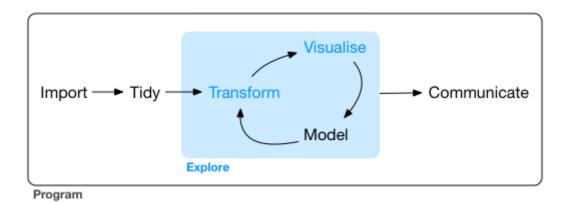
Cloud-native Geospatial

with Pangeo and the Microsoft Planetary Computer

Tom Augspurger May 2023



R for Data Science (Wickham & Grolemund)

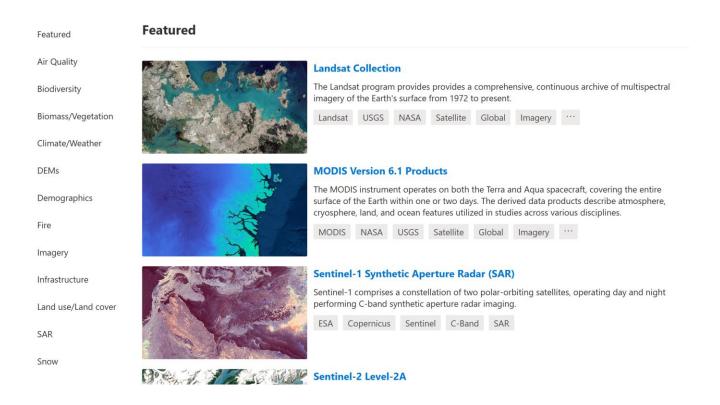
Components

Data, APIs, Compute

Data Catalog

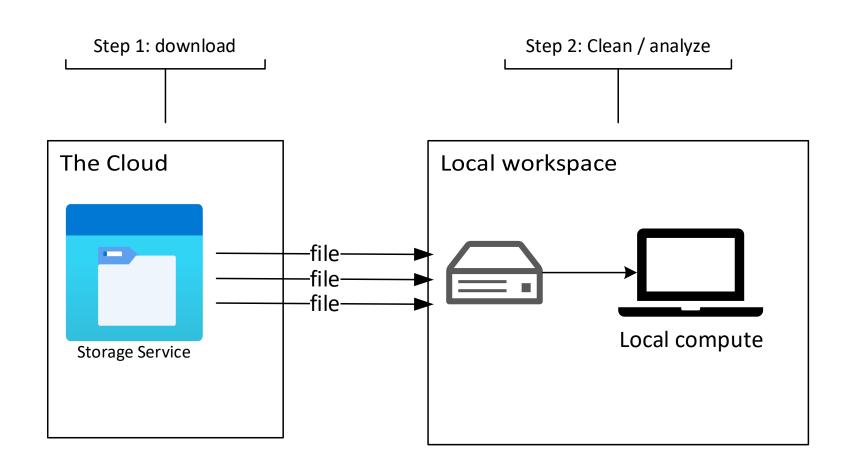
The Planetary Computer Data Catalog includes petabytes of environmental monitoring data, in consistent, analysis-ready formats. All of the datasets below can be accessed via Azure Blob Storage, and can be used by developers whether you're working within or outside of our Planetary Computer Hub.





Data Access

The "Download" Model



Azure Blob Storage Compute Jupyter -High bandwidth-Local client

The "Cloud-Native" Model

Files on Blob Storage isn't enough

- Find all the Sentinel-2 images over Wyoming in 2022
- Now do that for GOES-CMI, which has a completely different naming scheme

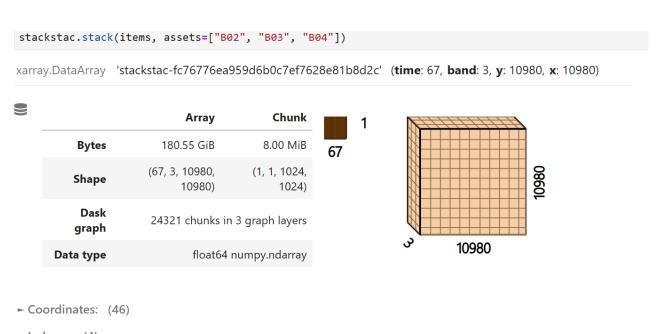


STAC for data access

 Find all the Sentinel-2 images over Wyoming in 2022

```
>>> import pystac_client
>>> catalog = pystac_client.Client.open(
        "https://planetarycomputer.microsoft.com/api/stac/v1/"
>>> items = catalog.search(
        collections="sentinel-2-l2a",
        intersects=aoi,
        datetime="2022",
        query={"eo:cloud_cover": {"lt": 10}}
```

STAC to Data Containers



► Indexes: (4)

▼ Attributes:

spec: RasterSpec(epsg=32613, bounds=(499980.0, 4490220.0, 609780.0, 4600020.0), resolutions_

xy=(10.0, 10.0))

crs: epsg:32613

transform: | 10.00, 0.00, 499980.00|

0.00,-10.00, 4600020.00

0.00, 0.00, 1.00

resolution: 10.0

Files + STAC isn't (always) enough

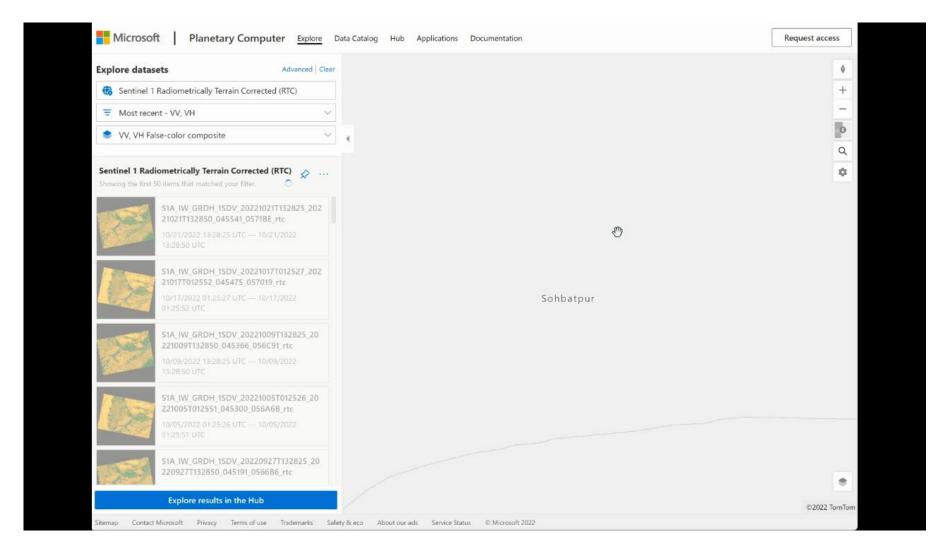
	Download	Cloud-native
Download	0.5	-
Metadata	0.02	1.5
Data	0.5	1.5
Total	1.02 seconds	3.0 seconds

Time to read a variable from a National Water Model file.

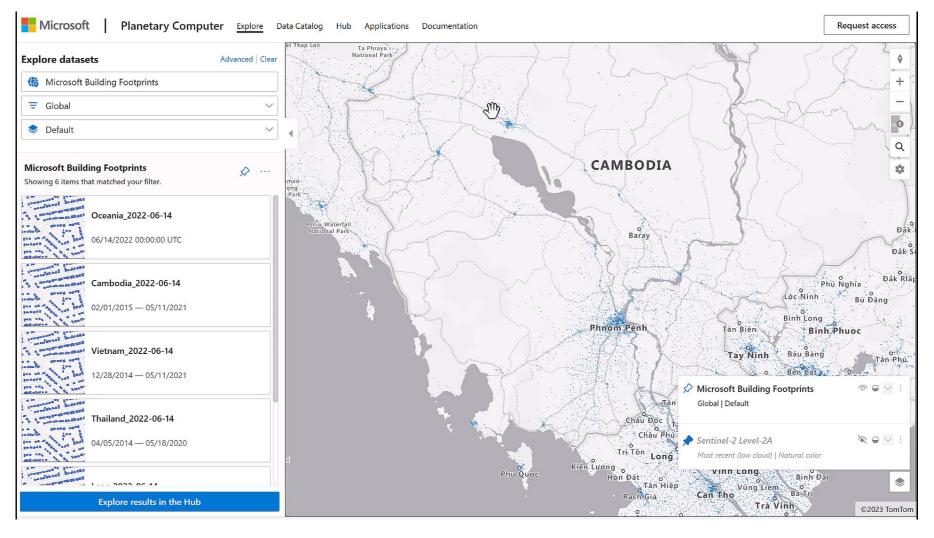
Cloud Optimized Assets (and clients)

- · We'll sometimes do data conversions
 - · Consolidated metadata
 - Efficient access subsets of the data
- Depends on the use-case

Explorer



Explorer



https://ai4edatasetspublicassets.azureedge.net/assets/pc_video/vector-tile-ms-buildings-feature.mp4

Planetary Computer Summary

- Data on Azure Blob Storage
- · STAC metadata for each dataset
- APIs for querying & visualizing the data
- · Cloud-Optimized assets (sometimes)

Thanks!

https://planetarycomputer.microsoft.com

taugspurger@microsoft.com