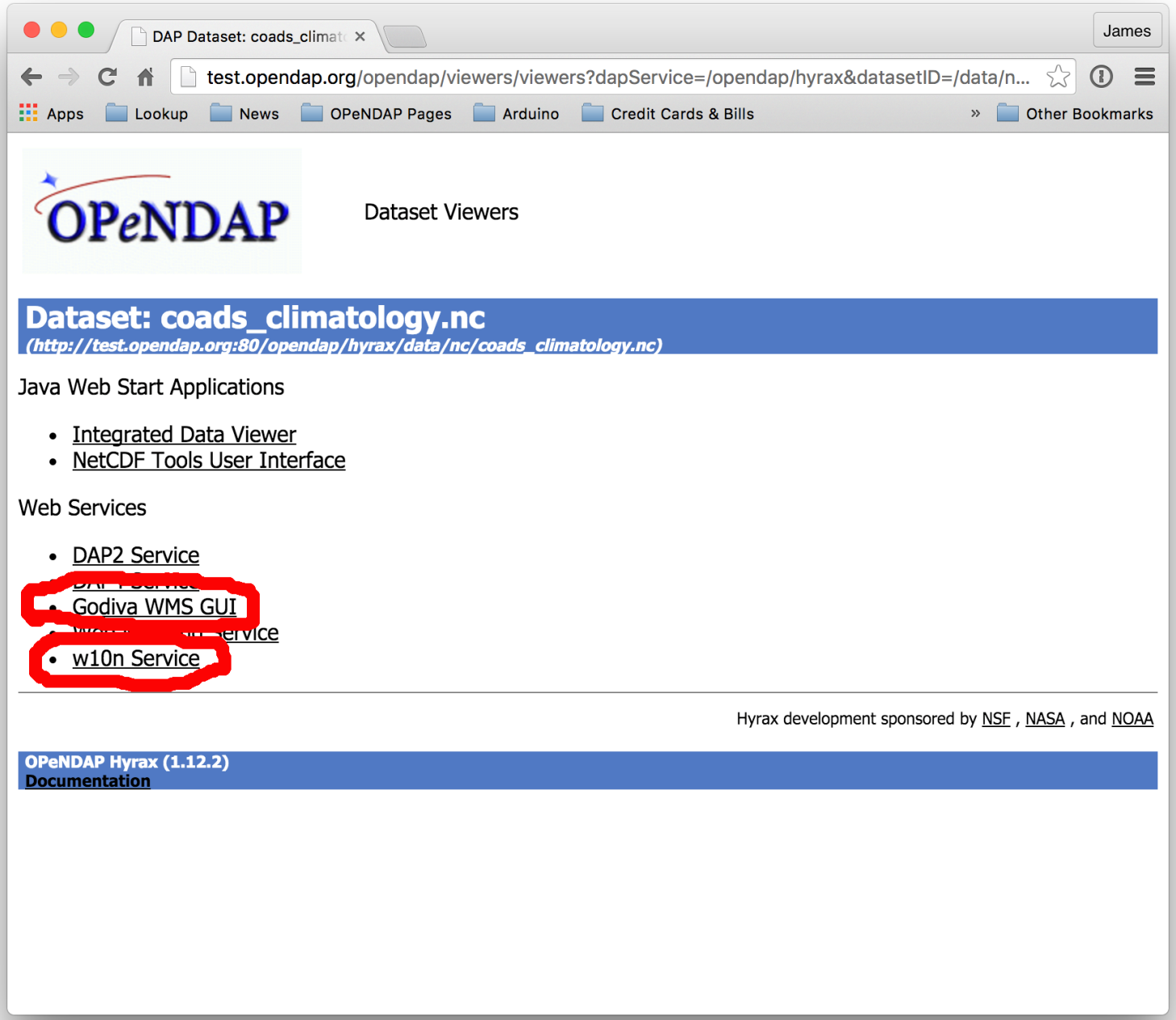

OPeNDAP Hyrax: Contents
James
test.opendap.org/dap/data/nc/contents.html
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
OPeNDAP
Contents of /data/nc
Name
Last Modified
Size
DAP Response Links
Dataset Viewers
bears.nc 2015-10-28T20:08:30 852 ddx dds das info html rdf viewers
bears.nc.das 2015-10-28T20:08:30 145 - - - - - -
coads_climatology.nc 2015-10-28T20:08:30 3114044 ddx dds das info html rdf viewers
feb.nc 2015-10-28T20:08:39 5992 ddx dds das info html rdf viewers
fnoc1.das 2015-10-28T20:08:30 162 - - - - - -
fnoc1.nc 2015-10-28T20:08:30 23944 ddx dds das info html rdf viewers
fnoc1.nc.html 2015-10-28T20:08:30 177 - - - - - -
jan.nc 2015-10-28T20:08:39 6580 ddx dds das info html rdf viewers
THREDDS Catalog XML
Hyrax development sponsored by NSF , NASA , and NOAA
OPeNDAP Hyrax (1.12.2)
Documentation

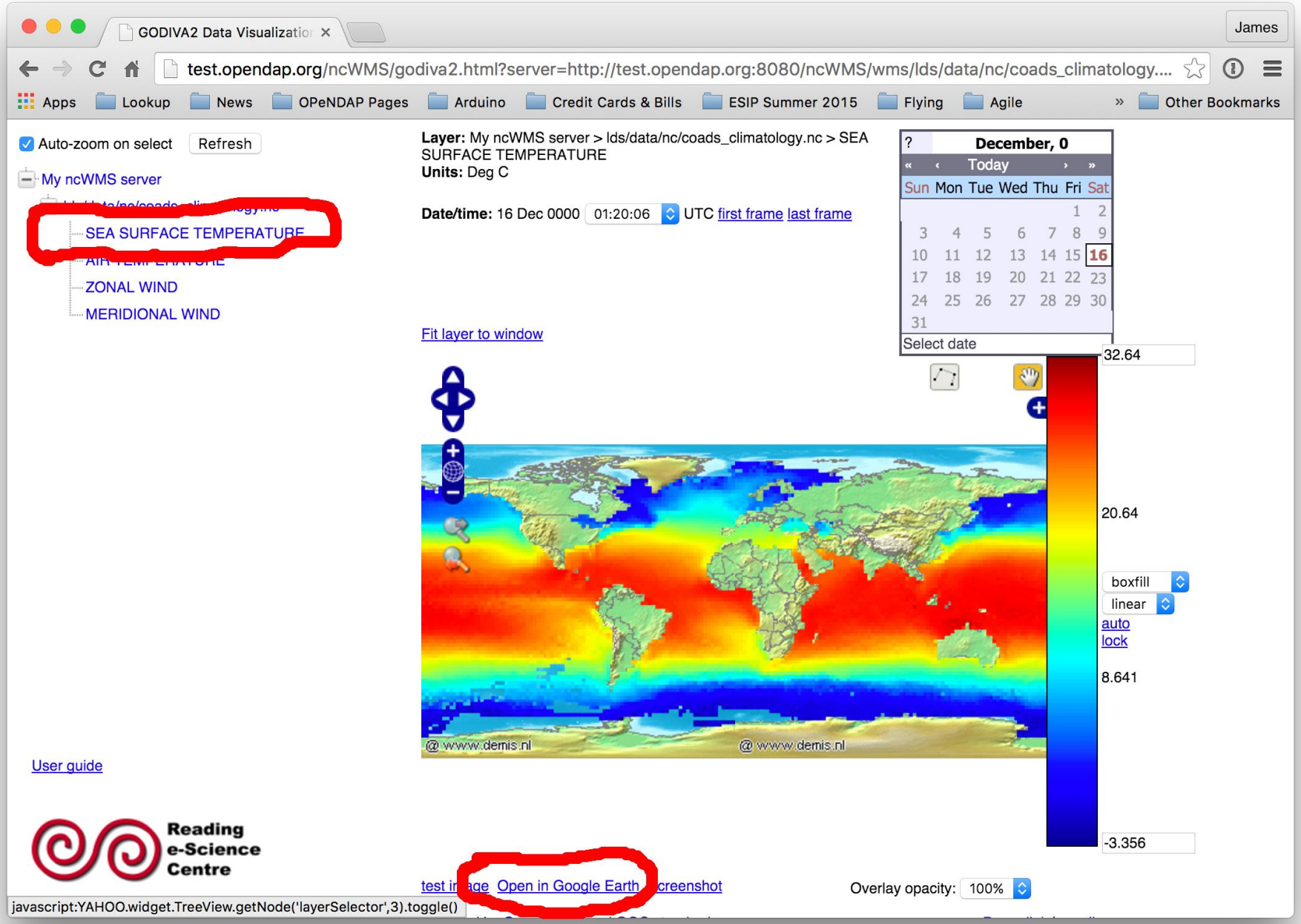

DAP Dataset: coads_climat
James
test.opendap.org/opendap/viewers/viewers?dapService=/opendap/hyrax&datasetID=/data/n...
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
OPeNDAP
Dataset Viewers
Dataset: coads_climatology.nc
(http://test.opendap.org:80/opendap/hyrax/data/nc/coads_climatology.nc)
Java Web Start Applications
Integrated Data Viewer
NetCDF Tools User Interface
Web Services
DAP2 Service
DAP4 Service
Godiva WMS GUI
Web Mapping Service
w10n Service
Hyrax development sponsored by NSF , NASA , and NOAA
OPeNDAP Hyrax (1.12.2)
Documentation

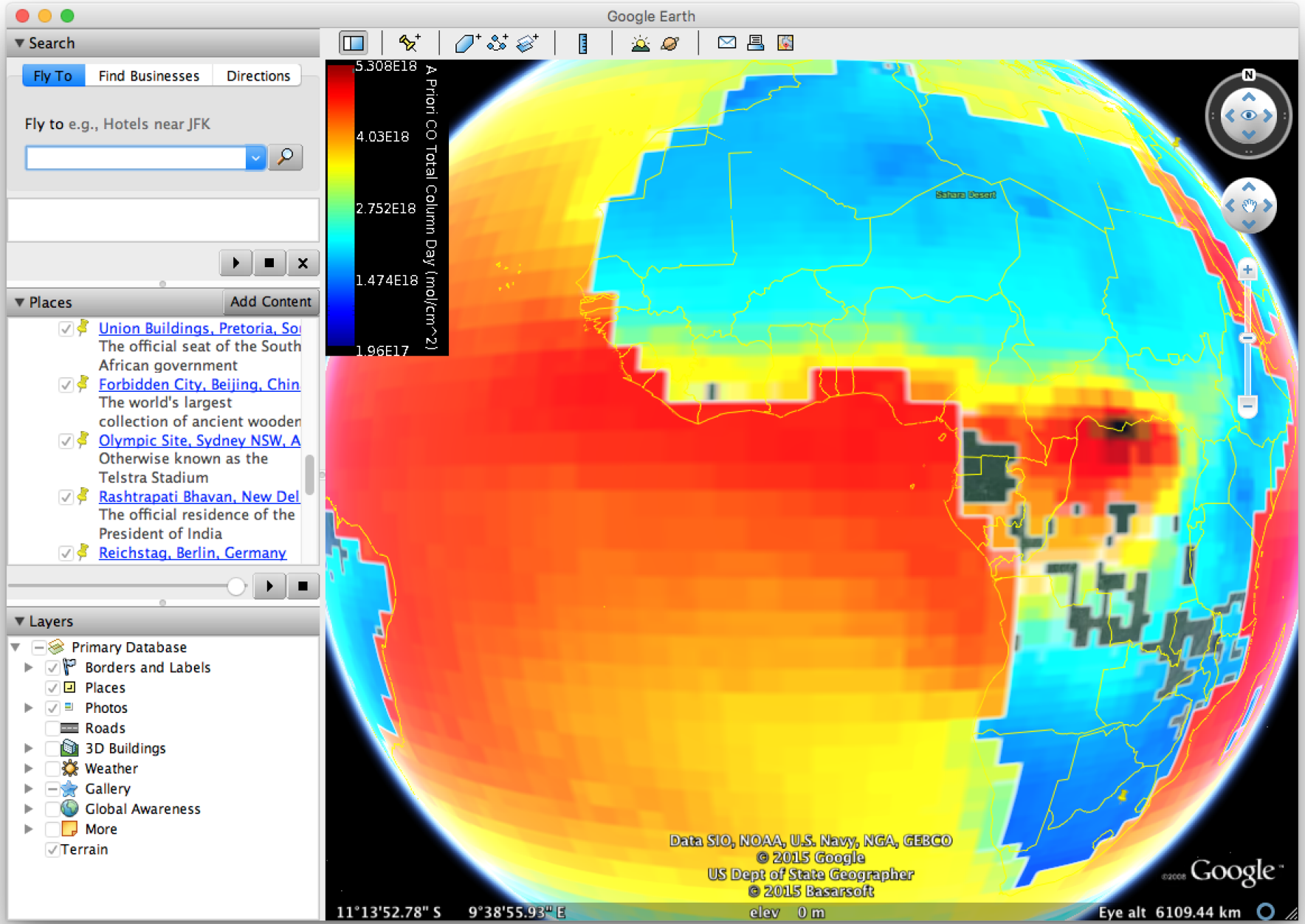

w10n Service: meta for noc
James
test.opendap.org/opendap/w10n/data/nc/coads_climatology.nc/
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
OPeNDAP
w10n Service
meta for node/opendap/w10n/data/nc/coads_climatology.nc
Parent Node
META: json html
Global Attributes
NC_GLOBAL
DODS_EXTRA
Variables
COADSX [COADSX=0..179] (Type is Float64)
COADSY [COADSY=0..89] (Type is Float64)
attributes
TIME [TIME=0..11] (Type is Float64)
attributes
SST (Type is node)
no attributes
members
AIRT (Type is node)
no attributes
members
UWND (Type is node)
no attributes
members
VWND (Type is node)
no attributes
members


w10n Service: meta for lea
James
test.opendap.org/opendap/w10n/data/nc/coads_climatology.nc/COADSX/
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
OPeNDAP
w10n Service
meta for leaf /opendap/w10n/data/nc/coads_climatology.nc/COADSX
Parent Node
META: json html DAT json nc3 nc4 dap4 wee-e big-e
COADSX[COADSX=0..179] (Type is Float64)
attributes
units
degrees_east
modulo
point_spacing
even
Hyrax development sponsored by NSF , NASA , and NOAA
OPeNDAP Hyrax (1.12.2)
Documentation
test.opendap.org/opendap/w10n/data/nc/coads_climatology.nc/COADSX[]?output=json


test.opendap.org/opendap/
James
test.opendap.org/opendap/w10n/data/nc/coads_climatology.nc/COADSX[]?output=json
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
{
"name": "COADSX",
"type": "Float64",
"attributes": [
{"name": "units", "value": ["degrees_east"]},
{"name": "modulo", "value": [" "]},
{"name": "point_spacing", "value": ["even"]}
],
"data": [21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65,
67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113,
115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155,
157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197,
199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239,
241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281,
283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323,
325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365,
367, 369, 371, 373, 375, 377, 379],
"w10n":[{"name":"spec","value":"draft-20091228"},{"name":"application","value":"Hyrax-1.12.2"},
{"name":"type","value":"dap.2"},{"name":"path","value":"/opendap/w10n/data/nc/coads_climatology.nc"},
{"name":"identifier","value":"/COADSX"},{"name":"output","value":
[{"name":"json","value":"application/json"},{"name":"dods","value":"application/octet-stream"},
{"name":"nc","value":"application/x-netcdf"},{"name":"nc4","value":"application/x-netcdf;ver\u003d4"}]}]
}


DAP Dataset: coads_climat
James
test.opendap.org/opendap/viewers/viewers?dapService=/opendap/hyrax&datasetID=/data/n...
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
Other Bookmarks
OPeNDAP
Dataset Viewers
Dataset: coads_climatology.nc
(http://test.opendap.org:80/opendap/hyrax/data/nc/coads_climatology.nc)
Java Web Start Applications
Integrated Data Viewer
NetCDF Tools User Interface
Web Services
DAP2 Service
Godiva WMS GUI
w10n Service
Hyrax development sponsored by NSF , NASA , and NOAA
OPeNDAP Hyrax (1.12.2)
Documentation


GODIVA2 Data Visualizatio
James
test.opendap.org/ncWMS/godiva2.html?server=http://test.opendap.org:8080/ncWMS/wms/lds/data/nc/coads_climatology....
Apps
Lookup
News
OPeNDAP Pages
Arduino
Credit Cards & Bills
ESIP Summer 2015
Flying
Agile
Other Bookmarks
Auto-zoom on select
Refresh
My ncWMS server
SEA SURFACE TEMPERATURE
ZONAL WIND
MERIDIONAL WIND
Layer: My ncWMS server > lds/data/nc/coads_climatology.nc > SEA SURFACE TEMPERATURE
Units: Deg C
Date/time: 16 Dec 0000 01:20:06 UTC first frame last frame
December, 0
Today
Sun Mon Tue Wed Thu Fri Sat
Select date
Fit layer to window
32.64
20.64
boxfill
linear
auto
lock
8.641
-3.356
@ www.demis.nl
User guide
Reading e-Science Centre
Open in Google Earth
Overlay opacity: 100%
javascript:YAHOO.widget.TreeView.getNode('layerSelector',3).toggle()


Google Earth
Search
Fly To
Find Businesses
Directions
Fly to e.g., Hotels near JFK
Places
Add Content
Union Buildings, Pretoria, So
The official seat of the South African government
Forbidden City, Beijing, Chin
The world's largest collection of ancient wooden
Olympic Site, Sydney NSW, A
Otherwise known as the Telstra Stadium
Rashtrapati Bhavan, New Del
The official residence of the President of India
Reichstag, Berlin, Germany
Layers
Primary Database
Borders and Labels
Places
Photos
Roads
3D Buildings
Weather
Gallery
Global Awareness
More
Terrain
5.308E18
4.03E18
2.752E18
1.474E18
1.96E17
A Priori CO Total Column Day (mol/cm^2)
Sahara Desert
Data SIO, NOAA, U.S. Navy, NGA, GEBCO
© 2015 Google
US Dept of State Geographer
© 2015 Basarsoft
Google
11°13'52.78" S 9°38'55.93" E
elev 0 m
Eye alt 6109.44 km

This work is supported by NASA/GSFC under Raytheon Co. contract number NNG15HZ39C