Making Satellite Datasets Accessible for Everyone

A look into my NASA Internship – summer 2015

Aaron Scott

University of North Dakota

NASA Internship

- NASA/Univ. of Virginia HPC workshop
- Langley Research Center – Hampton, VA
- Atmospheric Science Data Center (ASDC)

Atmospheric Science Data Center

Overview

- "Responsible for processing, archival, and distribution of NASA Earth science data"
  - Radiation budget, clouds, aerosols, and tropospheric chemistry
- > 45 projects
- > 1700 archived data sets
- > 3 Petabytes (PB) of data

Atmospheric Science Data Center

Support

- NASA operates several different satellite missions
  - CALIPSO
  - CERES
  - MISR
  - MOPITT
- ASDC works with science teams
  - Provide data support and Processing
  - Archiving and subsetting

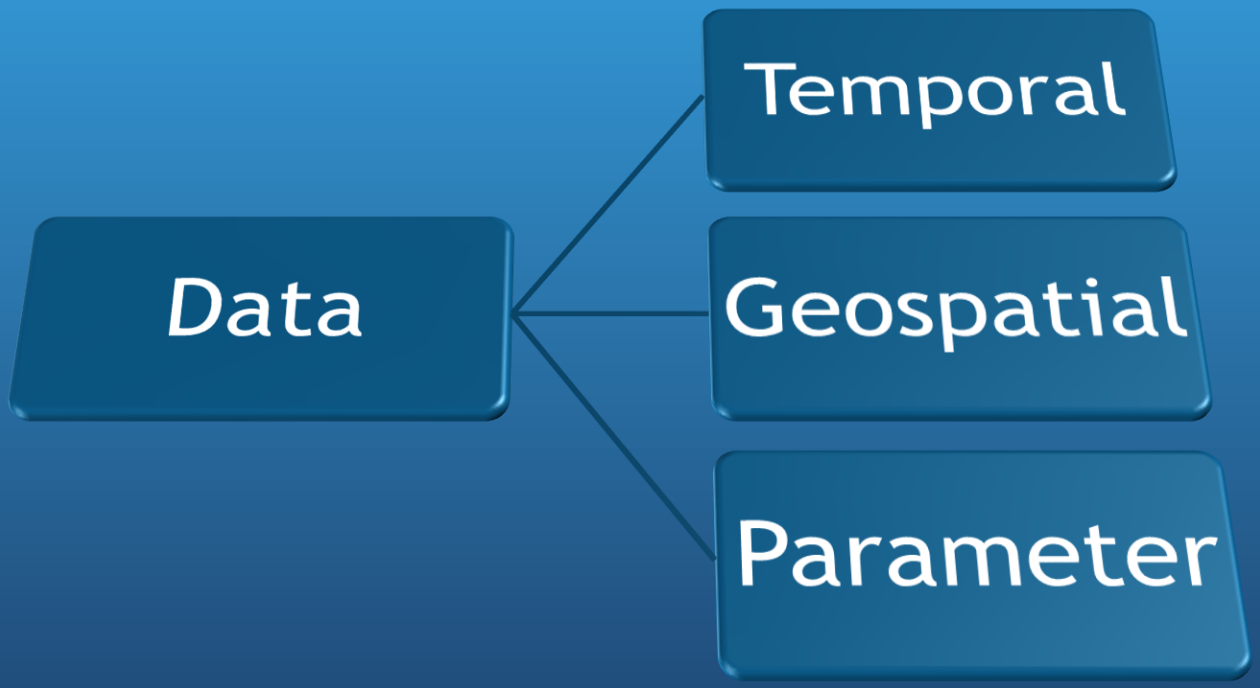
Atmospheric Science Data Center

Datasets

- Satellite datasets are inherently large
  - Data archives are growing daily
- Researchers need accessibility
- Let the scientists do science!
  - Analyze data not process
- Researchers only want necessary data

Atmospheric Science Data Center

Subsetting



Atmospheric Science Data Center

Subsetting

- Subset – a set that is part of a larger set

Atmospheric Science Data Center

Subsetting

- Subset – a set that is part of a larger set

Atmospheric Science Data Center

Subsetting

- Subset – a set that is part of a larger set

Atmospheric Science Data Center

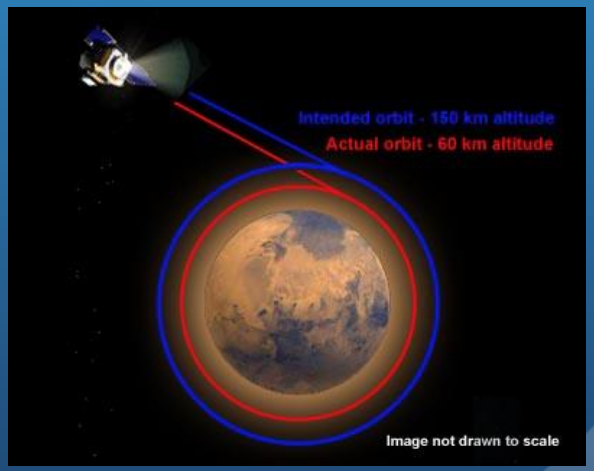
Making Data Useful

- Data collection without communication is … pointless?
- Data needs to be in some format that is useful
- Standard names
- Standard dimensions
- Units

Atmospheric Science Data Center

Making Data Useful

- And units are important!
- Mars Climate Orbiter
- Navigation error
  - English and metric units
  - 60 miles off trajectory



Source:
http://personal.victoria.ac.nz/stephen_marshall/SE/Failures/SE_MCO.html

Atmospheric Science Data Center

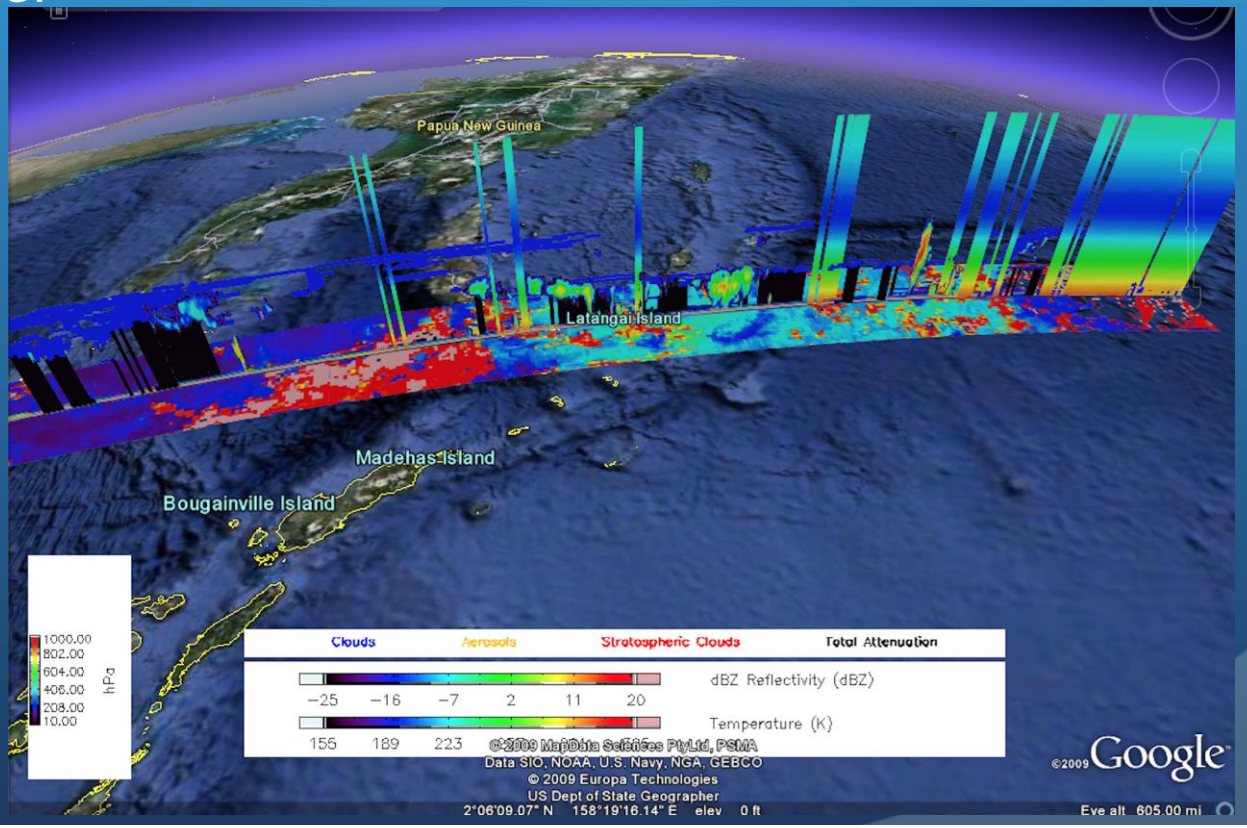
Making Data Useful

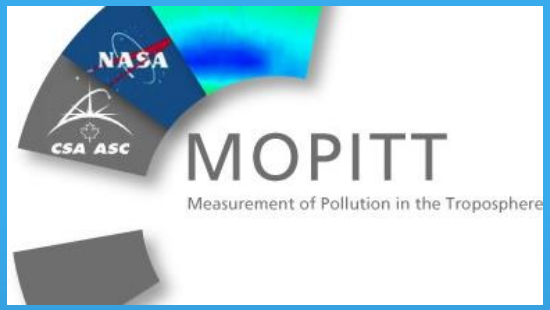
- CF (Climate and Forecast) Compliance
- Metada standards
  - Data that tells us about data (units, etc.)
- Standard names
- Standard Dimensions

Atmospheric Science Data Center

Making Data Useful

- Data compliance makes using data and visualization easier

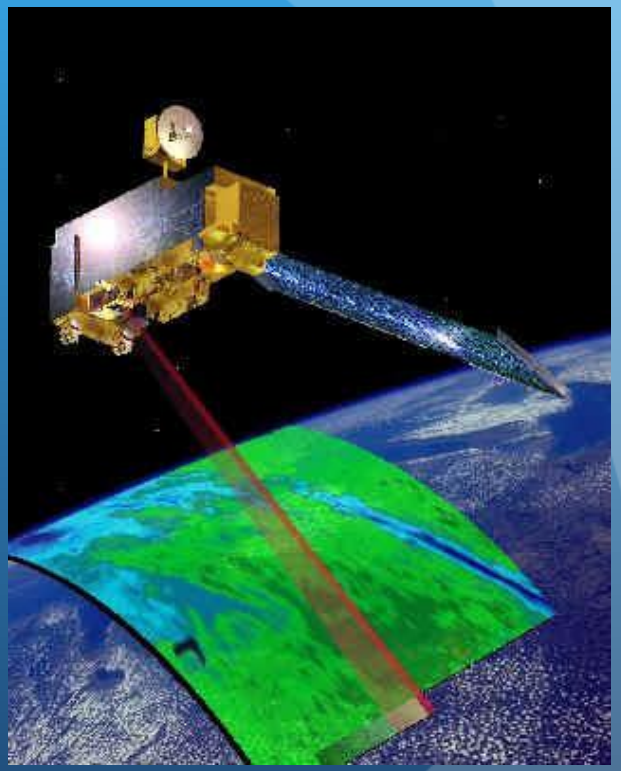




- Measurment of Pollution in the Troposphere
- Launched on TERRA in 1999
- Canadian Space Agency (CSA)
- National Center for Atmos. Research (NCAR)/NASA
- Measures carbon monoxide and methane

MOPITT
Measurement of Pollution in the Troposphere

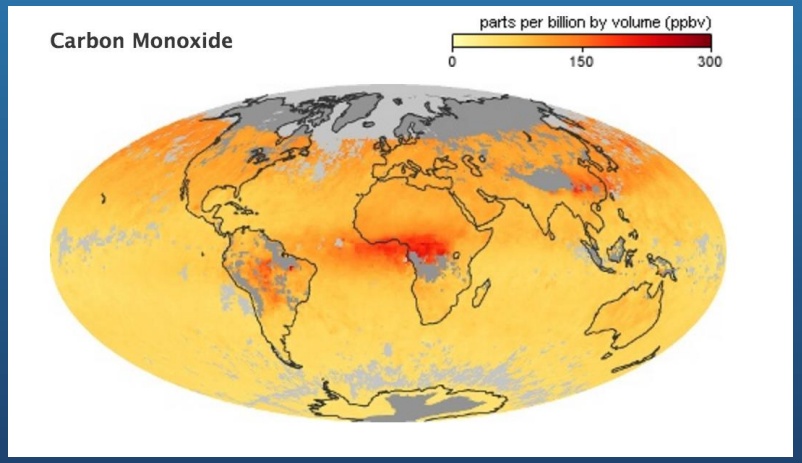
- Carbon monoxide profiles
  - Horizontal Resolution = 22 km
  - Vertical Resolution = 3 km
- High enough resolution to track the gas to certain cities or other sources
- Swath = 640 km wide



Source:
http://www.atmosp.physics.utoronto.ca/MOPITT/home.html

MOPITT
Measurement of Pollution in the Troposphere

- Carbon monoxide is a colorless, odorless, and poisonous gas
- Burning of fossile fuels
- Leads to atmospheric conditions such as smog



MOPITT Monthly Average for December 2015

MOPITT
Measurement of Pollution in the Troposphere

MOPITT Spatial Subset with Day-only Solar Zenith Angle Filter

Data granule:
MOP02T-20130811-L2V10.1.1.prov.hdf

Bounding Box (lon-lat pairs):
POLYGON((43 -12, 43 -26, 51 -26, 51 -12, 43 -12))

The original data granule (24 hours of data) contains 193192 observations.
The subsetted result file contains 1265 observations.

Atmospheric Science Data Center

HDF5/NetCDF4

- Hierarchical Data Format (HDF)
- Network Common Data Form (NetCDF)
- Common data structures used in the atmospheric sciences

Atmospheric Science Data Center

MOPITT Subsetter Algorithm

- MOPITT data subsetter at ASDC was upgraded to incorporate HDF5/NetCDF4
- Algorithm follows CF compliance
- Python code used as development language

Atmospheric Science Data Center

MOPITT Subsetter Algorithm

Data
- Not CF
- Large file

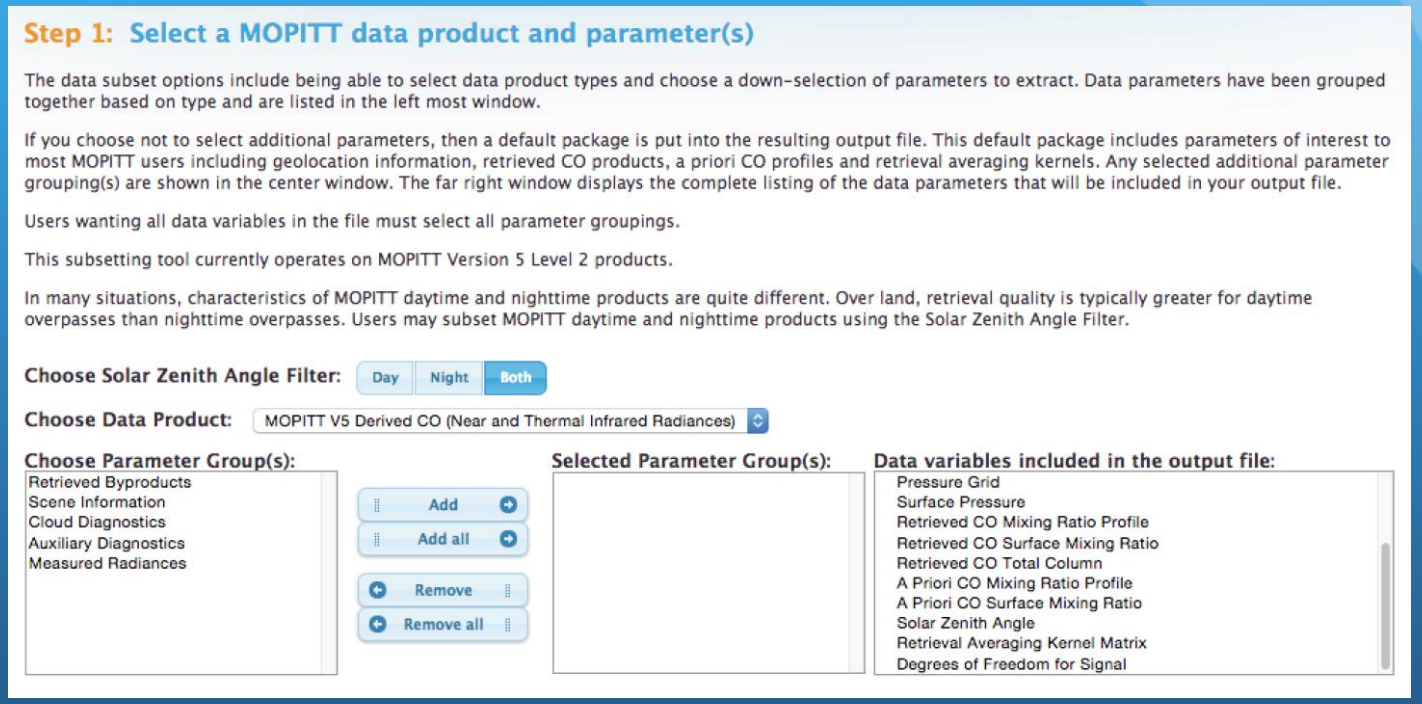
→

Algorithm
- Python code reads in xml order file

→

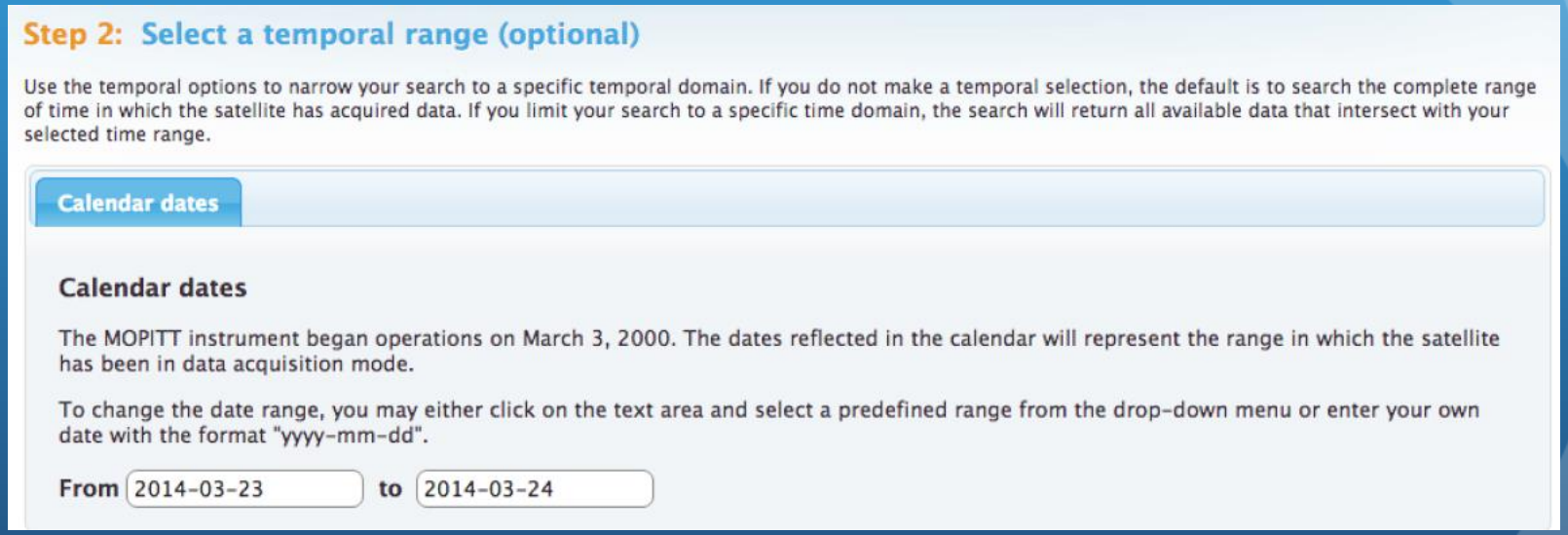
Data Subset
- CF
- Smaller file

MOPITT Subsetter Algorithm

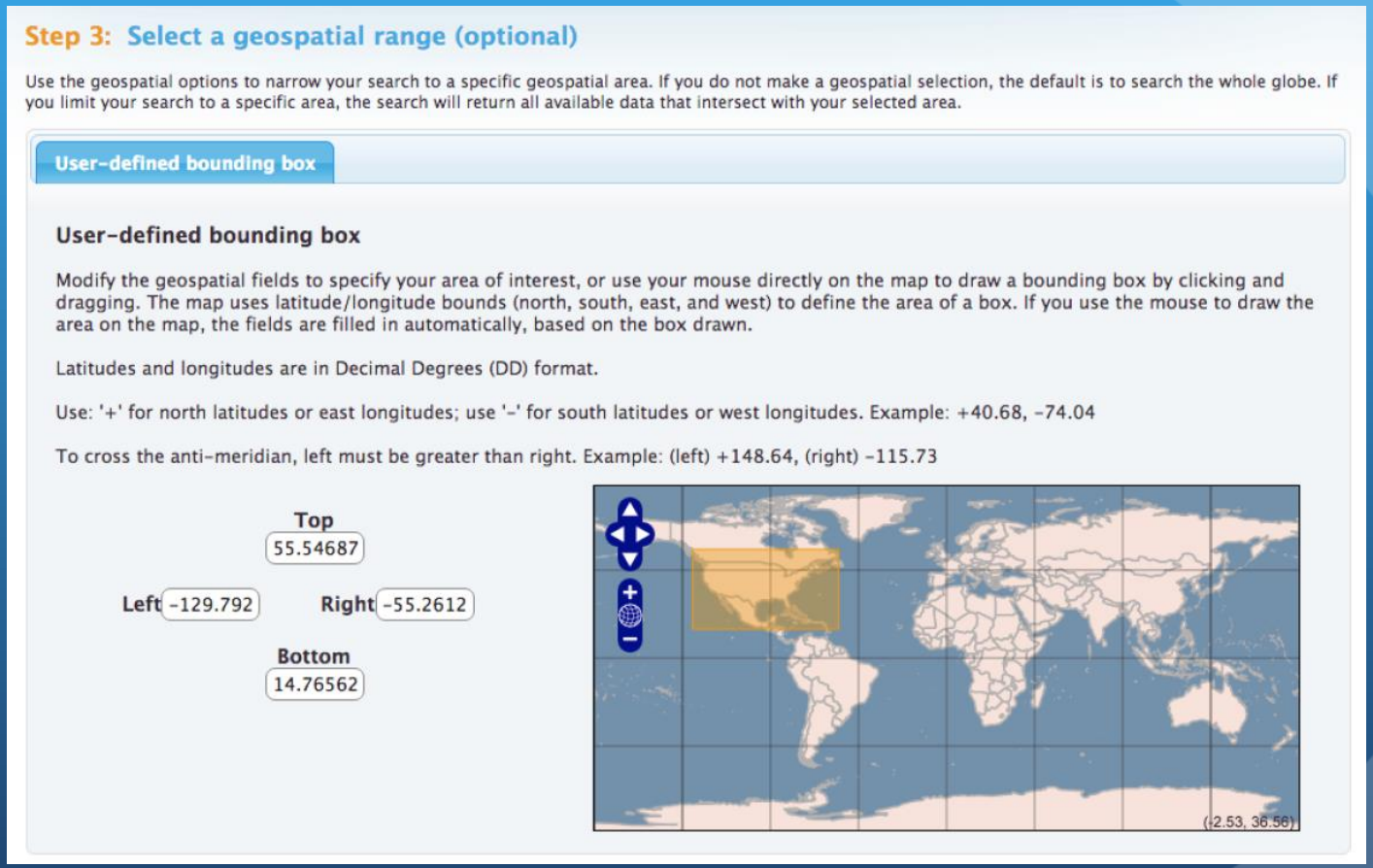


- Web based ordering platform
- It's like "take out" for scientists!

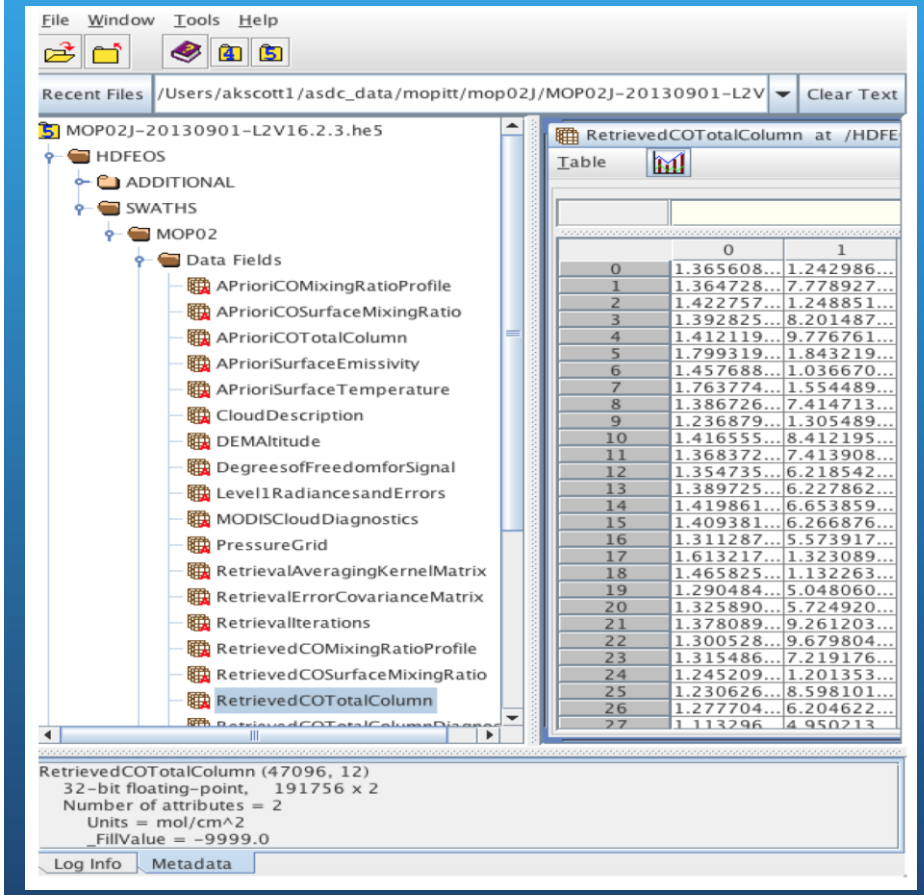
MOPITT Subsetter Algorithm

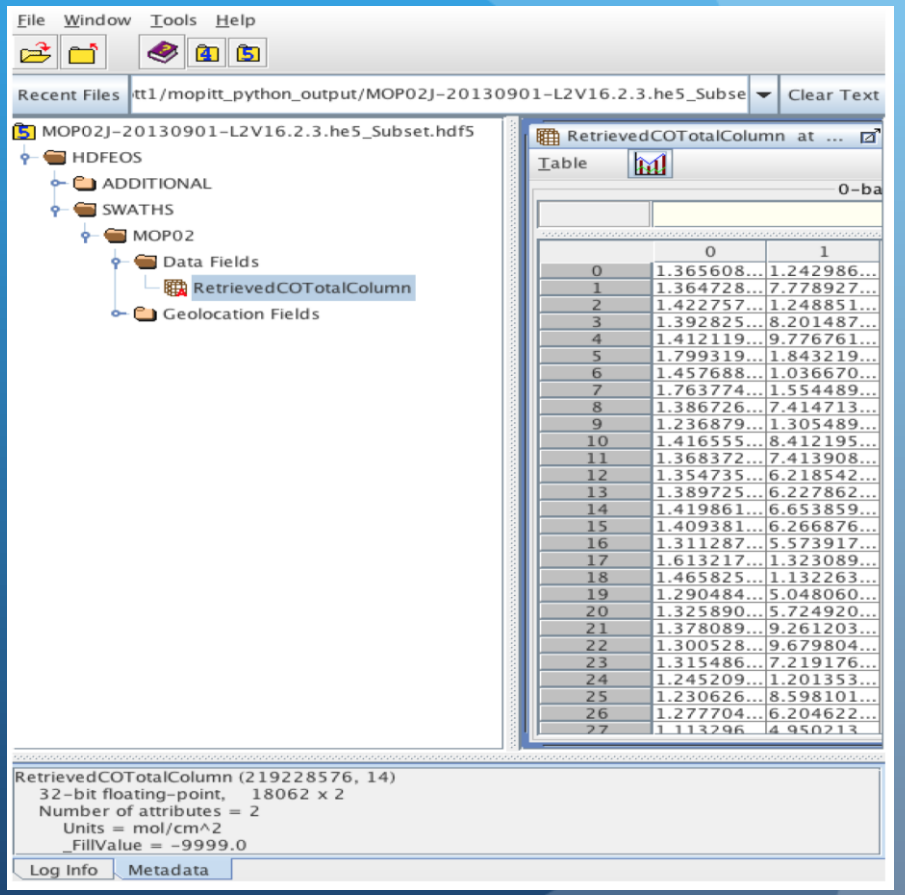


MOPITT Subsetter Algorithm



MOPITT Subsetter Algorithm





- View of data before and after passed through algorithm
- HDFView software

Atmospheric Science Data Center

Conclusions

- Making data more accessible requires:
  - Communication
  - Collaboration
  - Governance
- The efforts of the ASDC and similar organizations will continue to work to provide accessible data that can be used by everyone
- Furthering science starts with proper data management

Acknowledgements

- Walt Baskin – NASA Mentor
- ASDC
- ND Space Grant Consortium – THANK YOU!