Enabling Cloud Services and Enhanced Data Discovery with earthdata-varinfo

Eni Awowale[1,2], Carlee Loeser[1,2], Owen Littlejohns[3,4], Jennifer Adams[1,2], David Auty[3,5], Lena Iredell[1,2], Brianna Rita Pagán[1,2], Mahabal Hegde[1], Nicholas Lenssen IV[1,2]

[1] NASA Goddard Earth Sciences Data and Information Services Center (NASA GES DISC), [2] ADNET Systems, Inc.,
[3] NASA Earth Observing System Data and Information System (EOSDIS) Evolution and Development (NASA EED-3), [4] INNOVIM, [5] Raytheon Company

Introduction

NASA's Earth Observing System Data and Information System (EOSDIS) contains thousands of Earth science datasets from satellites, models, and field campaigns. Each of these collections can contain hundreds of variables that describe each measurement within the dataset, therefore an automated method for generating UMM-Var records is necessary. The Unified Metadata Model for Variables (UMM-Var) provides a framework for variable metadata records in NASA's Common Metadata Repository (CMR). The Python tool, earthdata-varinfo, was developed to solve this problem of automating the curation of UMM-Var records. Given either a collection DMR file or a netCDF-4 file, earthdata-varinfo can scrape variable metadata and return a CMR compliant UMM-Var record. Earthdata-varinfo can generate thousands of UMM-Var records in a matter of seconds, thus enabling subsetting capabilities and enhancing data discovery.

Motivation

1) Many scientists search for data based on variables or measurements, rather than collections or projects. Curating metadata for these variables is important for discovery.
2) Subsetting, regridding, and reformatting are important data transformation tools that allow scientists to reduce the size of data files and transform them for comparison to other data. To enable subsetting by variable, curating variable metadata is necessary.
3) Since some data collections have hundreds of variables in a single file, it is challenging to curate this metadata manually. Earthdata-varinfo extracts the CF attributes for all of the variables in a data file and outputs them into CMR-ready metadata records.

Workflow and Capabilities

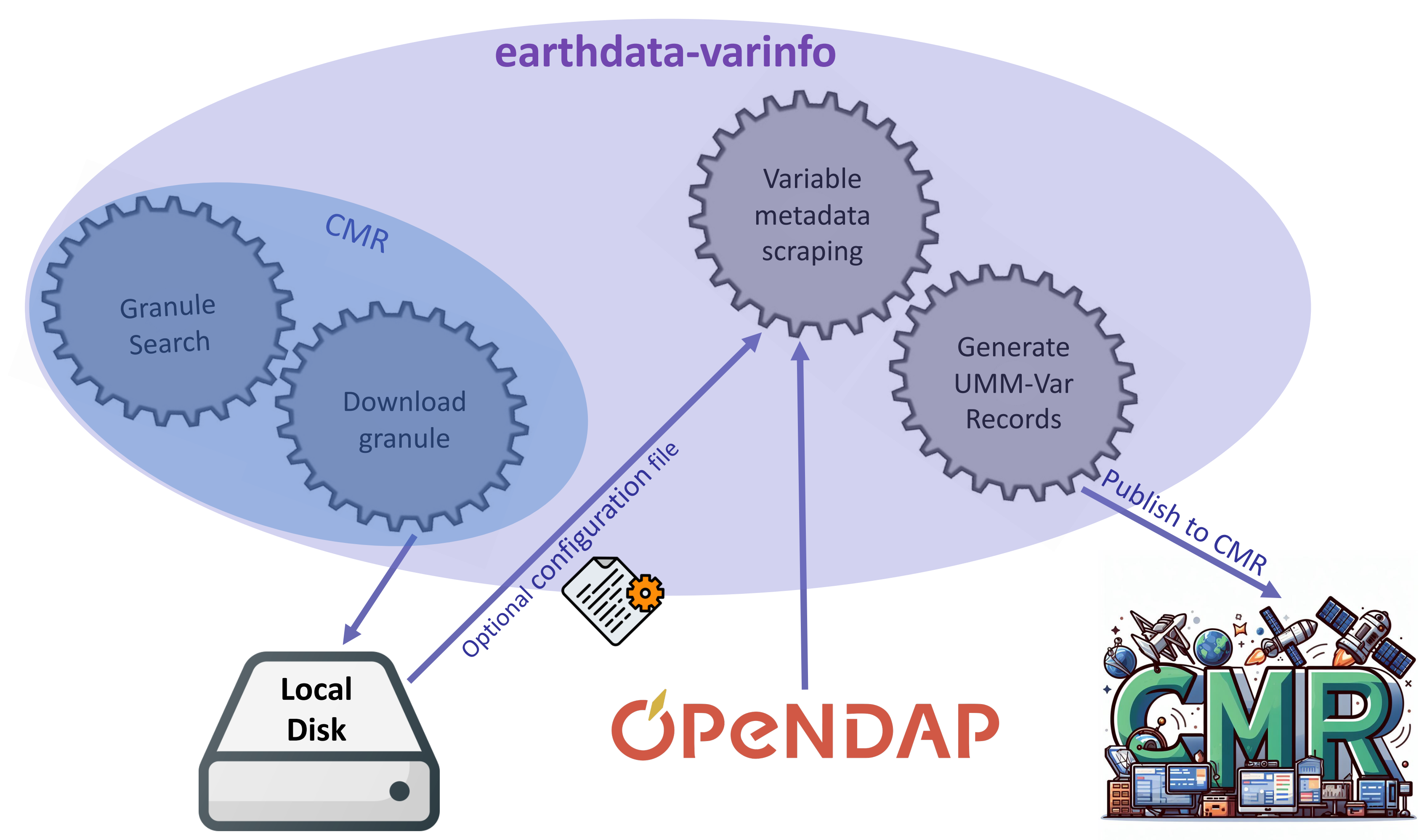


Figure 1. Workflow diagram for earthdata-varinfo

How does it work?

1) Earthdata-varinfo requires a single granule that is representative of the data collection and includes all its variables. There are three options:
   a) Conduct a CMR granule search to download the most recent granule from Earthdata Search locally.
   b) Choose the representative granule through OPeNDAP.
   c) Use a granule that is already stored locally.
2) Earthdata-varinfo can be run locally or in a Docker container. Collection level configuration files can be included to the workflow for known quirky datasets. Earthdata-varinfo returns a UMM-Var record for each variable in the granule that is CMR compliant.
3) UMM-Var records are published to CMR.

Alternatively, UMM-Var records can be generated via the CMR GraphQL API, which utilizes earthdata-varinfo.

What CF attributes are collected with earthdata-varinfo?

- long_name
- standard_name
- title (or definition, description)
- dtype (e.g. float32, int16)
- dimensions
- units
- _FillValue
- scale_factor
- add_offset
- valid_range

What are the compatible granule file formats?

- '.nc', '.nc4', '.grb', '.h5', '.he5', '.HDF5'
- OPeNDAP file types (e.g. '.dmr')

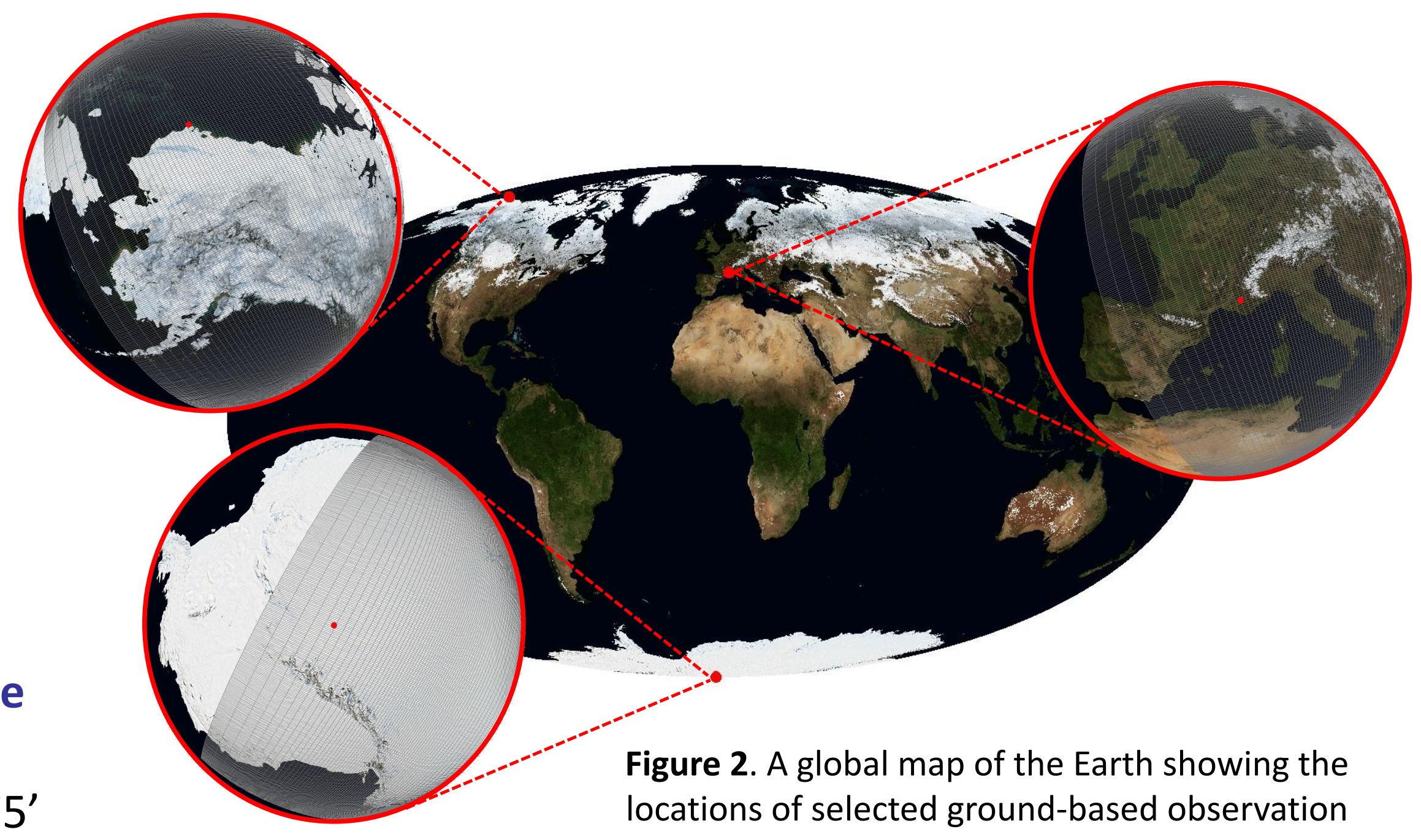


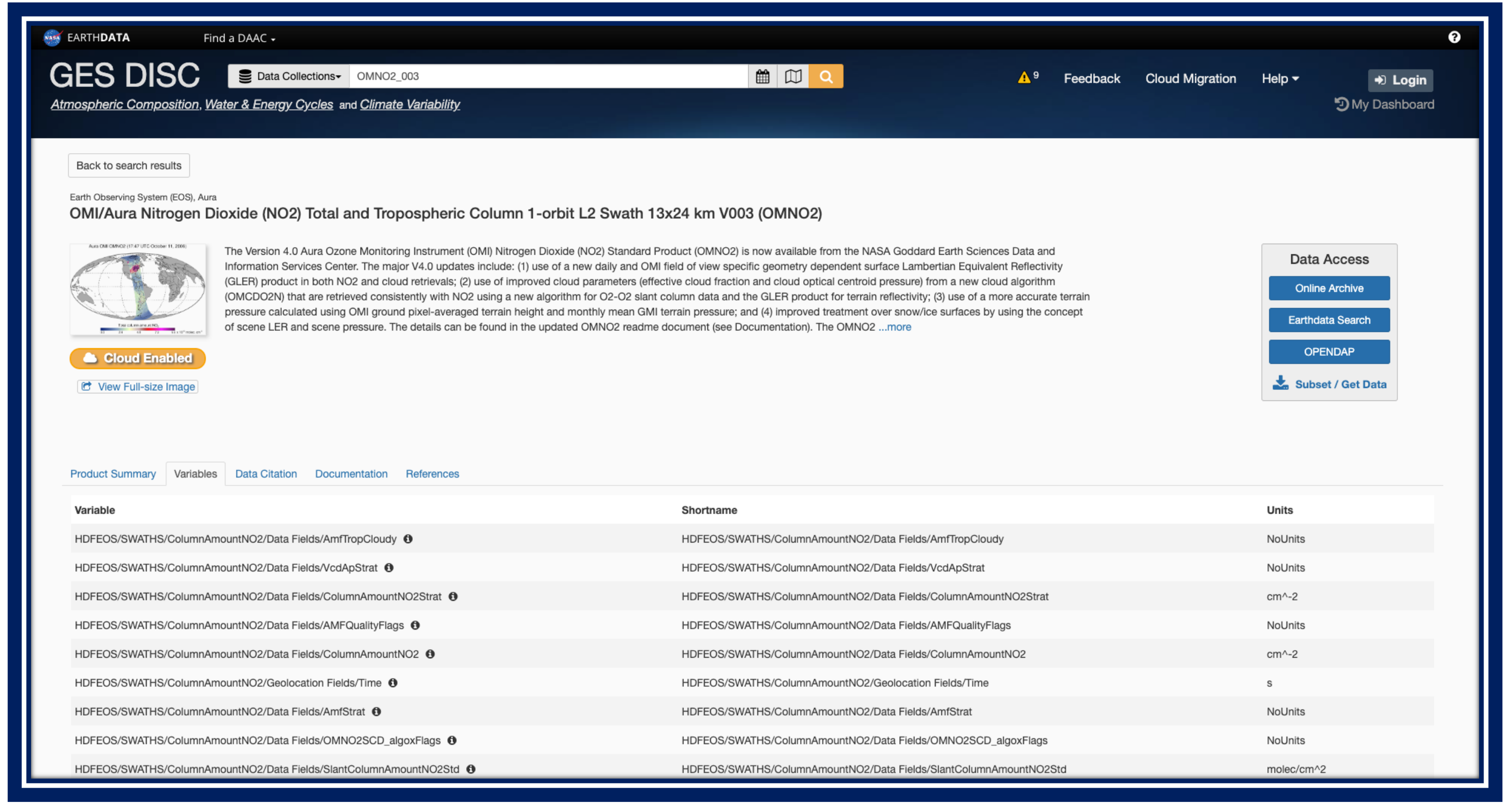
Figure 2. A global map of the Earth showing the locations of selected ground-based observation stations for OMI.

Acknowledgements

The authors thank the GES DISC and EOSDIS scientists and engineers for their support on the earthdata-varinfo tool. Special regards to Stephanie Stettz, Steven Hessenius, and Lucas Sterzinger of GES DISC for their help and advice with preparing this poster. Icon images obtained from Wikimedia Commons. Satellite data products are available at disc.gsfc.nasa.gov and search.earthdata.nasa.gov.

Find us and try it out on GitHub!

Data Discovery and Transformation



EARTHDATA Find a DAAC
GES DISC — Data Collections — OMNO2_003
Atmospheric Composition, Water & Energy Cycles and Climate Variability
Feedback | Cloud Migration | Help | Login | My Dashboard

Back to search results

Earth Observing System (EOS), Aura
OMI/Aura Nitrogen Dioxide (NO2) Total and Tropospheric Column 1-orbit L2 Swath 13x24 km V003 (OMNO2)

The Version 4.0 Aura Ozone Monitoring Instrument (OMI) Nitrogen Dioxide (NO2) Standard Product (OMNO2) is now available from the NASA Goddard Earth Sciences Data and Information Services Center. The major V4.0 updates include: (1) use of a new daily and OMI field of view specific geometry dependent surface Lambertian Equivalent Reflectivity (GLER) product in both NO2 and cloud retrievals; (2) use of improved cloud parameters (effective cloud fraction and cloud optical centroid pressure) from a new cloud algorithm (OMCDO2N) that are retrieved consistently with NO2 using a new algorithm for O2-O2 slant column data and the GLER product for terrain reflectivity; (3) use of a more accurate terrain pressure calculated using OMI ground pixel-averaged terrain height and monthly mean GMI terrain pressure; and (4) improved treatment over snow/ice surfaces by using the concept of scene LER and scene pressure. The details can be found in the updated OMNO2 readme document (see Documentation). The OMNO2 ...more

Cloud Enabled | View Full-size Image

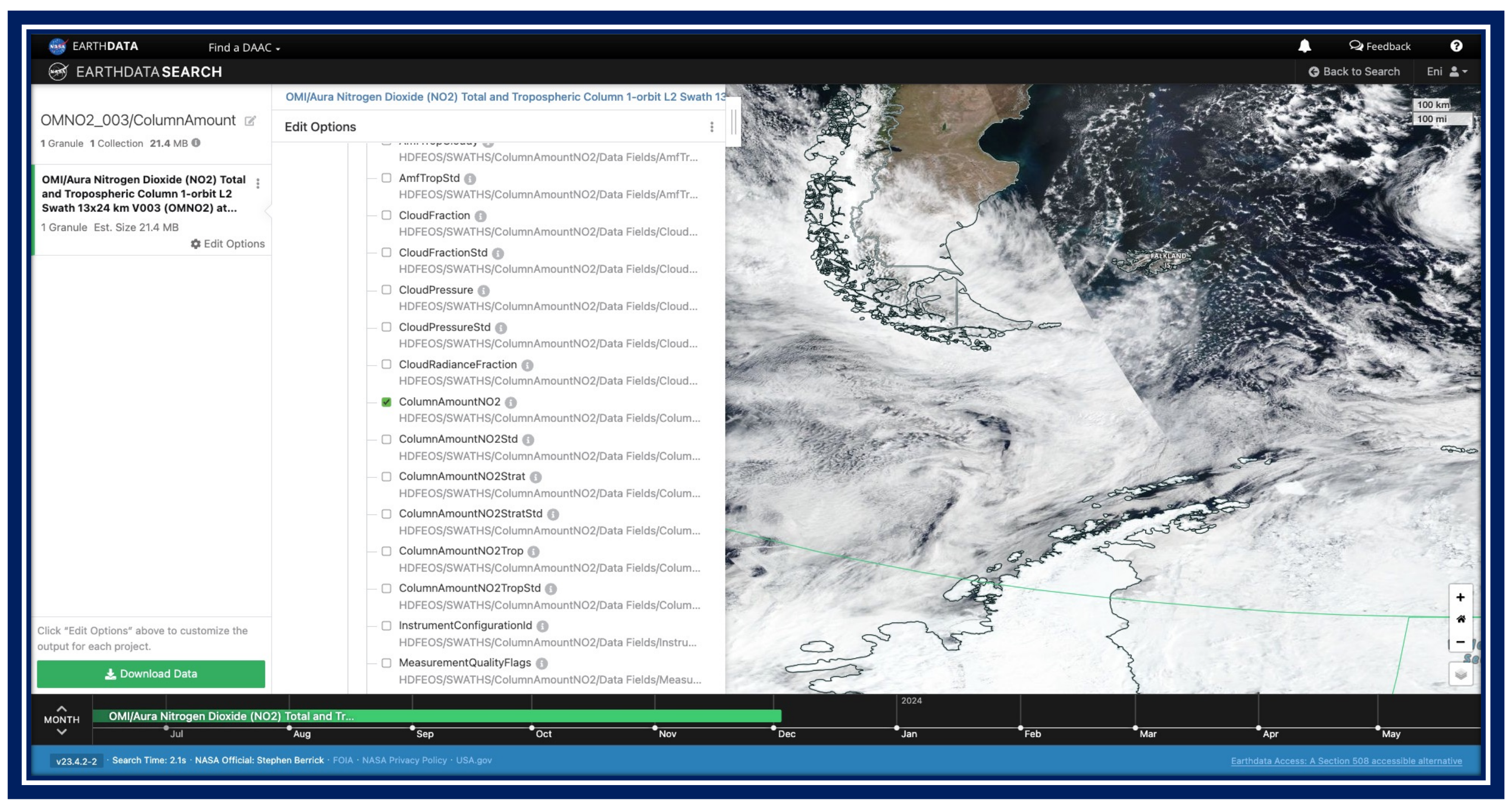
Data Access: Online Archive | Earthdata Search | OPENDAP | Subset / Get Data

Product Summary | Variables | Data Citation | Documentation | References

Variable	Shortname	Units
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfTropCloudy	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfTropCloudy	NoUnits
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/VcdApStrat	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/VcdApStrat	NoUnits
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ColumnAmountNO2Strat	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ColumnAmountNO2Strat	cm^-2
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AMFQualityFlags	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AMFQualityFlags	NoUnits
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ColumnAmountNO2	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/ColumnAmountNO2	cm^-2
HDFEOS/SWATHS/ColumnAmountNO2/Geolocation Fields/Time	HDFEOS/SWATHS/ColumnAmountNO2/Geolocation Fields/Time	s
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfStrat	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfStrat	NoUnits
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/OMNO2SCD_algoxFlags	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/OMNO2SCD_algoxFlags	NoUnits
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/SlantColumnAmountNO2Std	HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/SlantColumnAmountNO2Std	molec/cm^2

Figure 3. Dataset landing page for the *OMNO2_003* collection. Note the variables tab.

Once the UMM-Var records have been published, they are displayed on the GES DISC dataset landing pages under a "Variables" tab. Users can assess what variables are in the data before downloading it. Additionally, the short names are displayed for a more direct interpretation of the variables after downloading. In the future, this page will be expanded to provide more information for each variable.



EARTHDATA Find a DAAC | Feedback
EARTHDATA SEARCH | Back to Search | Eni

OMNO2_003/ColumnAmount
1 Granule 1 Collection 21.4 MB

OMI/Aura Nitrogen Dioxide (NO2) Total and Tropospheric Column 1-orbit L2 Swath 13x24 km V003 (OMNO2) at...
1 Granule Est. Size 21.4 MB
Edit Options

OMI/Aura Nitrogen Dioxide (NO2) Total and Tropospheric Column 1-orbit L2 Swath 13
Edit Options

HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfTr...
□ AmfTropStd
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/AmfTr...
□ CloudFraction
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Cloud...
□ CloudFractionStd
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Cloud...
□ CloudPressure
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Cloud...
□ CloudPressureStd
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Cloud...
□ CloudRadianceFraction
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Cloud...
☑ ColumnAmountNO2
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ ColumnAmountNO2Std
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ ColumnAmountNO2Strat
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ ColumnAmountNO2StratStd
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ ColumnAmountNO2Trop
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ ColumnAmountNO2TropStd
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Colum...
□ InstrumentConfigurationId
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Instru...
□ MeasurementQualityFlags
HDFEOS/SWATHS/ColumnAmountNO2/Data Fields/Measu...

Click "Edit Options" above to customize the output for each project.
Download Data

v23.4.2-2 · Search Time: 2.1s · NASA Official: Stephen Berrick · FOIA · NASA Privacy Policy · USA.gov
Earthdata Access: A Section 508 accessible alternative

Figure 4. A variable subsetting request for ColumnAmountNO2 in Earthdata Search with the Harmony Level 2 subsetter.

Once the UMM-Var records have been published, it is possible to perform a variable subset using Harmony services within Earthdata Search. By selecting only the variables that the user needs, the subsequent data downloaded is smaller, more manageable, and often yields a faster download. The Earthdata Harmony service can be accessed through the NASA Earthdata Search website. Look for the customization icons to determine which datasets have these services available.

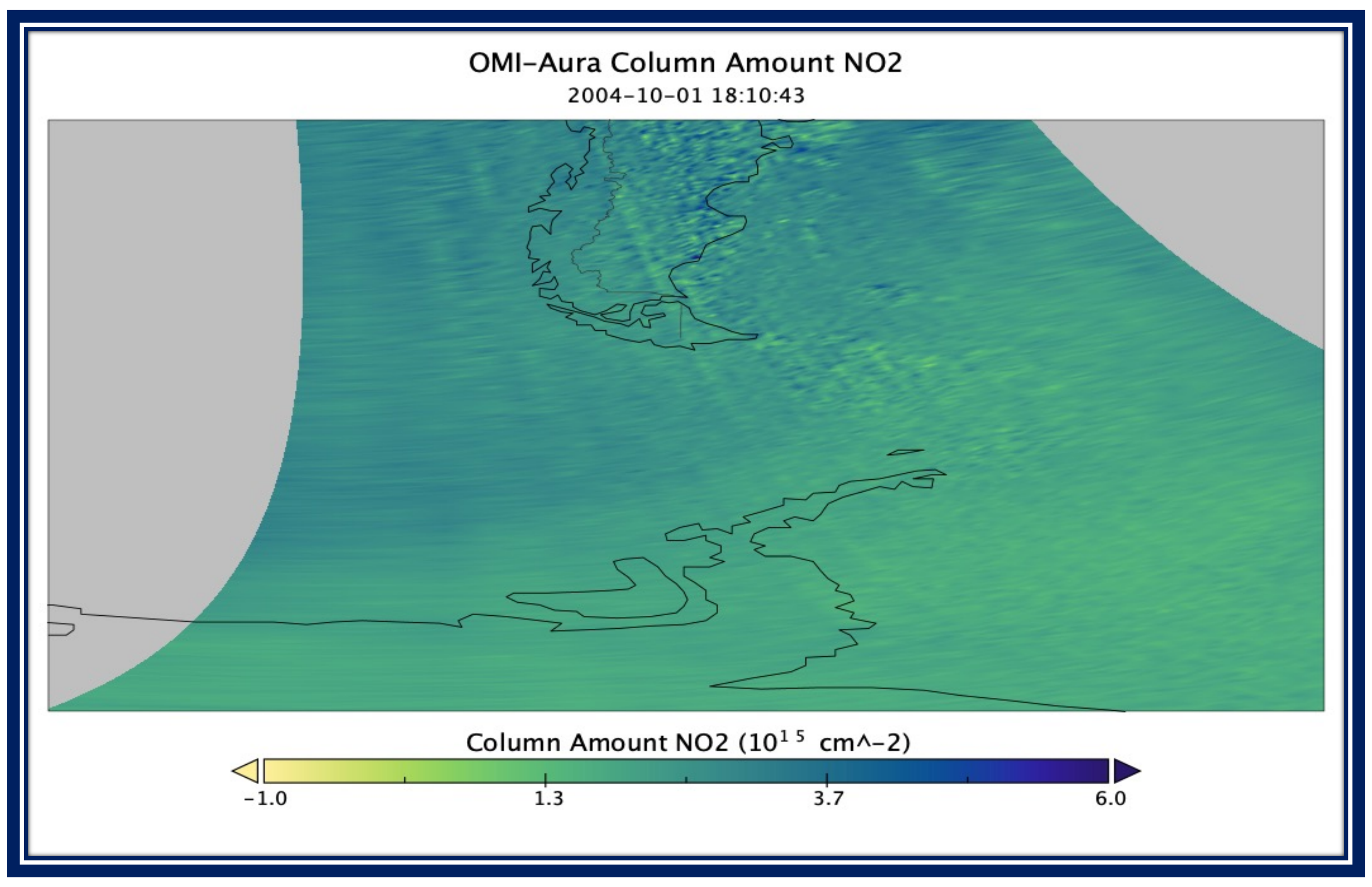

OMI–Aura Column Amount NO2
2004–10–01 18:10:43
Column Amount NO2 (10^{15} cm^–2)
–1.0
1.3
3.7
6.0

Figure 5. A spatial and variable subset for the variable ColumnAmountNO2 in the collection *OMNO2_003*.

Once the user selects the variable(s) and downloads the data, they can proceed with their data analysis as usual. This is one example of a Harmony subsetting output file, for the ColumnAmountNO2 variable from the collection *OMNO2_003*. This variable was also spatially subsetted to focus on southern Patagonia and the Antarctic peninsula region. The original file was ~ 20MB, and the output file after the variable and spatial subsetting was ~ 0.15MB.

What's next?

To date, we have published 7000+ variables from ~58 GES DISC collections. We will continue to use earthdata-varinfo to publish more variables while enhancing and expanding its functionality and usability.

Future plans:

- Improvements for identifying the classification of variables into: science variables, dimension variables, quality variables, or ancillary variables, which is important for both human- and machine-readability.
- Mapping the variable to a GCMD science keyword to enable data discovery through the GES DISC website and Earthdata Search.
- Adding Launchpad token authentication for publishing to CMR as ingest via Earthdata Login (EDL) Bearer Tokens are being disabled.

AGU Annual Meeting December 2023