



EE Datasets Tour and Data Upload

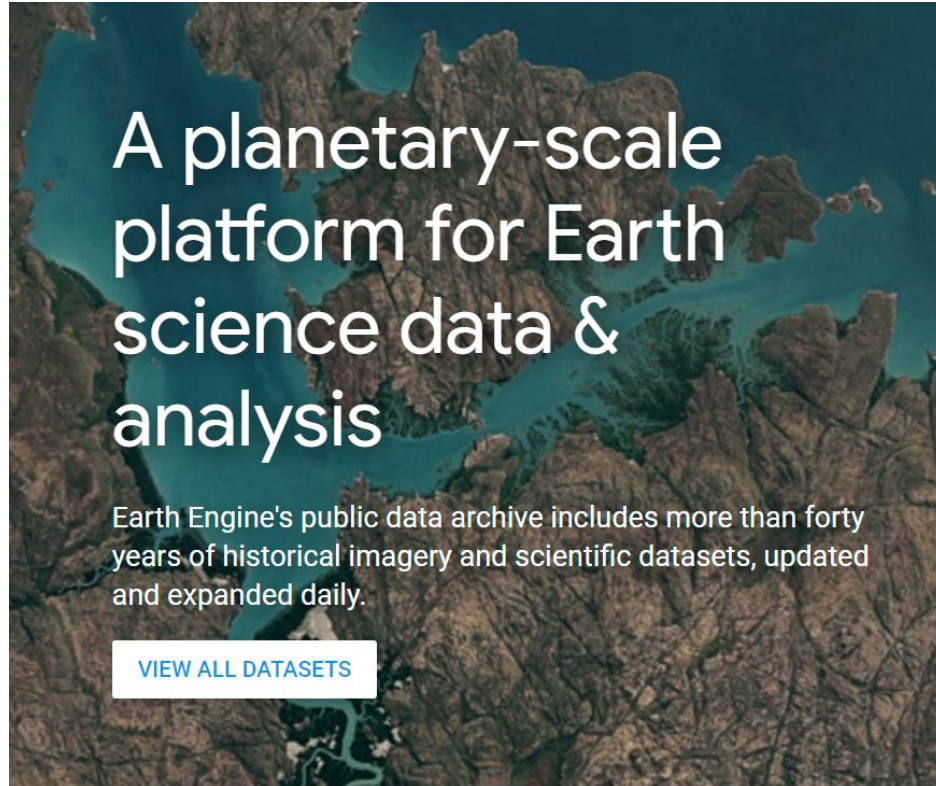
<http://bit.ly/eecatalog>

simonf@google.com / 2019-09-16

Agenda

- Brief tour
- New since June 2018
- Frequently asked questions
 - When will X be added?
 - Why is Y behind?
- Data distribution standards?
- User data upload
 - Manifests

Largest public repository of ARD (analysis-ready data) in the world?



29 petabytes total

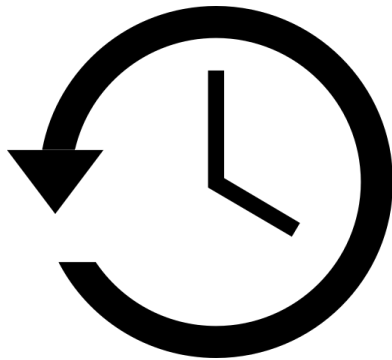
~1 Petabyte/month added
(50% of this is Sentinel-2)

>600 datasets

~100 datasets / year added

Highlights: Freshness

All datasets in the catalog are auto-updated daily (sometimes more often) by the Google Earth Engine team.



Highlights: Uniformity

All raster assets in the catalog are accessible in the same way.

All vector assets in the catalog are accessible in the same way.

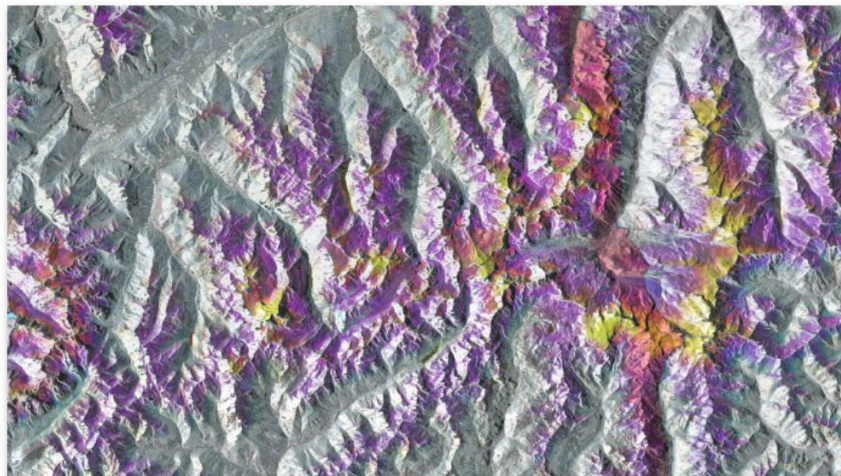


Highlights: catalog

Searchable [catalog](#) has detailed documentation for all datasets.



Dataset Highlights: Sentinel 1 GRD



Sentinel-1 SAR GRD: C-band Synthetic Aperture Radar

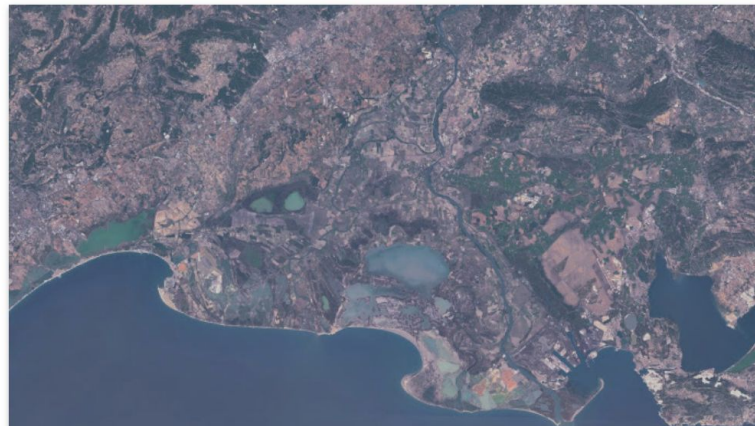
Data availability: 2014 – Present

The only public SAR (radar) dataset in the world. Useful for crop and flood monitoring.
5-day revisit cycle.

Dataset Highlights: Sentinel 2 L2 (SR) and L1 (TOA)



Surface Reflectance

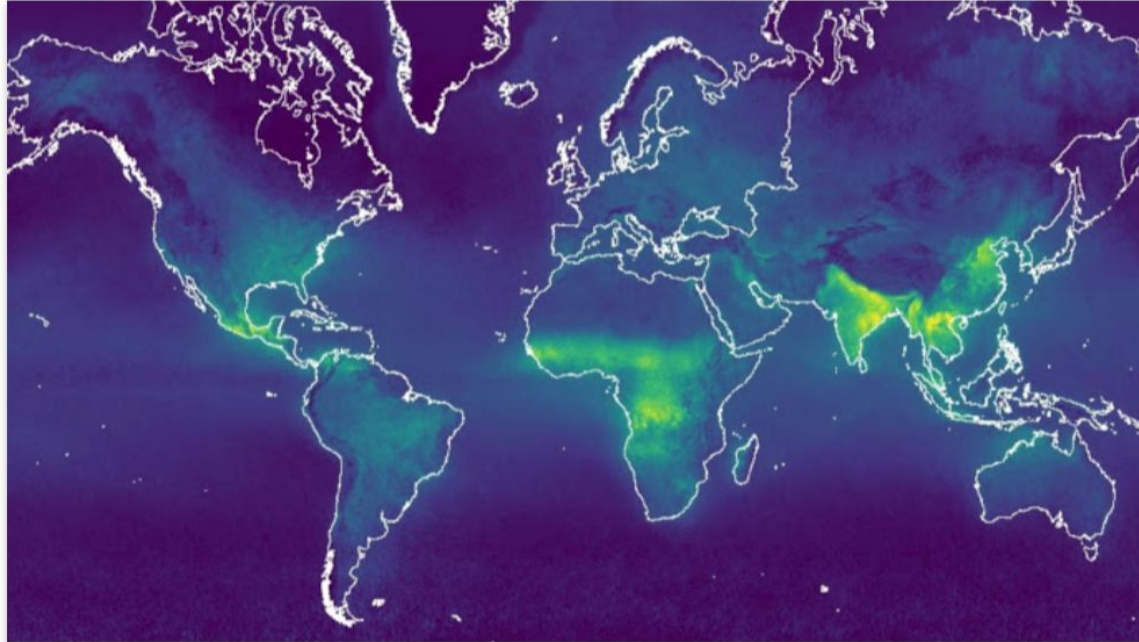


Top-of-Atmosphere Reflectance

Multiband optical imagery with up to 10m resolution (6-day revisit). Using SR products generated by ESA. [Mirrored to Cloud Public Datasets](#)

Dataset Highlights: Sentinel 5P (new!)

Gases: CO, CH₄, O₃, NO₂, SO₂, clouds, aerosoles,...



Sentinel-5P TROPOMI: TROPospheric Monitoring Instrument

Dataset Highlights: L4-L8 SR and TOA, L1-L5 MSS



Surface Reflectance

Atmospherically corrected surface reflectance from the Landsat 7 ETM+ sensor.

Dataset Availability: January 1999 - Present

TIER 1

TIER 2



Top of Atmosphere

Landsat 7 ETM+ Collection 1 calibrated top-of-atmosphere (TOA) reflectance.

Dataset Availability: January 1999 - Present

TIER 1

T1 + REAL-TIME

TIER 2

Multiband optical imagery with up to 30m resolution (16-day revisit). Using SR products generated by Google from a USGS docker image. [Mirrored to Cloud Public Datasets](#)

#GeoForGood19

Dataset Highlights: 26 MODIS v006 datasets

Daily composites of low-resolution images, with many derived products.

Daily Products



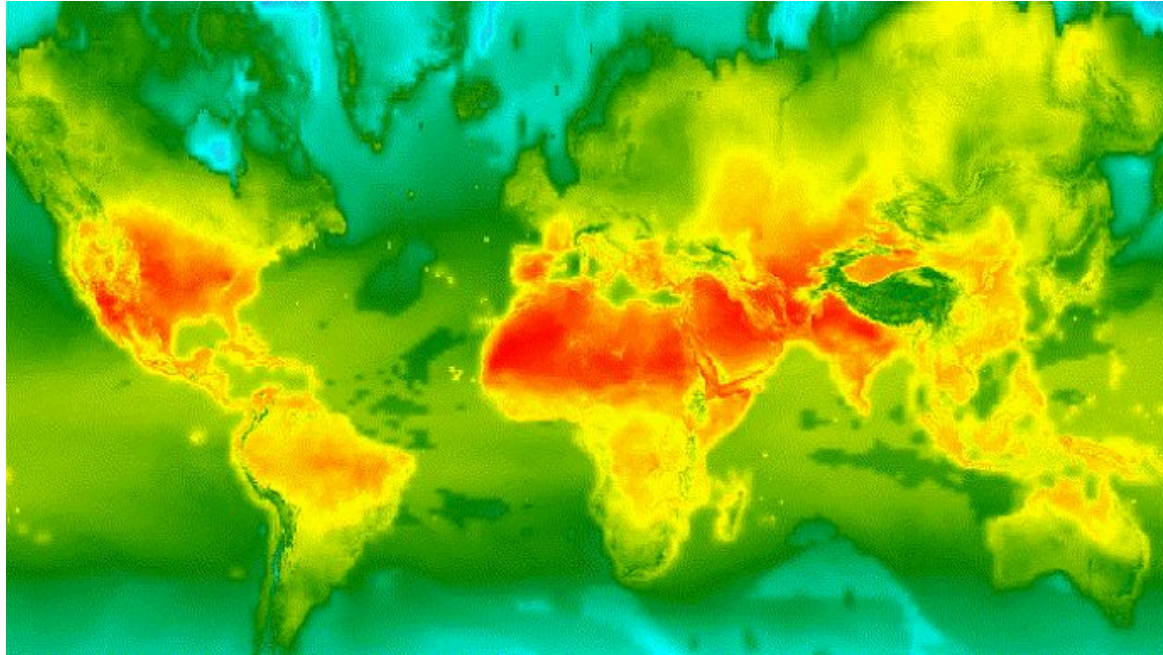
MCD43A4.006 MODIS Nadir BRDF-Adjusted Reflectance Daily 500m

The MCD43A4 V6 Nadir Bidirectional Reflectance Distribution Function Adjusted Reflectance Daily 500m. These are adjusted using a bidirectional reflectance distribution function to model the daily reflectance based on a 16-day retrieval period, with the image's date occurring on the 9th day. The image shows the best representative pixel from the 16-day period.

 [MODIS/006/MCD43A4](#)

Dataset Highlights: climate and weather datasets

Temperature, precipitation, drought, and many more layers



New since June 2018

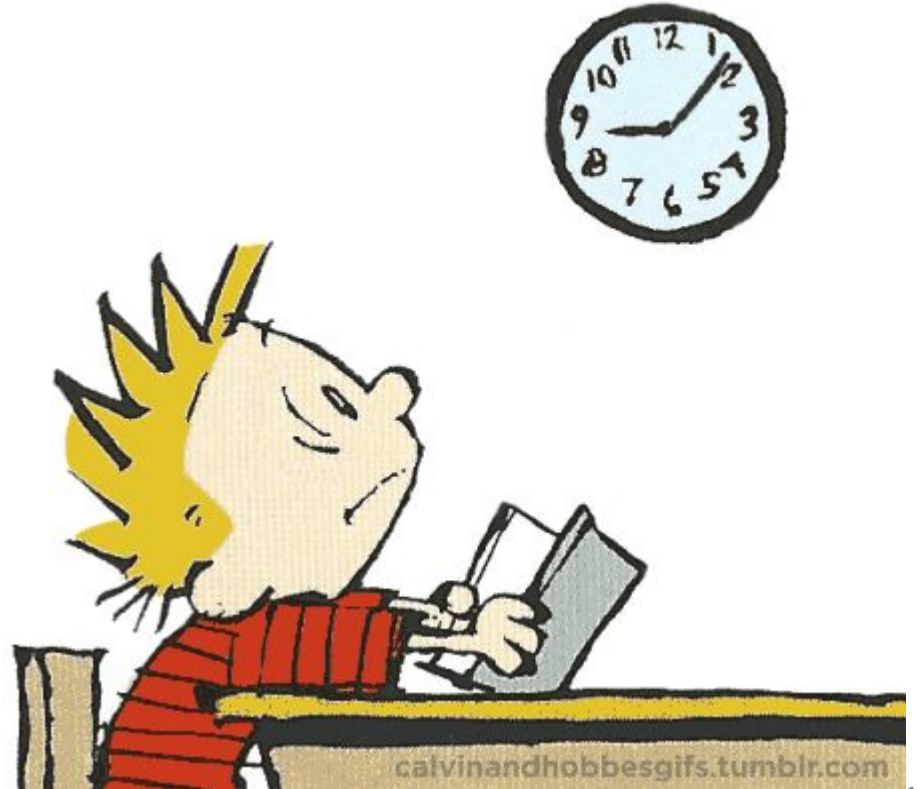
- [Sentinel 5P!](#) (Methane, aerosols...)
- [Sentinel 2 SR!](#) (public 10m imagery, corrected for atmospheric effects)
- [JAXA yearly SAR mosaics](#) (25m radar data from the PALSAR instruments)
- [Copernicus CORINE](#) (European 10m land cover)
- [Intertidal change](#) (historical classification of intertidal zones)
- [Arctic and Antarctic DEMs](#) (elevation for polar regions)
- [USGS GAP/Landfire ecosystems](#) (almost 600 land classes!)
- [OpenLandMap data](#) (soil/climate, user-uploaded)

Fun datasets you might not have seen

- [Land travel speed](#) (minutes / meter. Max is 87 min/m!)
- [50cm Netherlands DEM](#) (you can see individual ships)
- [CSP Global Human Modification](#) (fraction of each pixel altered by humans)
- [EO-1 Hyperion](#) (Hyperspectral imagery - 198 bands!)
- [Daily Fishing Hours](#) (Data from the [Global Fishing Watch](#) project)
- [MACA](#) and [NEX](#) (Temperature and precipitation projections to 2100)
- [GRACE monthly mass anomalies](#)

FAQ #1

When will dataset X be added???



FAQ 1.1

When will **GOES geostationary data** be added???

Answer 1.1

GOES 16 and 17 fire is almost ready to go.

FAQ 1.2

When will **Sentinel-1 SLC** be added???

Answer 1.2

Complex numbers are hard to average.

FAQ 1.3

When will **MODIS Level 2 (swath data)** be added???

Answer 1.3

Swath projection takes time to implement. Please give us more use cases for L2 data that L3 cannot satisfy.

FAQ 1.4

When will **CMIP** be added?

Answer 1.4

What subset of CMIP?

FAQ 1.5

When will **<a simple dataset>** be added???

You have [305 open dataset requests](#)!

Answer 1.5

Help us prioritize them! Star the request or add a comment to the bug.

FAQ 2

Why is dataset Y behind?

Answer 2

Many things can go wrong without warning:

- Provider site went down
- Provider changed catalog or data format
- Provider released a new dataset version
- Google infrastructure had an issue

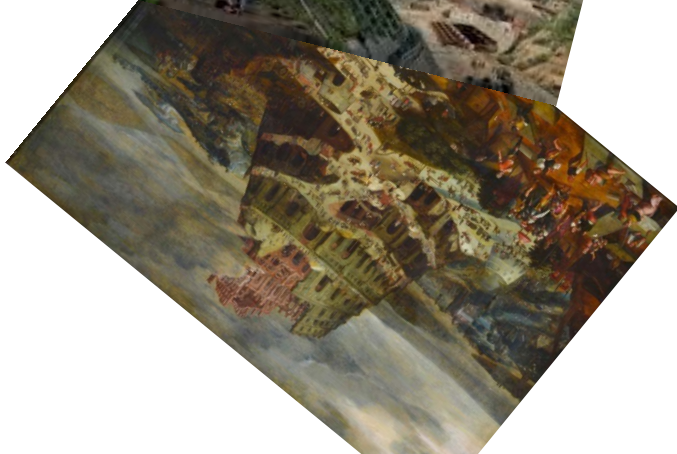
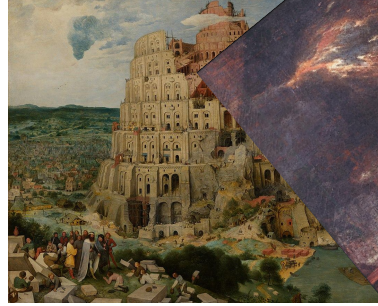
FAQ 2.1

Why do I have to ask on the developers list about why dataset Y is behind? Can I see a dashboard?

Answer 2.1



Answer 2.1



Answer 2.1



Serious Answer 2.1: Data Distribution Standards?

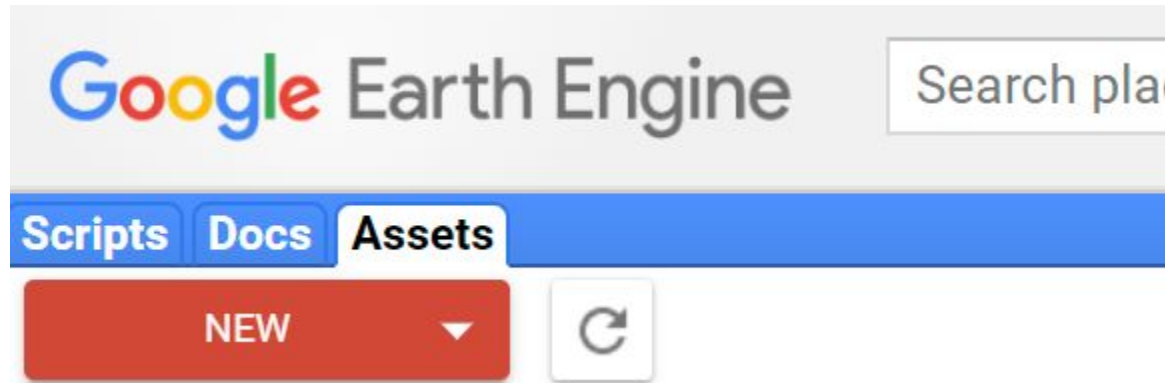
If data providers followed a standard for describing what files a dataset has, download pipelines could be simplified a lot, which would leave us more time to ingest new datasets.

I like [STAC \(SpatioTemporal Asset Catalogs\)](#). We export the EE catalog to <gs://earthengine-stac/catalog>

But really, any standards would help.

User data upload

Meanwhile, you can upload your own files using Code Editor or the [CLI \(Command-Line Interface\)](#).



Raster upload: geotiffs

Upload a new image asset

Source files [?]

SELECT

Please drag and drop or select files for this asset.

Asset ID [?]

users/simonf/ ▾

assetid

▼ Properties [?]

system:time_start

[+ Add property](#)

▼ Advanced

Pyramiding policy [?]

Mean ▾

Masking mode [?]

None ▾

[Learn more](#) [?] about how uploaded files are processed.

Vector upload: shapefiles or CSVs

Upload a new table asset

Source files [?]

SELECT

Please drag and drop or select files for this asset.

Asset ID [?]

users/simonf/ ▼

assetid

▼ Advanced

Character Encoding [?]

UTF-8



Maximum Error [?]

1.0

Maximum Vertices [?]

1000000

[Learn more](#) about how uploaded files are processed.

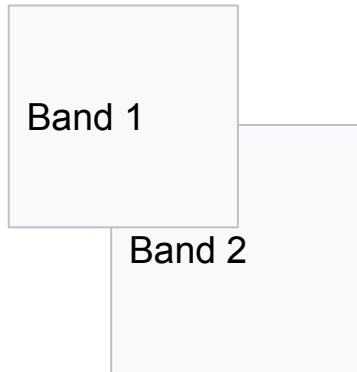
Manifest uploads

[See public documentation](#)

```
{
  "name": "projects/earthengine-legacy/assets/users/bucket/folder/asset",
  "tilesets": [
    {
      "sources": [
        {
          "uris": [
            "gs://earthengine-test/small.tif"
          ]
        }
      ]
    }
  ]
}
```

Manifest uploads - why?

- file1.tif
- file2.tif



Manifest uploads - why?

- file1.tif
- file2.tif



Survey results for top EE Data feature requests (most upvoted first)

- Faster updates to existing datasets.
- Adding temporal and spatial filters to dataset search.
- Better prioritization and triage of new datasets.
- Dashboard for ingestion pipelines
- Support for user-curated catalogs
- Pubsub notifications for new assets

Questions?

earthengine-data@google.com

