PO.DAAC Migrates to the Cloud and the River

Cassandra Nickles[1], Catalina Oaida[1], Alireza Farahmand[1], Suresh Vannan[2], Jinbo Wang[1], Mike Gangl[1], Frank Greguska[1], and Cedric David[3]

[1]NASA Jet Propulsion Laboratory
[2]Jet Propulsion Laboratory, California Institute of Technology
[3]Jet Propulsion Laboratory

November 22, 2022

Abstract

The Physical Oceanography Distributed Active Archive Center (PO.DAAC) has traditionally hosted NASA's Earth Observing System oceanography datasets, but is expanding its archive to include hydrology datasets from satellites like the upcoming Surface Water and Ocean Topography (SWOT) mission. The SWOT mission, expected to launch later this year (2022), will deliver approximately 20 TB of data per day! Though hydrologic and water resources applications will be enabled at a greater scale than ever before, an increase in data volume requires more efficient and scalable data management technologies. Cloud computing tools and services can help pave the way toward efficiency. By June 2022, PO.DAAC will have enabled all its data to be accessed in the NASA Earthdata Cloud hosted in Amazon Web Services (AWS). Other NASA DAACs are also in the process of migrating their Earth observations to the Earthdata Cloud, which will support seamless access across DAACs and disciplines. PO.DAAC desires to make data access, pre-processing, and analysis as seamless as possible for data users, supporting science and applications users alike with relevant tools and resources. In this presentation, after introducing the PO.DAAC, we highlight a new SWOT-specific data search mechanism (searching via the SWOT River Database (SWORD) pre-defined river reaches) and showcase a cloud computing workflow in the context of hydrologic applications by accessing and analyzing a proxy SWOT dataset, Pre-SWOT Making Earth System Data Records for Use in Research Environments (MEaSUREs) river heights. This cloud workflow can be easily adapted to other PO.DAAC datasets, or further developed with other DAAC data, offering effective guidance and support for a variety of science use cases and applications.

Jet Propulsion Laboratory
California Institute of Technology

podaac
Physical Oceanography Distributed Active Archive Center

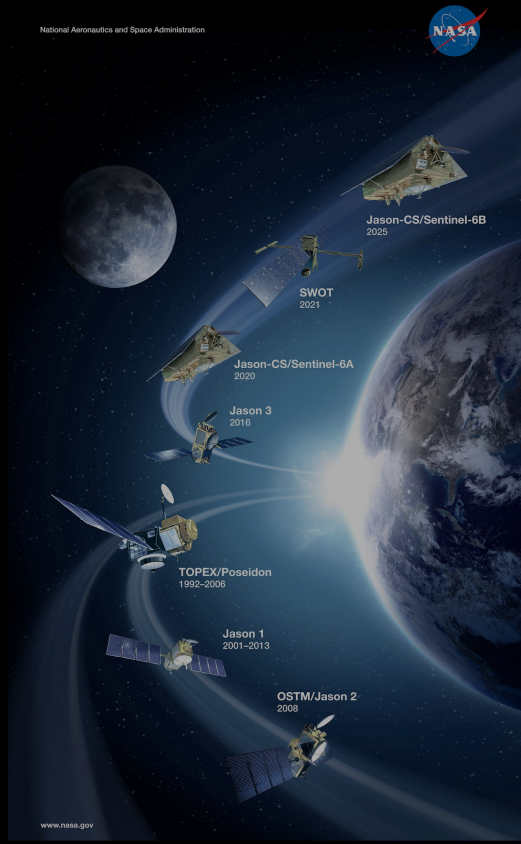
PO.DAAC Migrates to the Cloud and the River

Cassandra Nickles[1], Catalina Oaida[1], Alireza Farahmand[1], Suresh Vannan[1], Jinbo Wang[1], Michael Gangl[1], Frank Greguska[1], Cedric David[1]

1 Jet Propulsion Laboratory, California Institute of Technology

https://podaac.jpl.nasa.gov/

23 June 2022
Frontiers in Hydrology Meeting: The Potential of the SWOT Satellite Mission for Hydrologic Science



Physical Oceanography Distributed Active Archive Center (PO.DAAC)

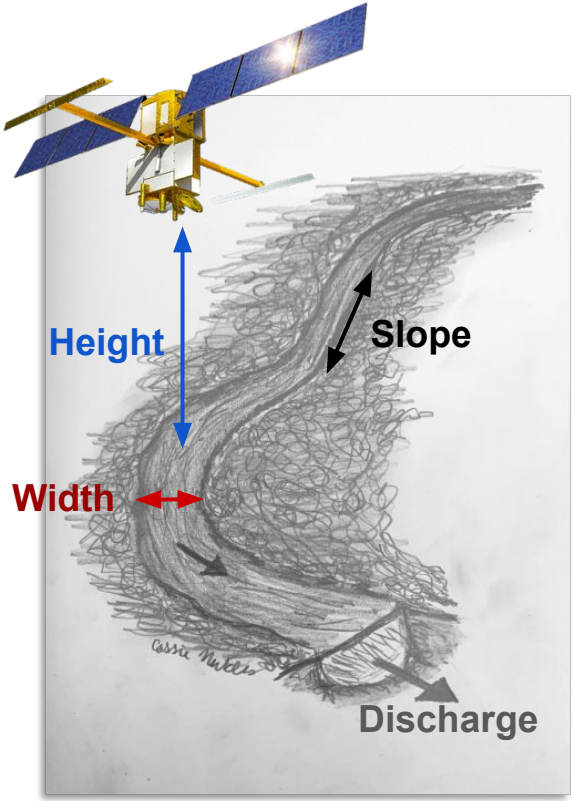
https://podaac.jpl.nasa.gov/

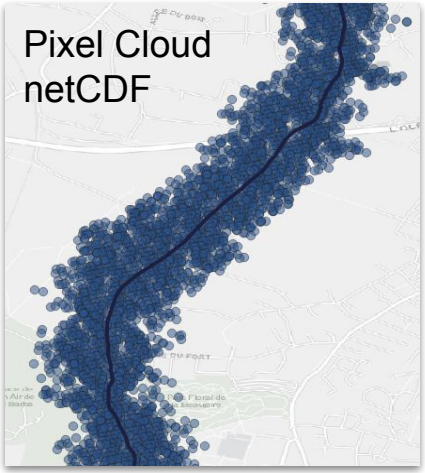
NASA data archive hosting:

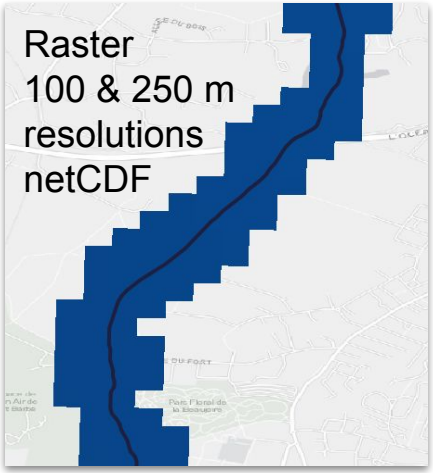
- Ocean Datasets including
  - SMAP
  - JASON Series
  - MODIS: AQUA/TERRA
  - GRACE/GRACE-FO
  - ECCO
  - Sentinel-6 Michael Freilich
  - SWOT (Launching Nov. 2022)
  - etc.
- Terrestrial Hydrosphere Datasets
  - GRACE/GRACE-FO
  - SWOT (Launching Nov. 2022)

PO.DAAC SWOT Website: https://podaac.jpl.nasa.gov/SWOT

SWOT Hydrology Measurements & Data Products







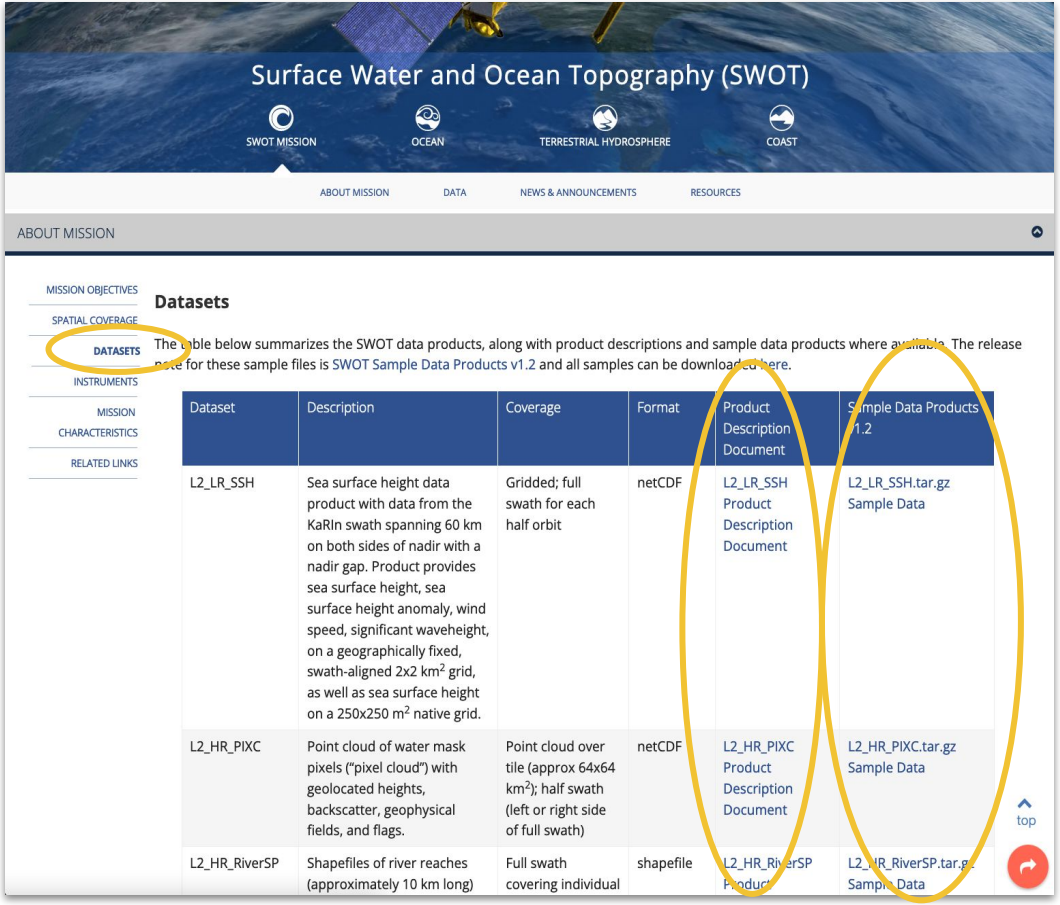
SWOT Data Product Descriptions & Sample Data

Product Description Documents

Sample files

- For getting accustomed with the format only
- 1-2 sample files for most of the products available

Will be getting some more comprehensive sample files in the cloud of products for an example 21-day cycle over the US soon!



PO.DAAC Tool: On-demand Raster Generation

- *Under Development*

- SWOT Standard Data Products (SDP) are limited to 100m and 250m resolutions
- PO.DAAC will offer On-demand Data Product (ODP) raster generation allowing projection, resolution, and overlap customization
- PO.DAAC to integrate output products with existing value added services (e.g. reformatting to cloud optimized geotiff)

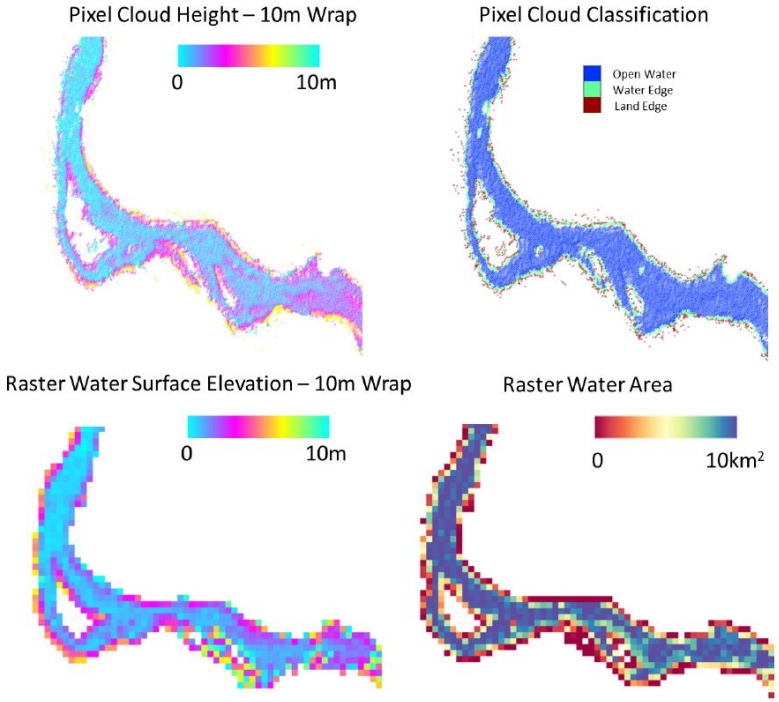


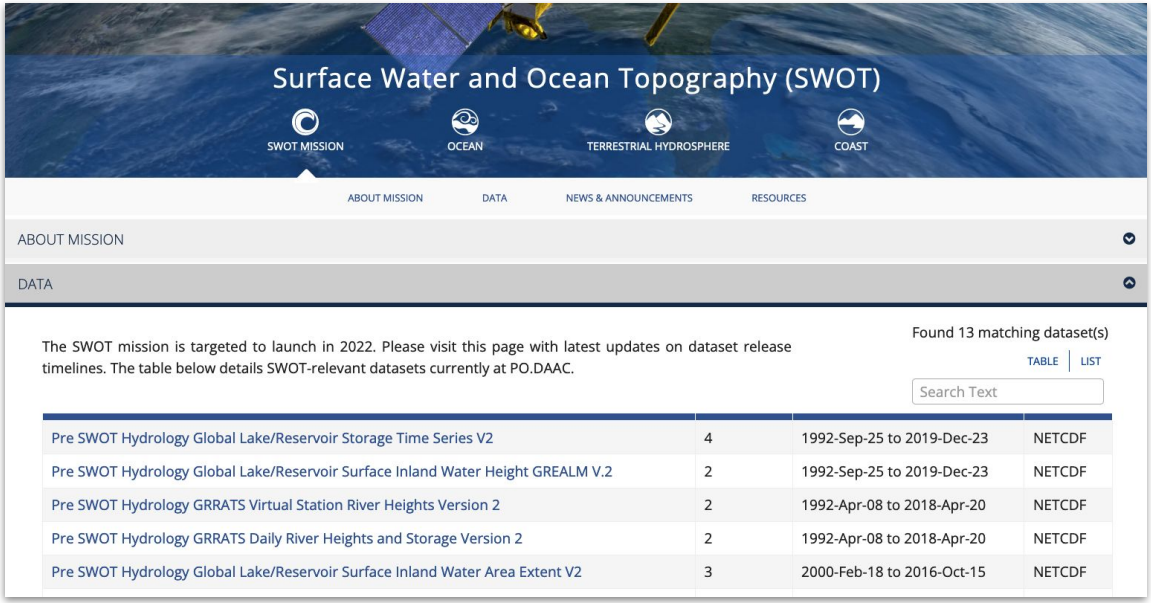
Figure 1. Example pixel cloud height (top left), and classification (top right) from which a raster product is produced, and the resulting raster water surface elevation (bottom left) and water area (bottom right). Note that this image shows layers from a 100 m raster in the UTM projection. Via D-56416_SWOT_Product_Description_L2_HR_Raster_20201105.pdf

Current SWOT-Relevant Data

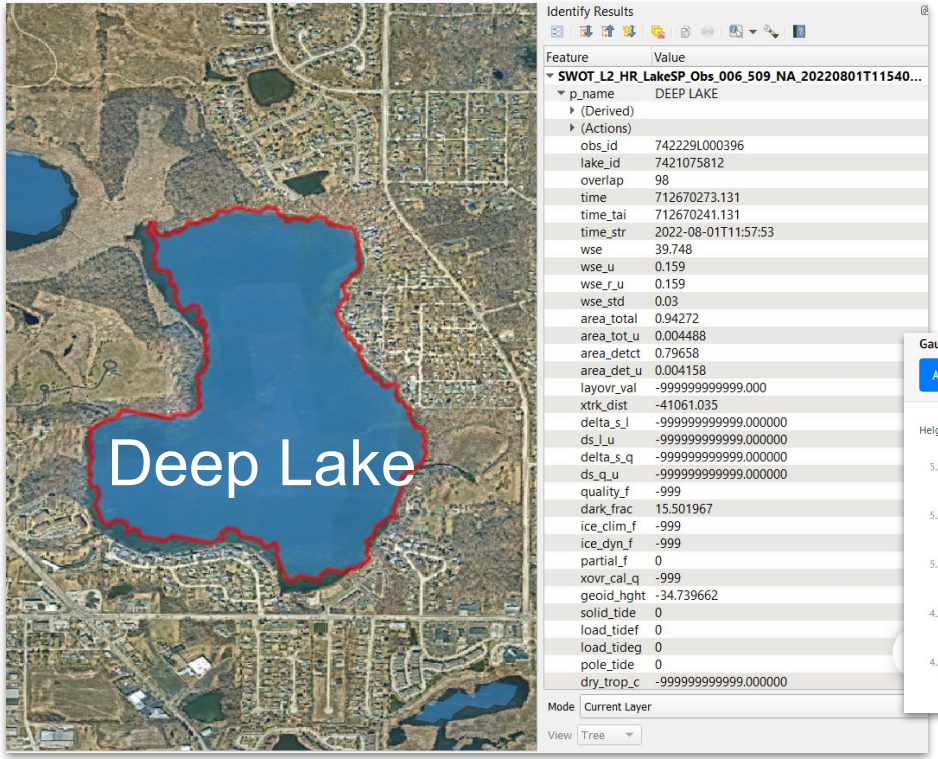
Hydrology Focus

- MEaSUREs Pre-SWOT lake extent, virtual station river height
- GRACE/GRACE-FO Water Equivalent Thickness Surface Mass Anomaly
- CYGNSS soil moisture
- LOCSS - lake observations by citizen scientists & satellites - *coming to PO.DAAC this summer*

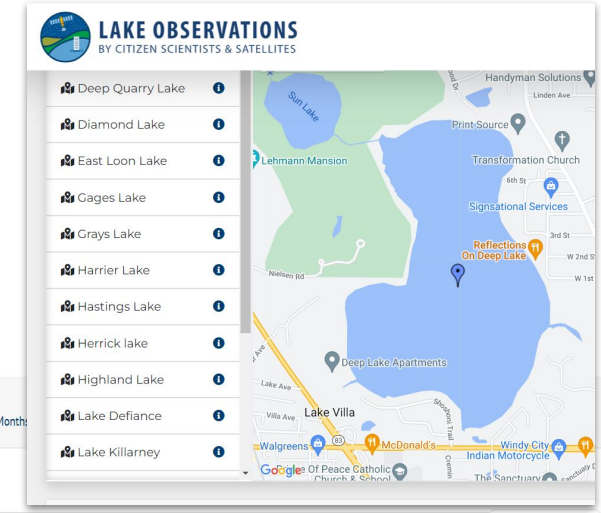
LAKE OBSERVATIONS
BY CITIZEN SCIENTISTS & SATELLITES

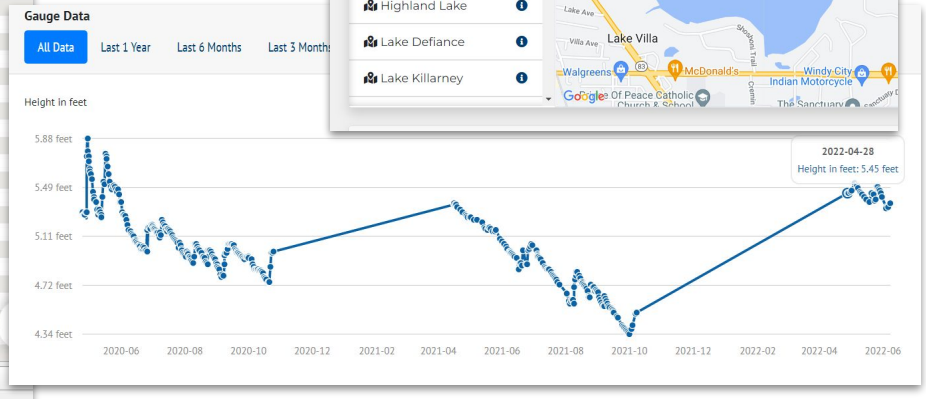


Example Future Use Case



Compare to LOCSS Data!





https://www.locss.org/

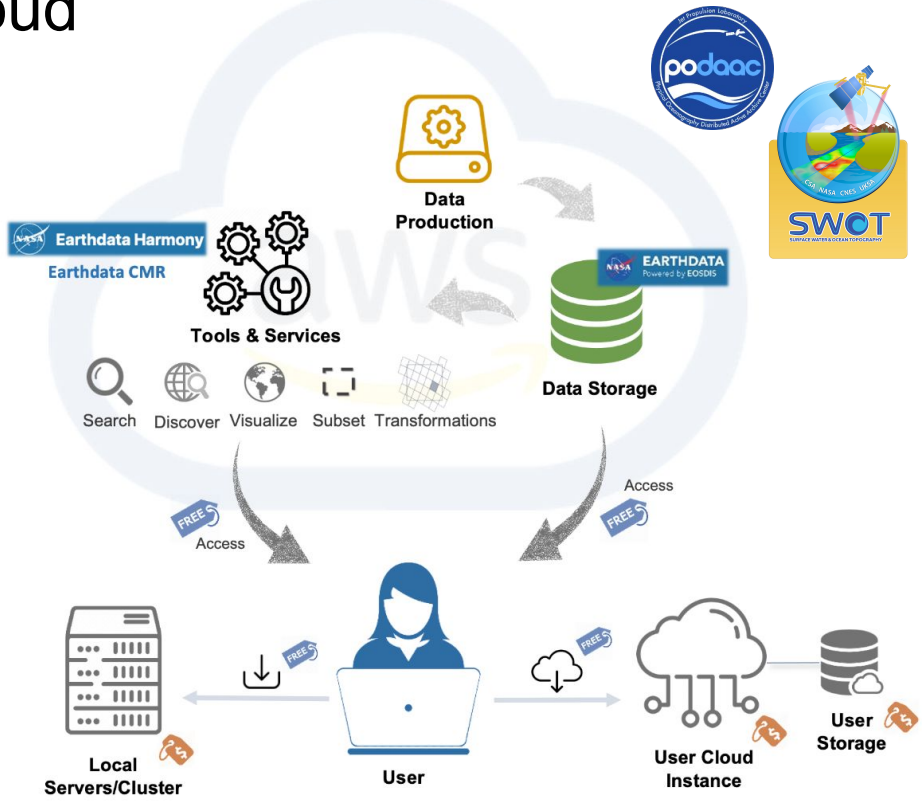
PO.DAAC Migrates to the Cloud

NASA Earthdata Cloud
Hosted in Amazon Web Services (AWS)

- Can **download data to local machines** or perform analysis **in the cloud**
- NASA data **access** and **download** is and will continue to be **free**, per NASA's open data policy.

Cloud offers the opportunity:

- to manage large data volumes
- to co-locate synergistic datasets
- for user analysis next to the data
- Leverage AWS cloud parallel computing



Cloud Data **Tools & Services**

Services that enable science and applications from and within the cloud, for oceanography, coastal and hydrology applications

Search and Access:

- **Geospatial search** in Earthdata Search by
  - River gauge location (point)
  - HUCs (polygons)
  - User-defined shapefile (polygons)
  - SWORD IDs (SWOT river IDs)
- **PO.DAAC Data Subscriber and Downloader**
  - https://github.com/podaac/data-subscriber
- **API based access** to data - Common Metadata Repository (CMR)

Subsetting data:

- Subset SWOT data by: time, space, variables, shapefiles (*Under development - Expect Aug./Sept. 2022*)
- APIs to subset, merge subsetted datasets, and transform data

SWORD River Reach Database

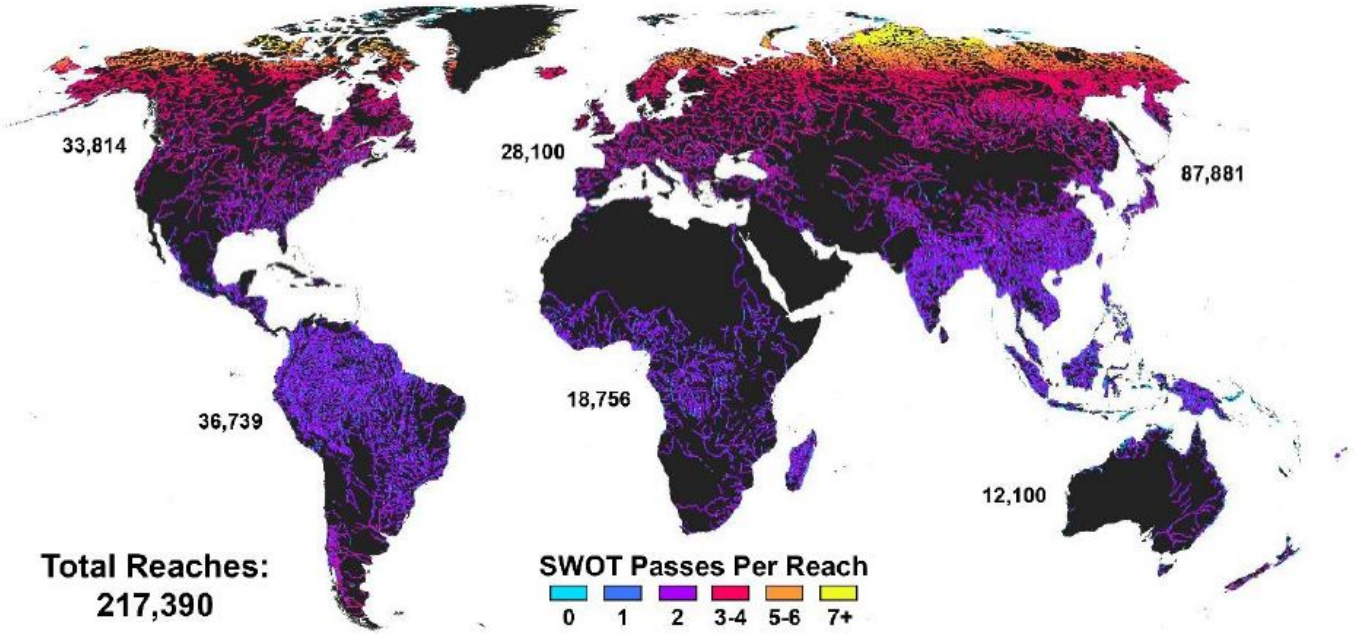
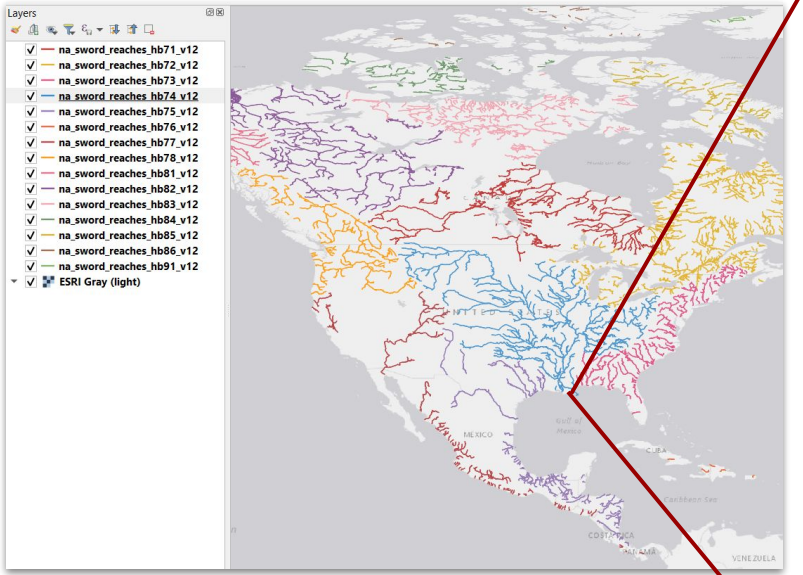
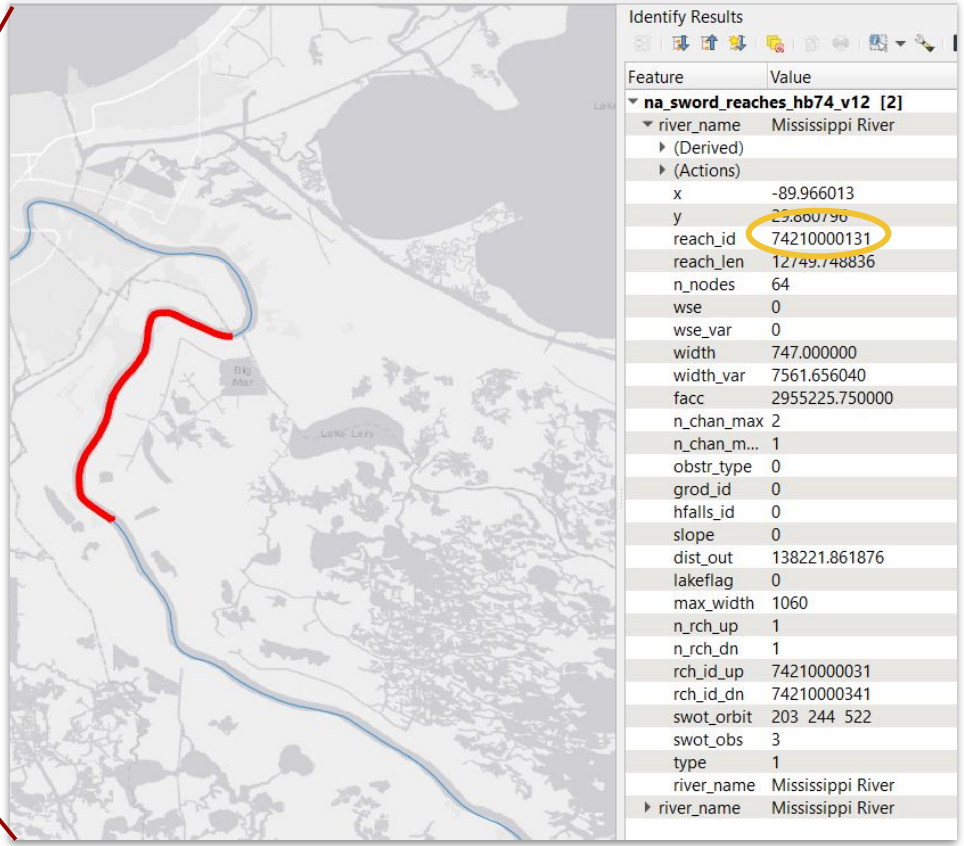


***Figure 1**: SWORD reach numbers per continent (not including ghost reaches). Colors display the number of SWOT passes per reach during the 21-day orbit cycle.*

Altenau et al. 2021

SWORD River Reaches - finding IDs

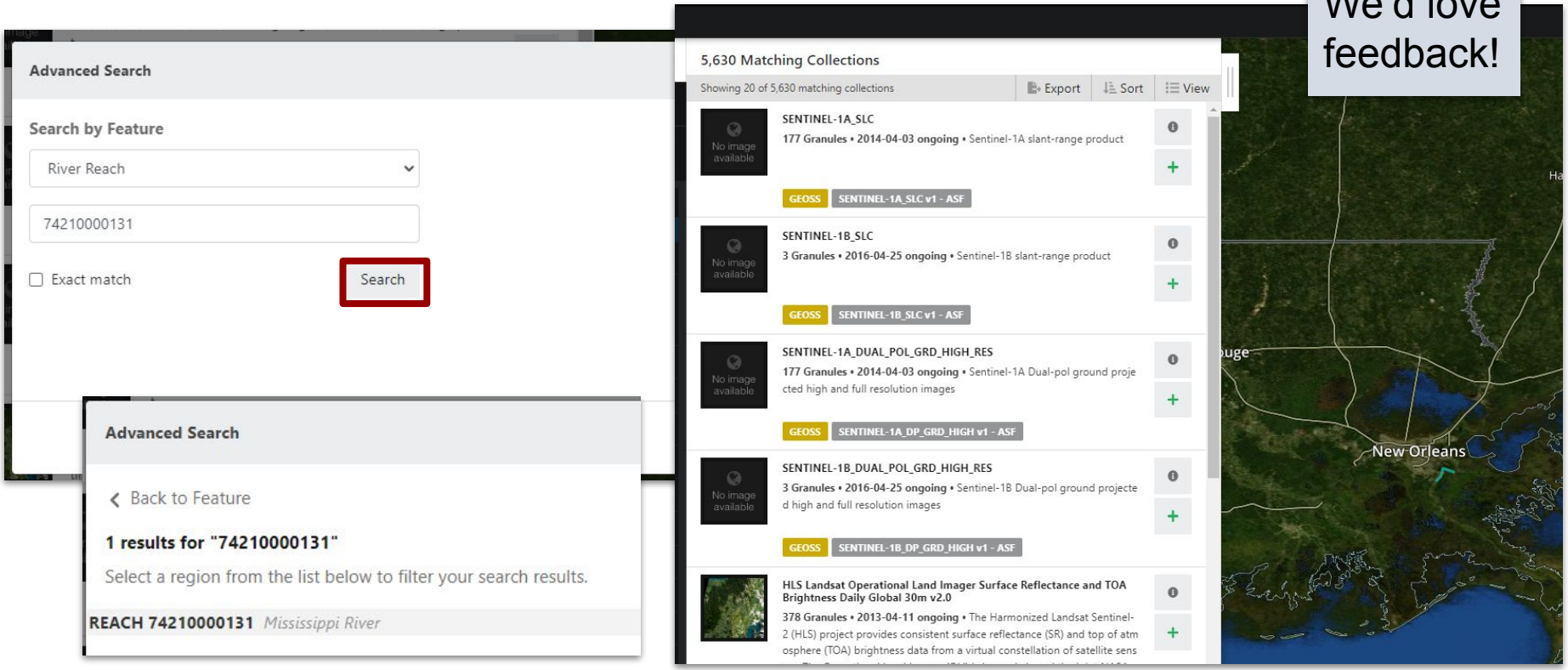




Shape files from
https://zenodo.org/record/5643392#.YpfUJ6jMJPY

Earthdata Search by SWORD River Reach ID

Try it out! We'd love feedback!



Programmatic Search via SWORD IDs

Plot a Single River Reach

In this section, we query the Feature Translation Service (FTS) SWORD service using a single Reach ID (from SWORD). In this example, we use the river Reach ID `13227000061`. This ID represents a specific reach along the Kasai River, a tributary of the Congo River in Africa.

```
[5]: response = requests.get("https://fts.podaac.earthdata.nasa.gov/rivers/reach/13227000061")
featureCollection = response_to_FeatureCollection(response)

pprint.pprint(response.json(), compact=True, width=60,
```

```
{'hits': 1,
 'results': {'13227000061': {...}},
 'search on': {'exact': False,
               'page_number': 1,
               'page_size': 100,
               'parameter': 'reach'},
 'status': '200 OK',
 'time': '8.768 ms.'}
```

Now we can plot this reach and calculate the center of the reac

```
[7]: fig.show()
```

https://github.com/podaac/tutorials/blob/master/notebooks/SWORD_River_Demo.ipynb

Can query datasets spatially programmatically via SWORD IDs (global)

Useful for regional studies over time accessing multiple files

Searching via HUC watersheds

HUC Feature Translation Service (FTS) Examples

This Jupyter Notebook contains examples related to geospatial search using the PO.DAAC *HUC Feature Translation Service (FTS)*, previewing (viusualizing) the queried region of interest, and using FTS results to query data through NASA's Common Metadata Repository (CMR).

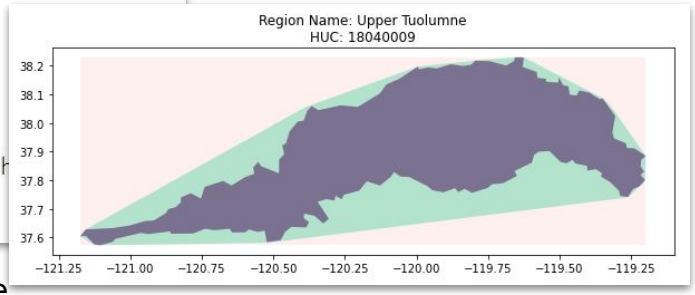
Example Use Case: Check if data is available over my region of interest using HUCs. In this example we are using the FTS-HUC API (https://fts.podaac.earthdata.nasa.gov/) to geospatially define our region of interest, namely the Upper Tuolumne River Basin in the San Joaquin River Basin in California's Sierra Nevada Mountains, searching by HUC or region name, and then using those geospatial bounds (coordinates) to query Sentinel-1 data in CMR.

1. use FTS to define geographic region of interest (query by partial or exact HUC or HUC region name)
2. preview query
3. use coordinates returned by FTS to query Sentinel-1 data in CMR, by polygon or bounding box.

Resources

USGS Hydrologic unit map to help identifiy region of interest (e.g. HUC value or name) can be found h
https://water.usgs.gov/GIS/regions.html

Can query names/partial names of basin or exact HUC ID matches (over U.S.)



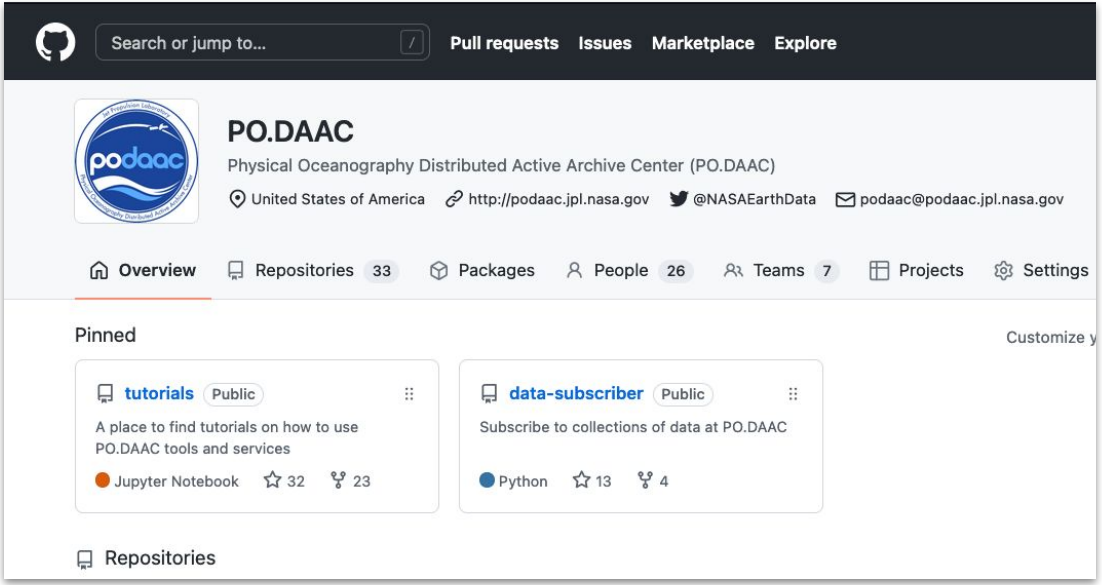
https://github.com/podaac/tutorials/blob/master/notebooks/HUC%20Feature%20Translation%20Service%20Examples-updated-20210804.ipynb

THANK YOU!

Check out our tutorials at https://github.com/podaac

Example Tutorial:

River Heights Jupyter Notebook
River_Heights_in_the_Cloud.ipynb



Contact us:
podaac@podaac.jpl.nasa.gov
cassandra.l.nickles@jpl.nasa.gov
catalina.oaida@jpl.nasa.gov

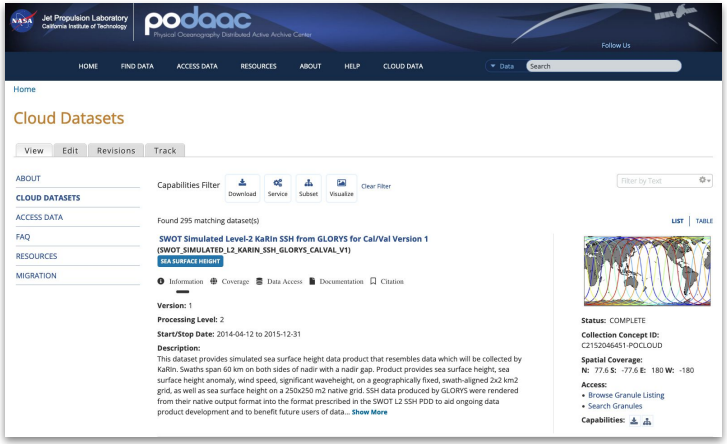
Reference

Altenau, E. H., Pavelsky, T. M., Durand, M. T., Yang, X., Frasson, R. P. d. M., & Bendezu, L. (2021). The surface water and ocean topography (SWOT) mission river database (SWORD): A global river network for satellite data products. *Water Resources Research*, 57, e2021WR030054. https://doi.org/10.1029/2021WR030054

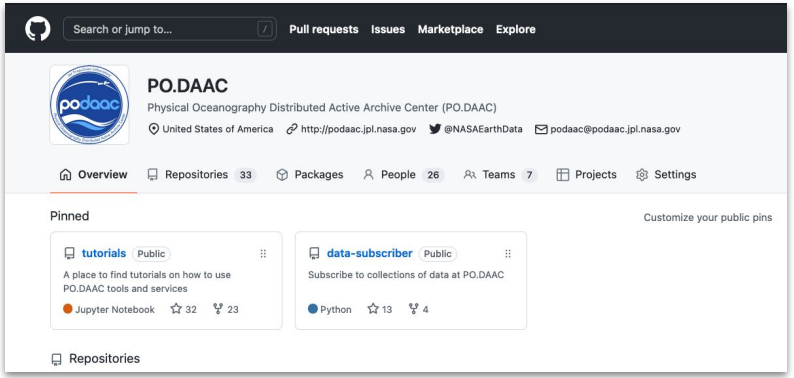
Extra Slides

Resources & User Community Support

- **One stop for PO.DAAC Cloud Information**: Cloud Data page with About, Cloud Datasets, Access Data, FAQs, Resources and Migration information
- Ask questions or find resources: PO.DAAC in the CLOUD Forum
- Cloud user migration overview, guidance, and resources: PO.DAAC Webinar
- Search and get access links: Earthdata Search Client and guide
- Search and get access links: PO.DAAC Cloud Earthdata Search Portal
- Browse cloud data in web-based browser: CMR Virtual Browse and guiding video
- Scripted data search end-point: Earthdata Common Metadata Repository (CMR) API
- Enable data download or access: Obtain Earthdata Login Account
- **Download data regularly**: PO.DAAC Data Subscriber Access video and PO.DAAC Data Subscriber instructions
- Bulk Download guide
- OPeNDAP in the cloud
- PO.DAAC scripts and notebooks: PO.DAAC Github
- How to get started in the AWS cloud (e.g. set up an instance): Earthdata Cloud Primer
- How to set up your own Jupyter Hub, Jupyter Lab, or Jupyter Notebooks in AWS cloud.
- Basic How-To tutorials for searching for cloud data and accessing data in the cloud (AGU workshop 2021):
  - Search and get access links from Earthdata Search
  - Earthdata login Authentication (scripted)
  - Direct data access in the cloud (without download)

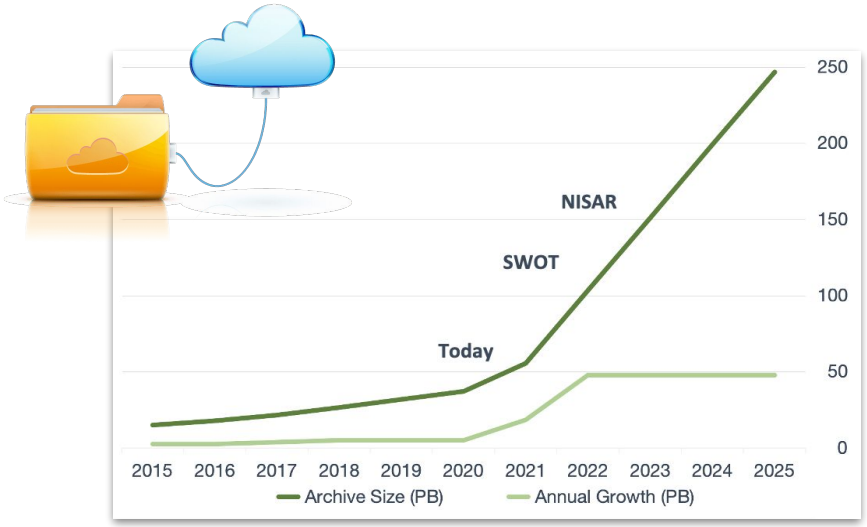


https://podaac.jpl.nasa.gov/cloud-datasets



https://github.com/podaac

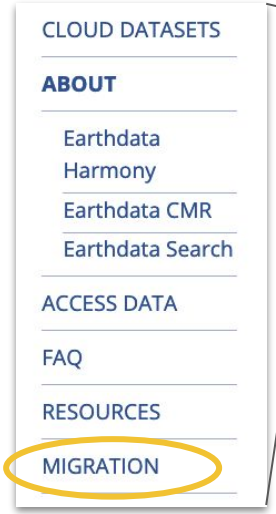
Physical Oceanography Distributed Active Archive Center (PO.DAAC) https://podaac.jpl.nasa.gov/

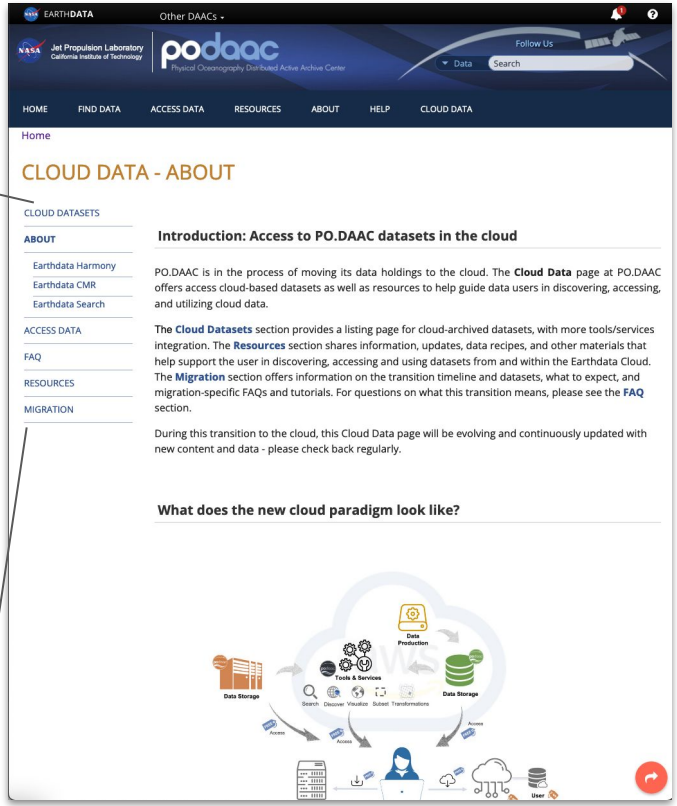


All (678) datasets (~30 PB) archived at PO.DAAC will be in the cloud by June 2022.

Learn more at

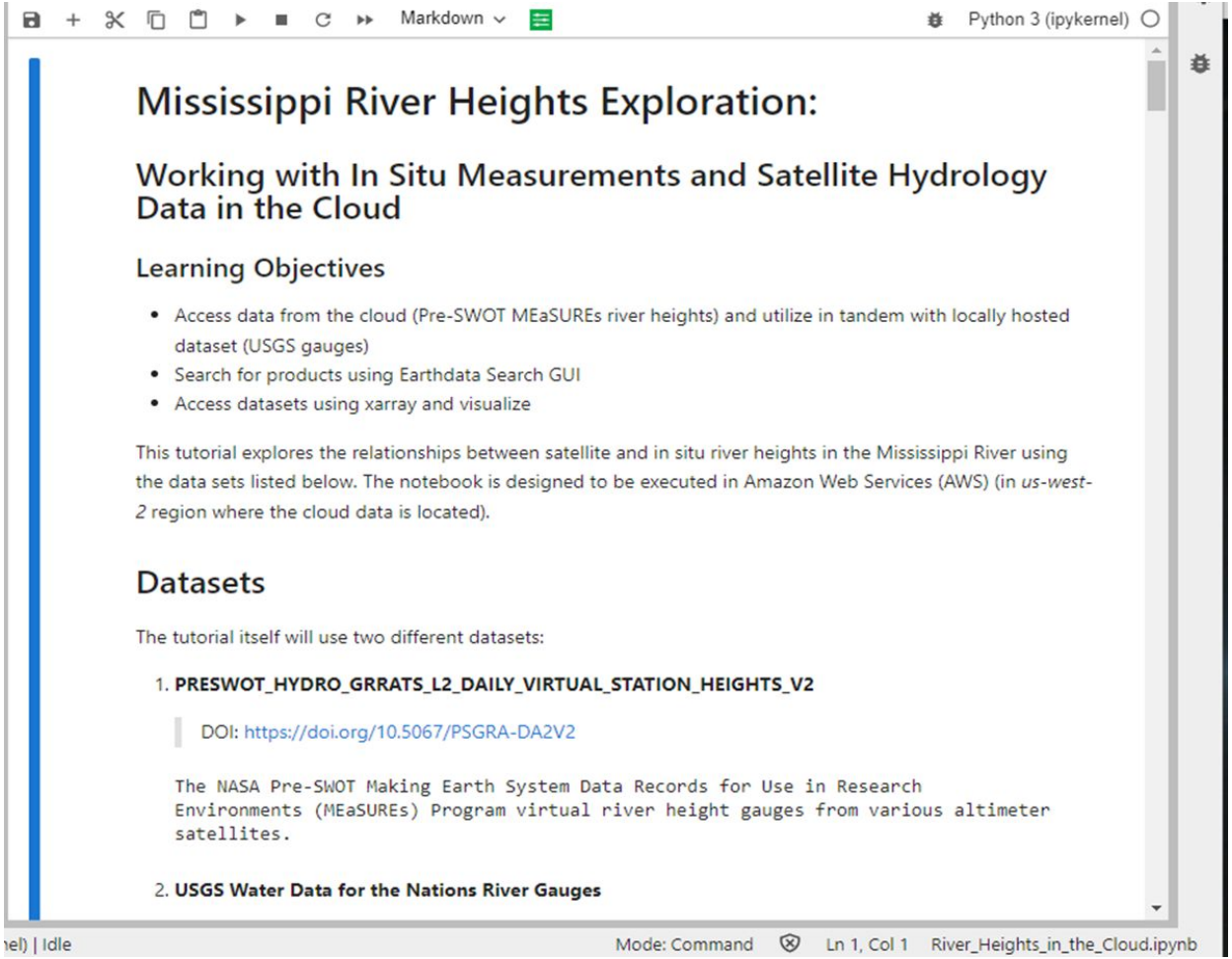
https://podaac.jpl.nasa.gov/cloud-datasets/migration



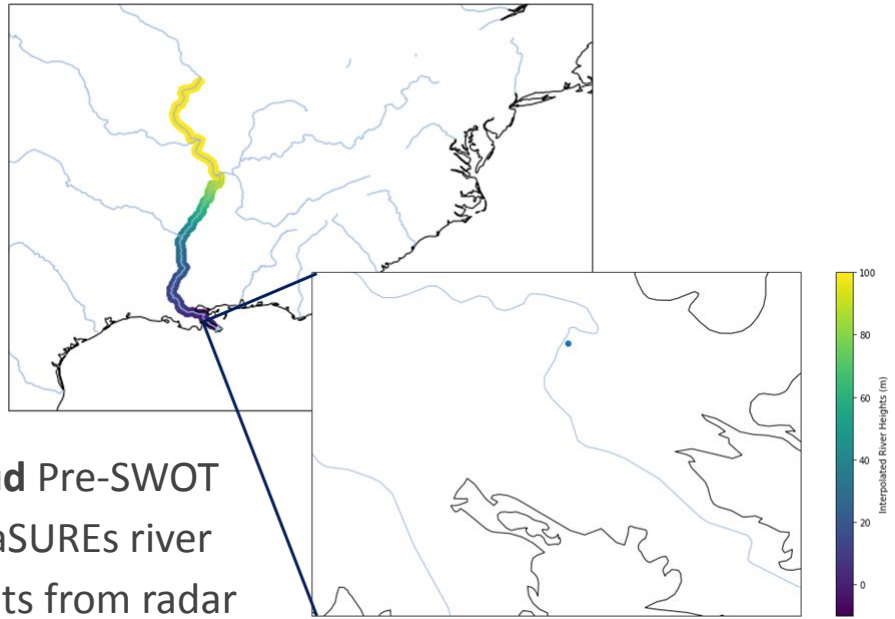


Tutorial introducing working in the AWS cloud

Potential workflow: River Heights [Jupyter Notebook]

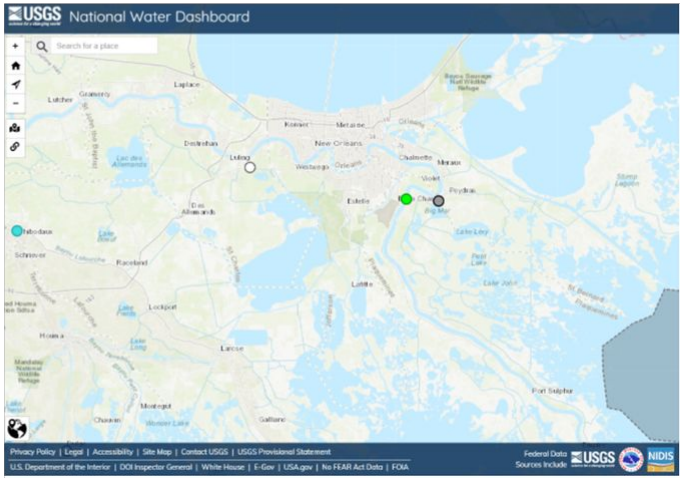


Cloud Use Case: Dataset Validation



Cloud Pre-SWOT MEaSUREs river heights from radar altimetry

In-situ river heights from United States Geological Survey (USGS) gage



CHOOSE YOUR CLOUD ADVENTURE

Jet Propulsion Laboratory
California Institute of Technology