EOSDIS
NASA'S EARTH OBSERVING SYSTEM
DATA AND INFORMATION SYSTEM

Efficiently serving HDF5 via OPeNDAP

Kent Yang

The HDF Group

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

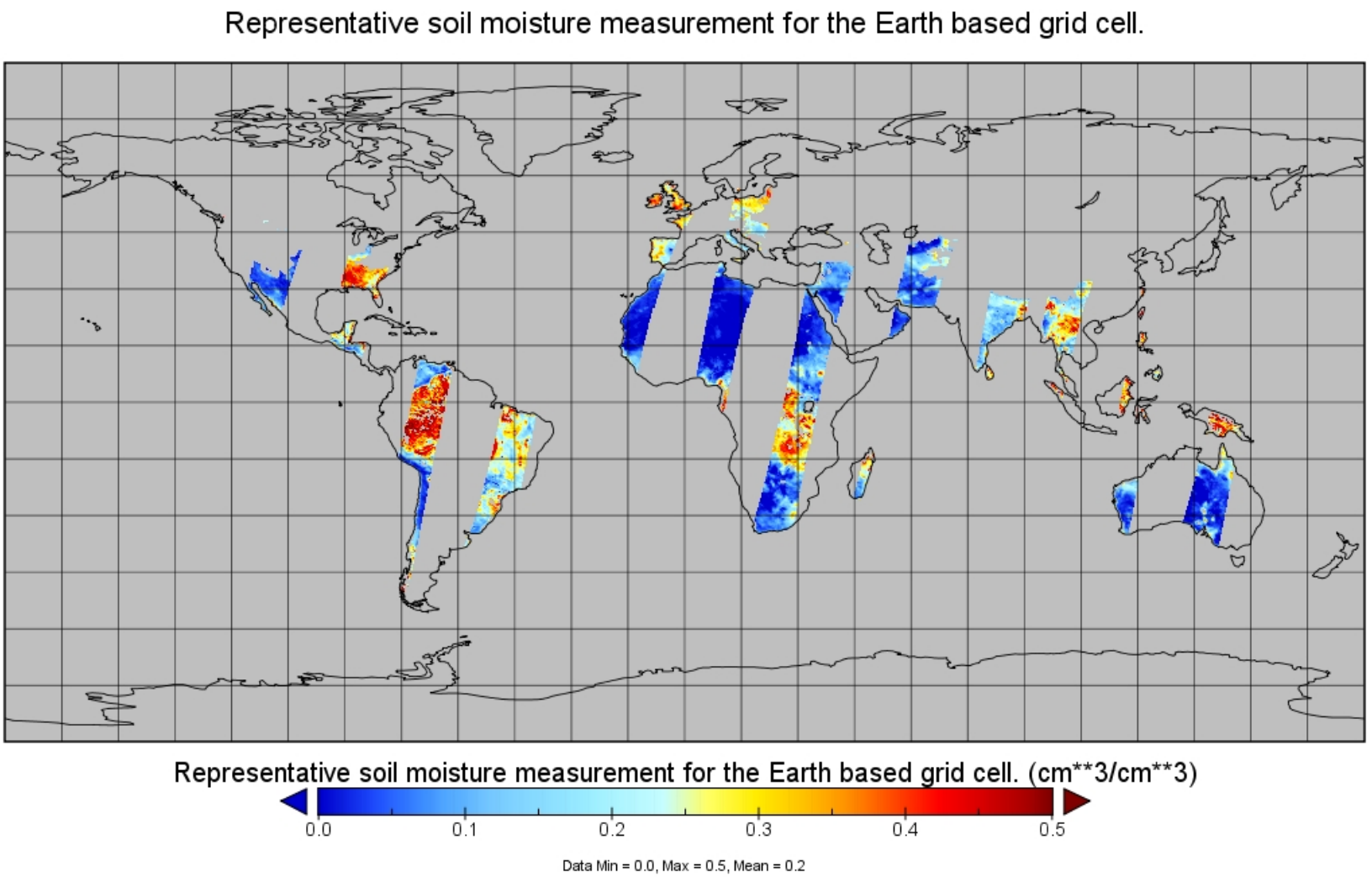
Why OPeNDAP[1]?

- Check metadata remotely (in various forms)
- Obtain the subset of data easily and efficiently
- Hide the original data sources
  - Hierarchical Data Format (HDF) versions 4 and 5
  - Network Common Data Form (NetCDF)
  - Geospatial Tagged Image File Format (GeoTiFF)
- Many popular earth science tools can visualize and analyze the data via OPeNDAP
- OPeNDAP output (including subsets) can be downloaded as other formats

HDF(5) Hyrax modules

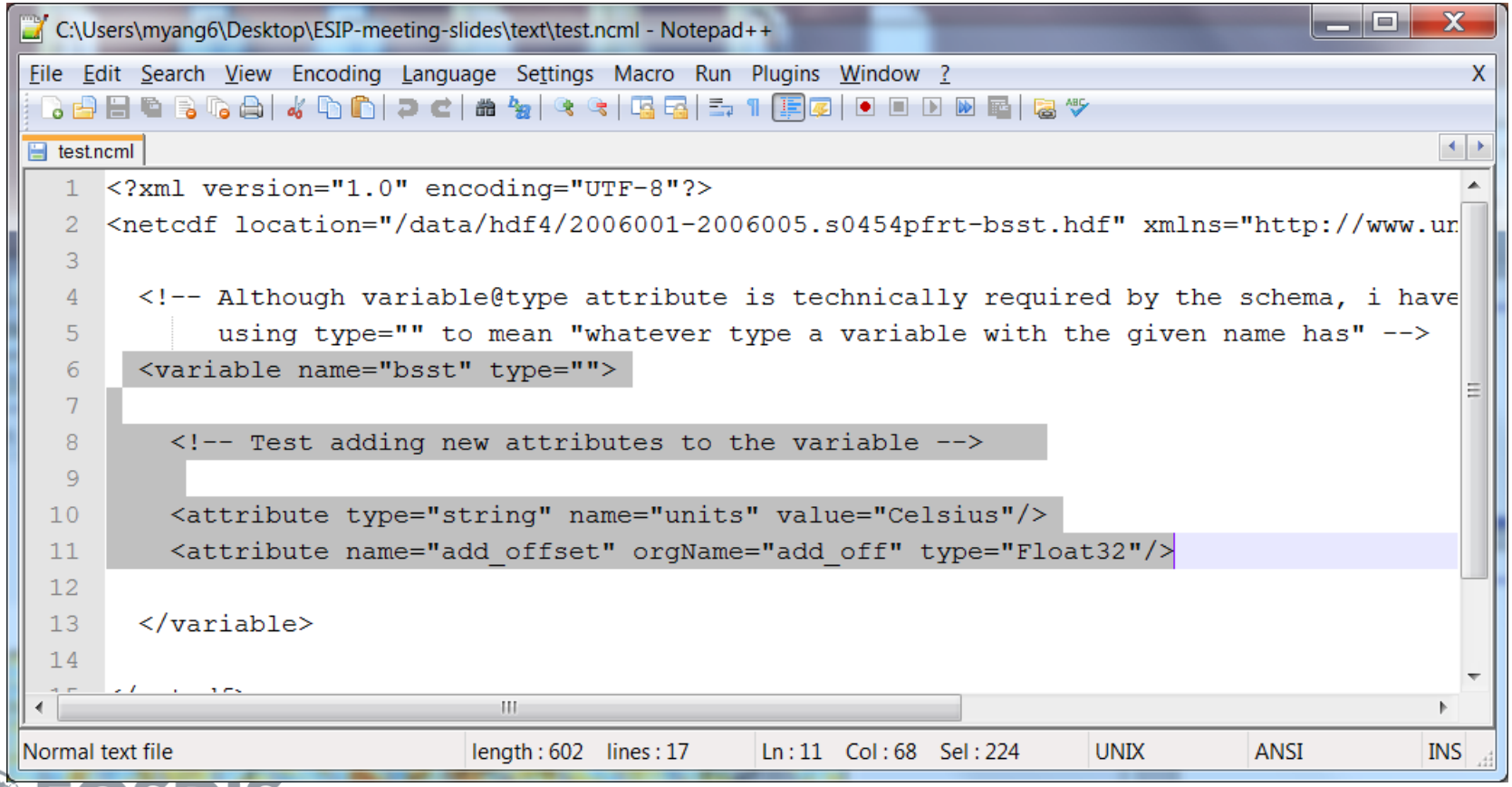
- The visualization of HDF(5) data via Hyrax
  - What NASA users request the most

Visualize A **S**oil **M**oisture **A**ctive **P**assive (SMAP) HDF5 variable via Hyrax



HDF5 handler and NcML[1]

- NcML module can be used with HDF handlers to provide the missing Climate and Forecast Metadata (CF) conventions information

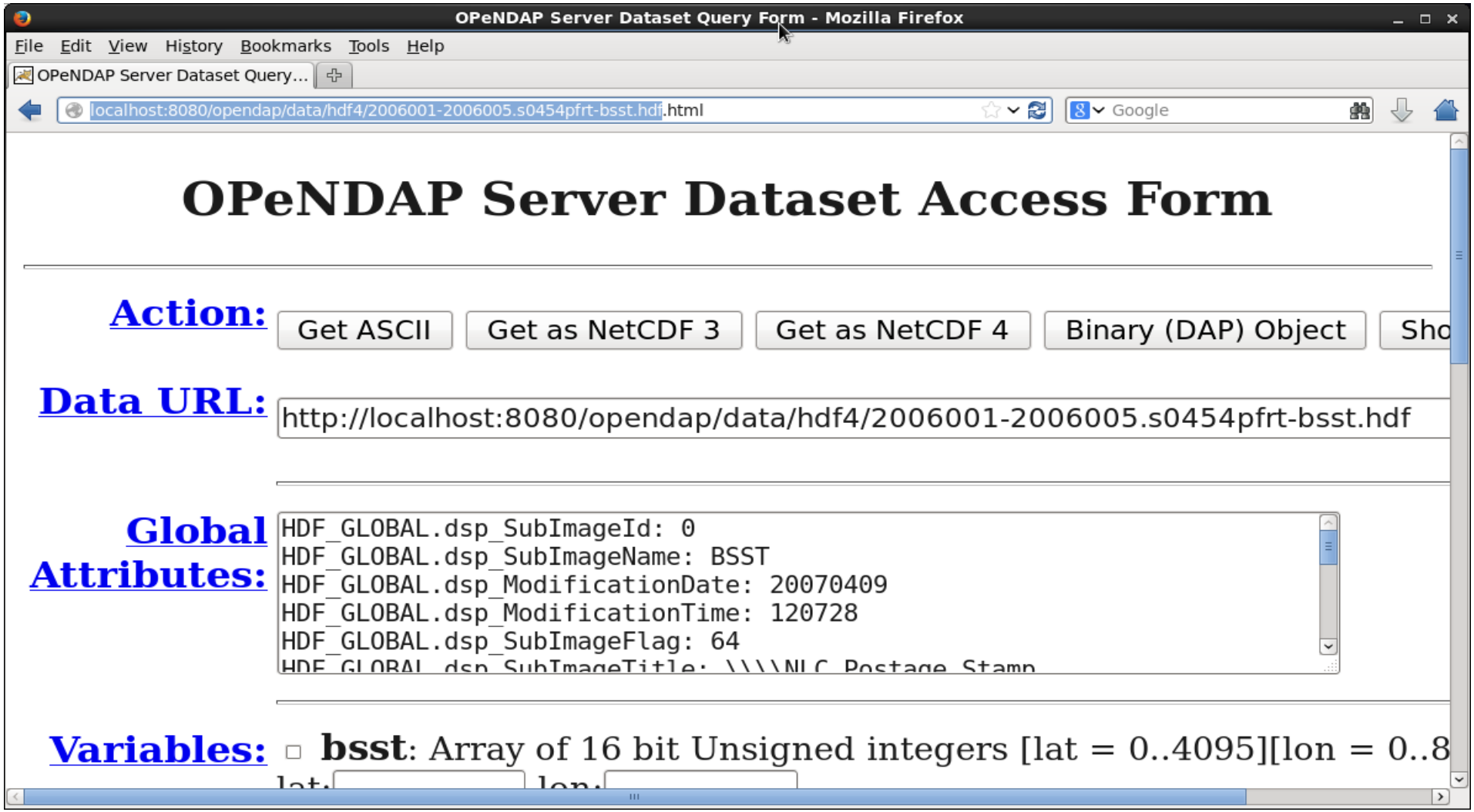


[1] **N**et**C**DF **M**arkup **L**anguage

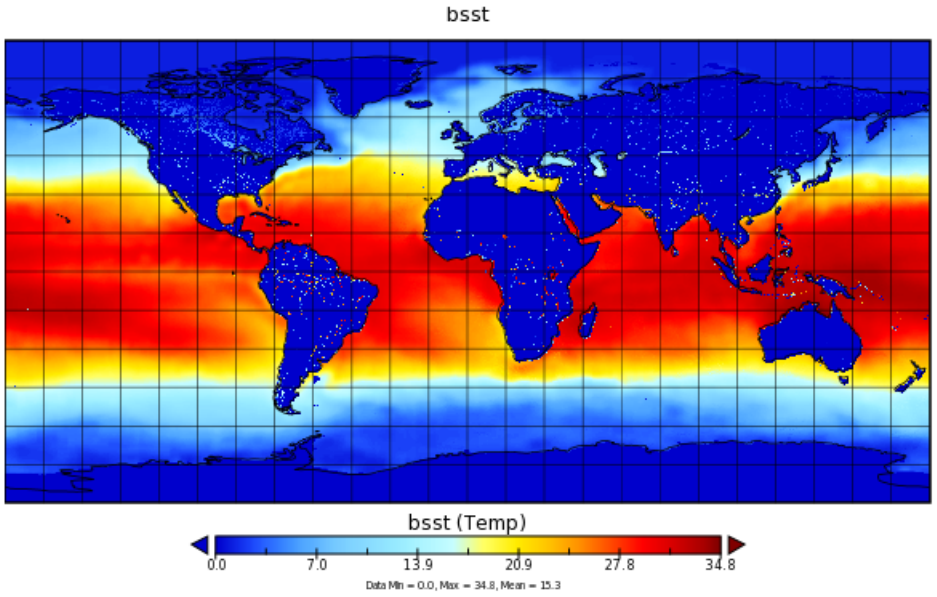
HDF5 handler and file NetCDF

- File NetCDF module can work with HDF handlers to convert HDF files to NetCDF-3 or NetCDF-4 classic files that follow the CF conventions.

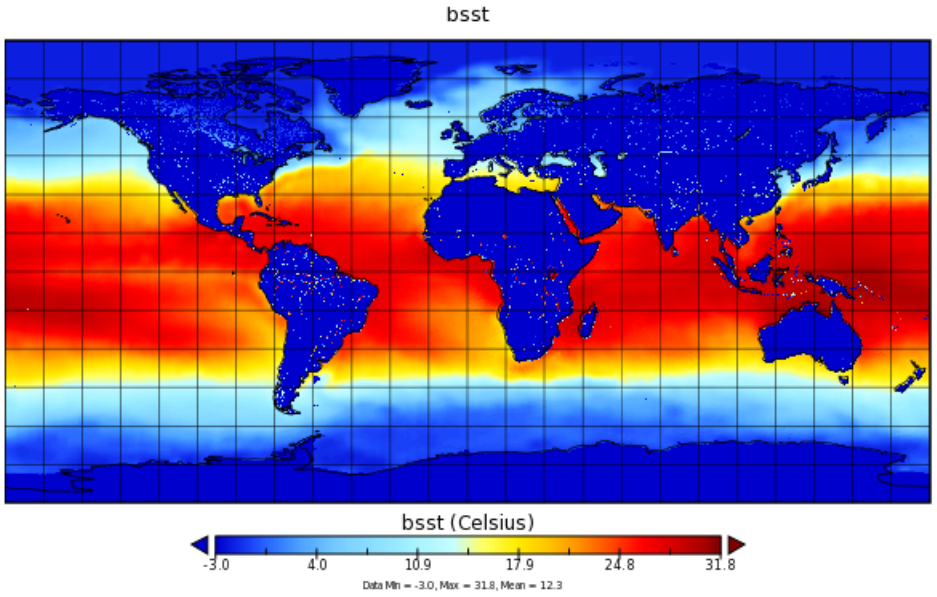
HDF5 handler and file NetCDF



Use NcML and file NetCDF to work with HDF5 handler



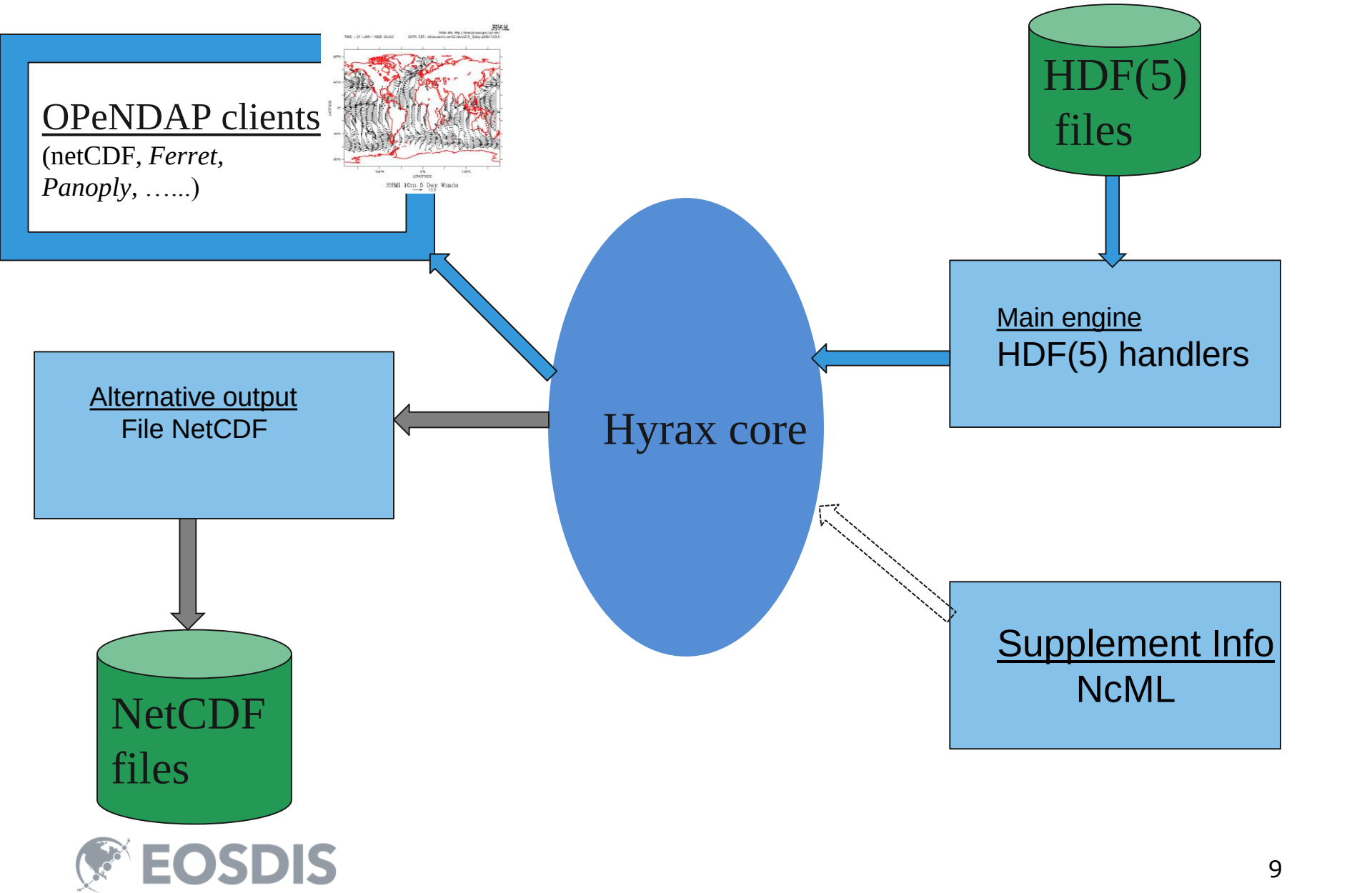
AVHRR[1] via Hyrax directly



The NetCDF file of AVHRR[1]
via Hyrax file NetCDF module

[1] **A**dvanced **V**ery **H**igh **R**esolution **R**adiometer

Service Chain to access HDF5 via Hyrax



DAP[1]4 support in the HDF5 handler

- CF option
  - DAP4 strictly mapped from DAP2
  - Dataset Metadata Response (DMR) replaces Dataset Descriptor Structure (DDS) and Dataset Attribute Structure (DAS)
- Default option
  - HDF5 group to DAP4 group
  - HDF5 signed 8-bit and 64-bit integers to DAP4
  - HDF5 dimensions following the NetCDF-4 to DAP4 mapping
    - Fully support NetCDF-4 (both classic and enhanced)

[1] **D**ata **A**ccess **P**rotocol

Interoperability enhancement

- CF option with DAP2
  - Provide a way for service providers to check if there are any objects ignored when mapping from HDF5 to DAP2
    - An example: DAP2 doesn't support 64-bit integer, HDF5 supports

Performance Improvement

- Reducing DDS and DAS access time
  - Caching DDS and DAS in memory
  - Caching DAS in the disk
- Reducing data access time
  - Caching the raw in the disk
  - Best if the data is compressed

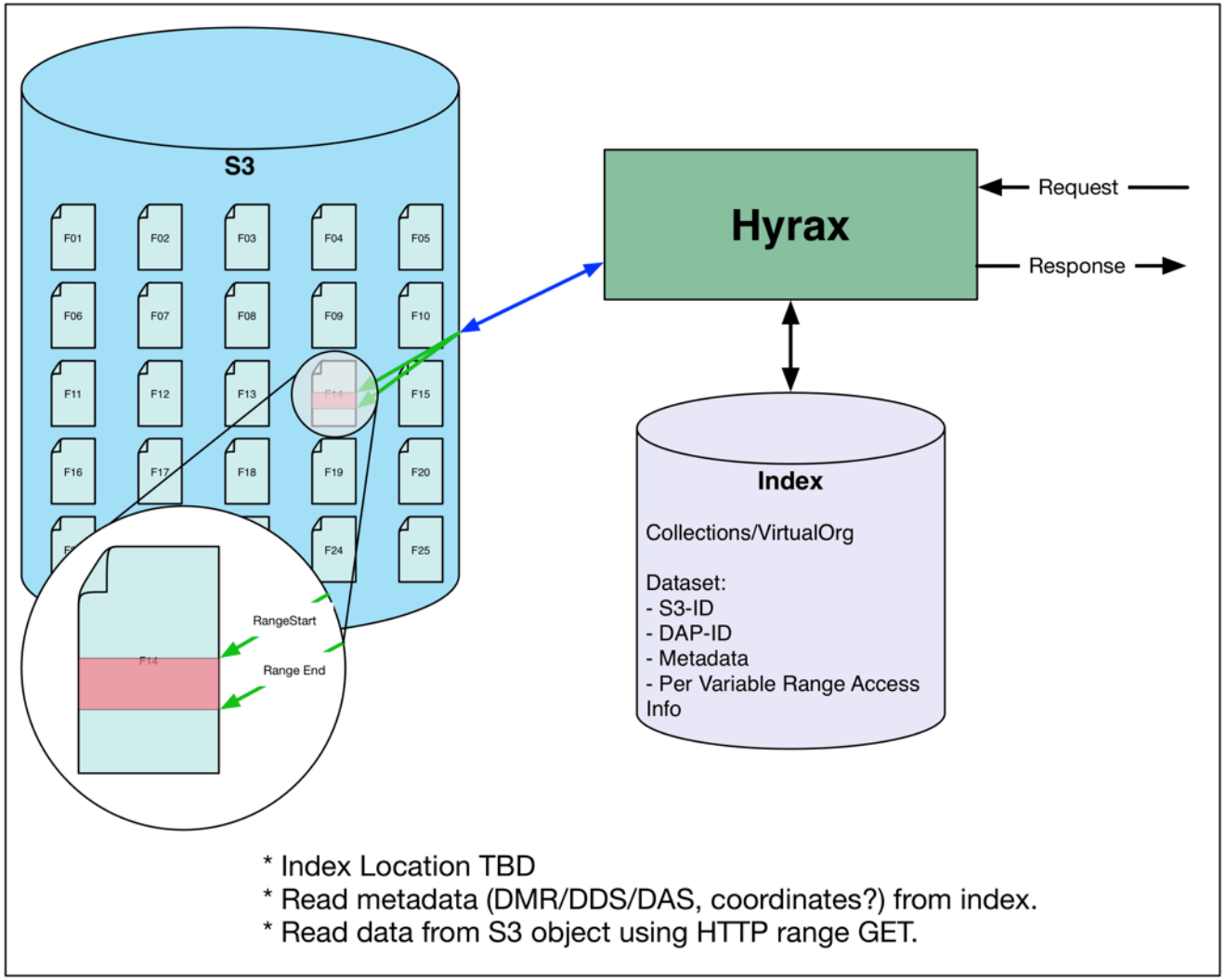
Other New Features

- Support the access of HDF-EOS5 sinusoidal projection in the HDF5 OPeNDAP handler

ACCESS HDF5 via Hyrax in Cloud

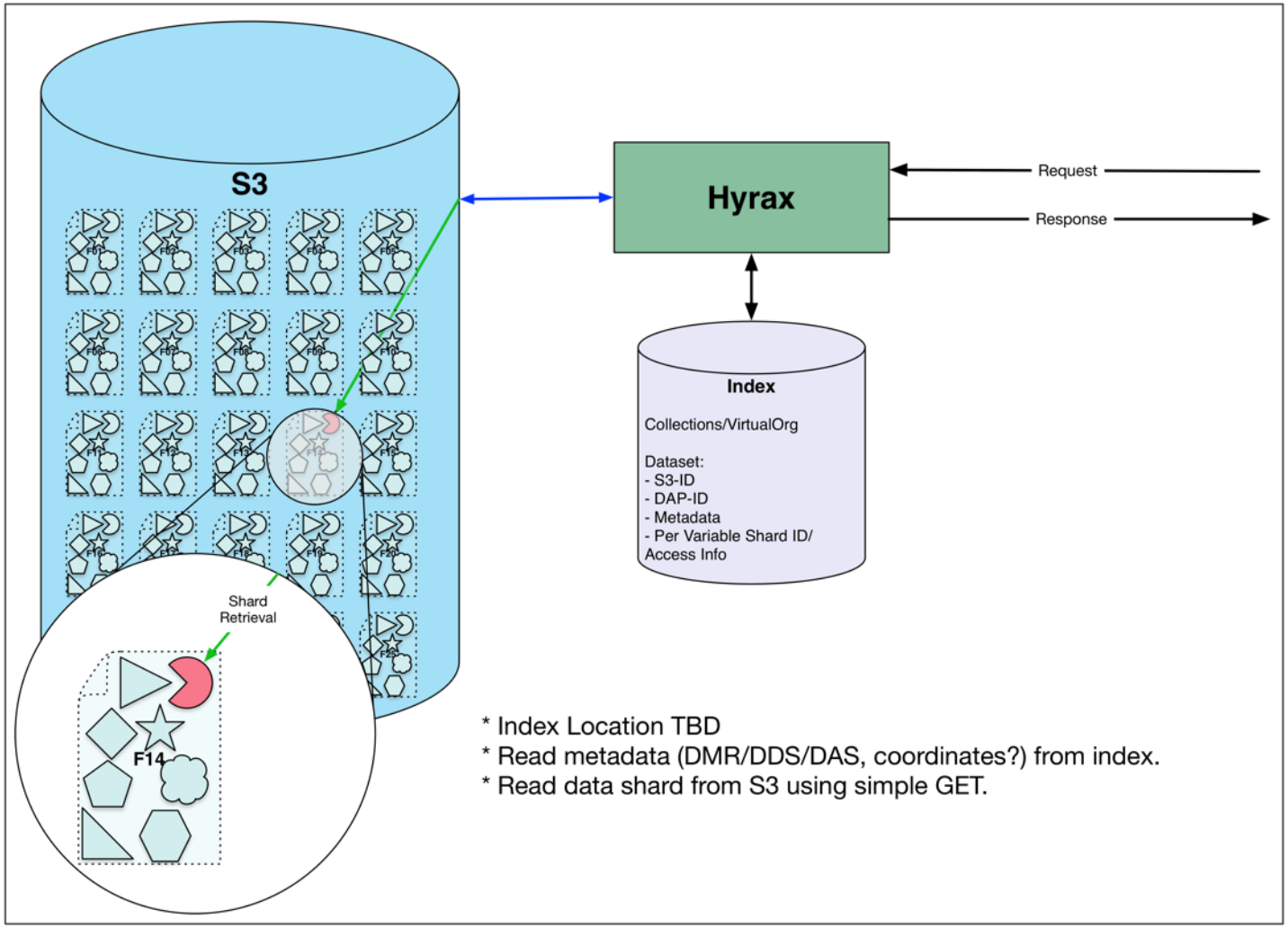
- Three architectures
- HDF5 handler can be enhanced for the future work of Architectures 2 and 3

Archit. #2: Files With HTTP[1] Range-Gets



1 Hypertext Transfer Protocol

Archit. #3: HDF5 Datasets as S3[1] Objects



What can be improved?

- Current
  - A2: Range-Gets index per chunk
  - A3: make an HDF5 chunk in a variable as an S3 object
- Future
  - Add an option to the HDF5 handler
    - Range-Gets index per variable
    - Make an HDF5 variable as an S3 object

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