

PO.DAAC Migrates to the Cloud and the River

Cassandra Nickles¹, Catalina Oaida¹, Alireza Farahmand¹, Suresh Vannan², Jinbo Wang¹, Mike Gangl¹, Frank Greguska¹, and Cedric David³

¹NASA Jet Propulsion Laboratory

²Jet Propulsion Laboratory, California Institute of Technology

³Jet 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.



PO.DAAC Migrates to the Cloud and the River

Cassandra Nickles¹, Catalina Oaida¹, Alireza Farahmand¹,
Suresh Vannan¹, Jinbo Wang¹, Michael Gangl¹, Frank
Greguska¹, Cedric David¹

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>

NASA Jet Propulsion Laboratory California Institute of Technology

podaac Physical Oceanography Distributed Active Archive Center

Follow Us

Data Search

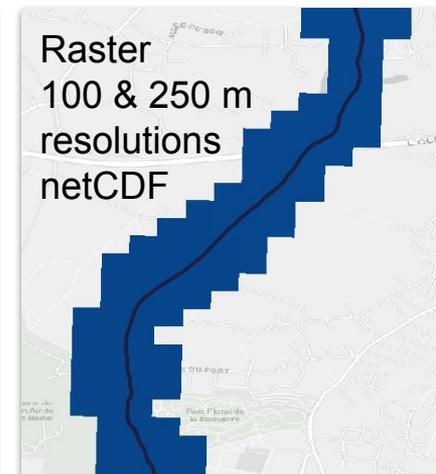
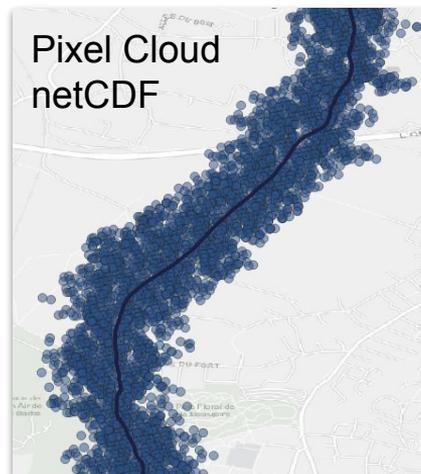
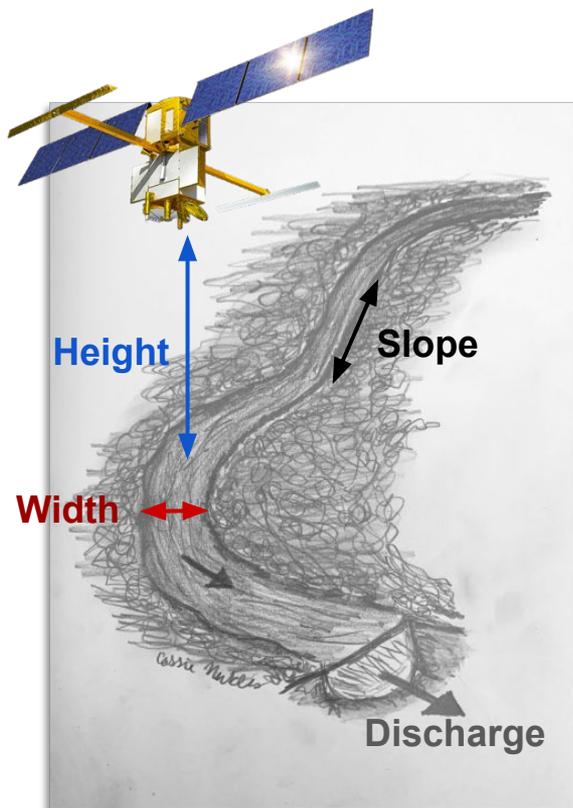
HOME FIND DATA ACCESS DATA RESOURCES ABOUT HELP CLOUD DATA

Surface Water and Ocean Topography (SWOT)

SWOT MISSION OCEAN TERRESTRIAL HYDROSPHERE COAST

ABOUT MISSION DATA NEWS & ANNOUNCEMENTS RESOURCES

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!

Surface Water and Ocean Topography (SWOT)

SWOT MISSION OCEAN TERRESTRIAL HYDROSPHERE COAST

ABOUT MISSION DATA NEWS & ANNOUNCEMENTS RESOURCES

ABOUT MISSION

MISSION OBJECTIVES
SPATIAL COVERAGE
DATASETS
INSTRUMENTS
MISSION CHARACTERISTICS
RELATED LINKS

Datasets

The table below summarizes the SWOT data products, along with product descriptions and sample data products where available. The release date for these sample files is SWOT Sample Data Products v1.2 and all samples can be downloaded [here](#).

Dataset	Description	Coverage	Format	Product Description Document	Sample Data Products v1.2
L2_LR_SSH	Sea surface height data product with data from the KaRIn swath spanning 60 km on both sides of nadir with a nadir gap. Product provides sea surface height, sea surface height anomaly, wind speed, significant waveheight, on a geographically fixed, swath-aligned 2x2 km ² grid, as well as sea surface height on a 250x250 m ² native grid.	Gridded; full swath for each half orbit	netCDF	L2_LR_SSH Product Description Document	L2_LR_SSH.tar.gz Sample Data
L2_HR_PIXC	Point cloud of water mask pixels ("pixel cloud") with geolocated heights, backscatter, geophysical fields, and flags.	Point cloud over tile (approx 64x64 km ²); half swath (left or right side of full swath)	netCDF	L2_HR_PIXC Product Description Document	L2_HR_PIXC.tar.gz Sample Data
L2_HR_RiverSP	Shapefiles of river reaches (approximately 10 km long)	Full swath covering individual	shapefile	L2_HR_RiverSP Product	L2_HR_RiverSP.tar.gz Sample Data

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)

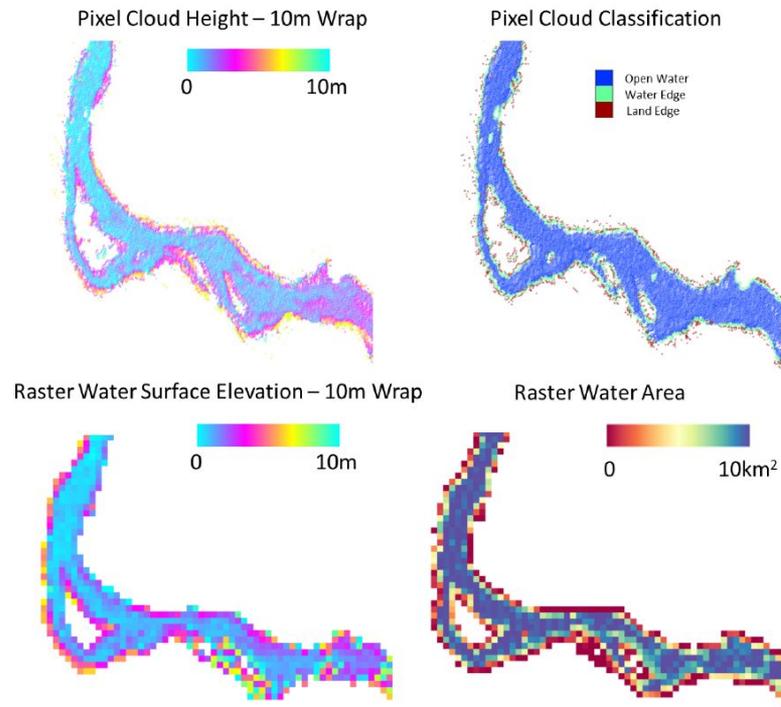


Figure 1. Example pixel cloud height (top left), and classification (top right) from which a raster product is produced, and the resulting raster water surface elevation (bottom left) and water area (bottom right). Note that this image shows layers from a 100 m raster in the UTM projection. Via [D-56416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



The screenshot shows the website for the Surface Water and Ocean Topography (SWOT) mission. The header includes the mission name and four main categories: SWOT MISSION, OCEAN, TERRESTRIAL HYDROSPHERE, and COAST. Below the header is a navigation menu with links for ABOUT MISSION, DATA, NEWS & ANNOUNCEMENTS, and RESOURCES. The main content area displays a message about the mission's launch in 2022 and a table of 13 matching datasets currently available at PO.DAAC. The table lists dataset names, the number of datasets, the time range, and the data format (NETCDF).

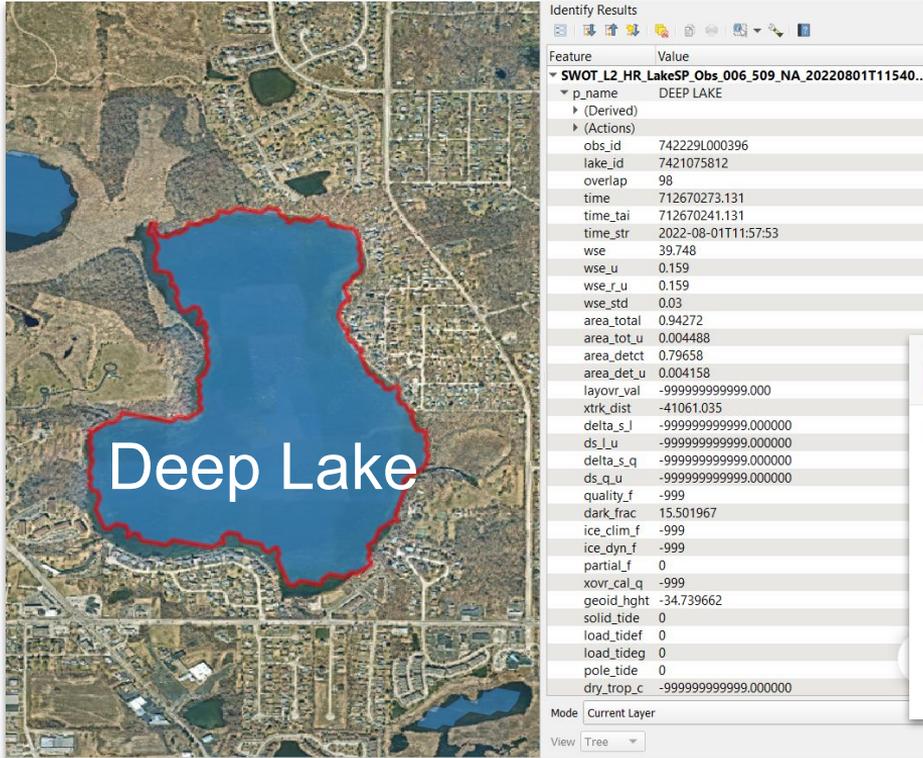
Found 13 matching dataset(s)

TABLE | LIST

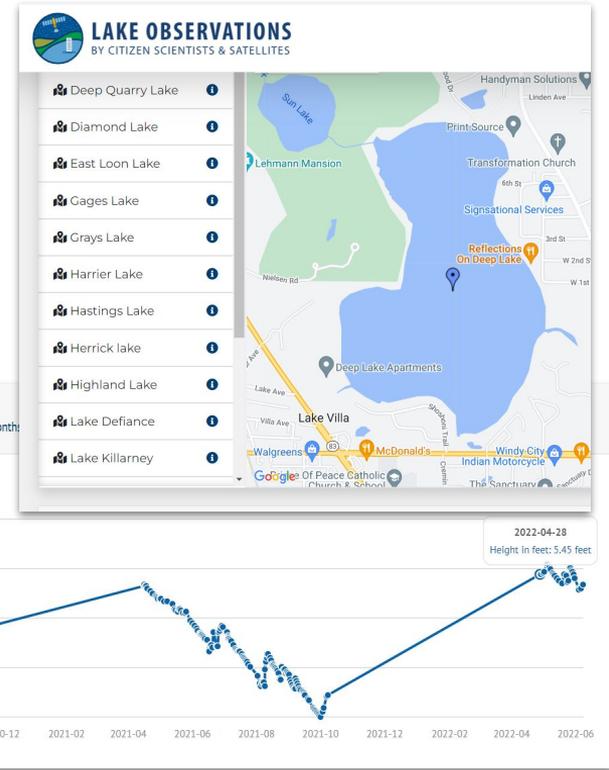
Search Text

Pre SWOT Hydrology Global Lake/Reservoir Storage Time Series V2	4	1992-Sep-25 to 2019-Dec-23	NETCDF
Pre SWOT Hydrology Global Lake/Reservoir Surface Inland Water Height GREALM V.2	2	1992-Sep-25 to 2019-Dec-23	NETCDF
Pre SWOT Hydrology GRRATS Virtual Station River Heights Version 2	2	1992-Apr-08 to 2018-Apr-20	NETCDF
Pre SWOT Hydrology GRRATS Daily River Heights and Storage Version 2	2	1992-Apr-08 to 2018-Apr-20	NETCDF
Pre SWOT Hydrology Global Lake/Reservoir Surface Inland Water Area Extent V2	3	2000-Feb-18 to 2016-Oct-15	NETCDF

Example Future Use Case



Compare to LOCSS Data!



<https://www.locss.org/>

PO.DAAC Migrates to the Cloud

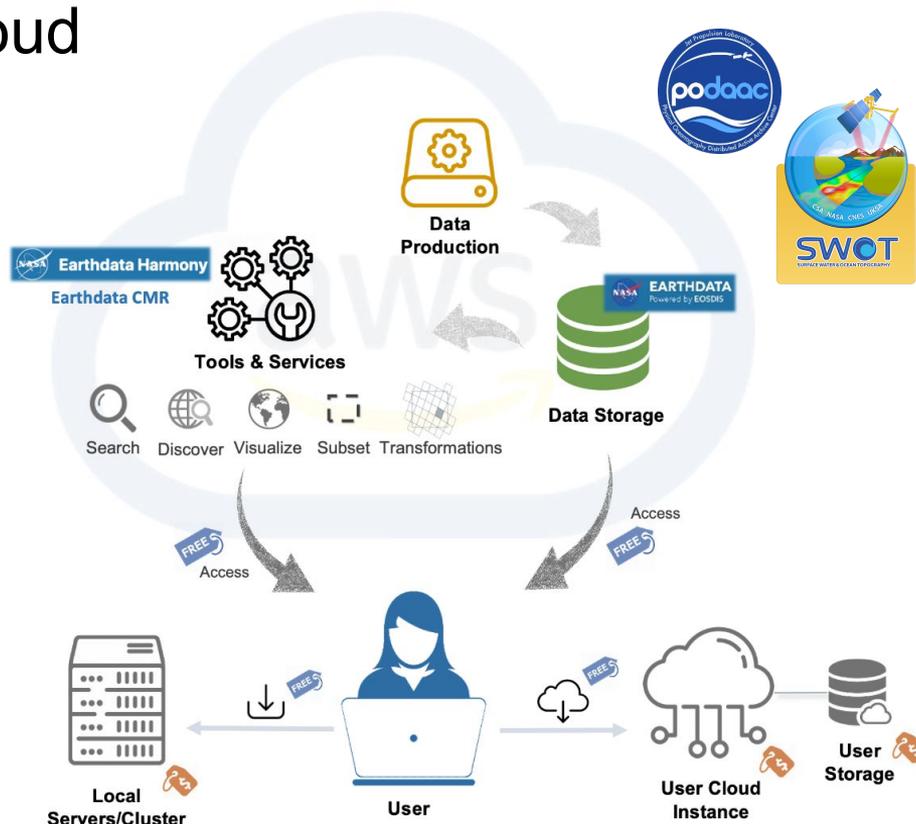
NASA Earthdata Cloud

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

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

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

SWORD River Reach Database

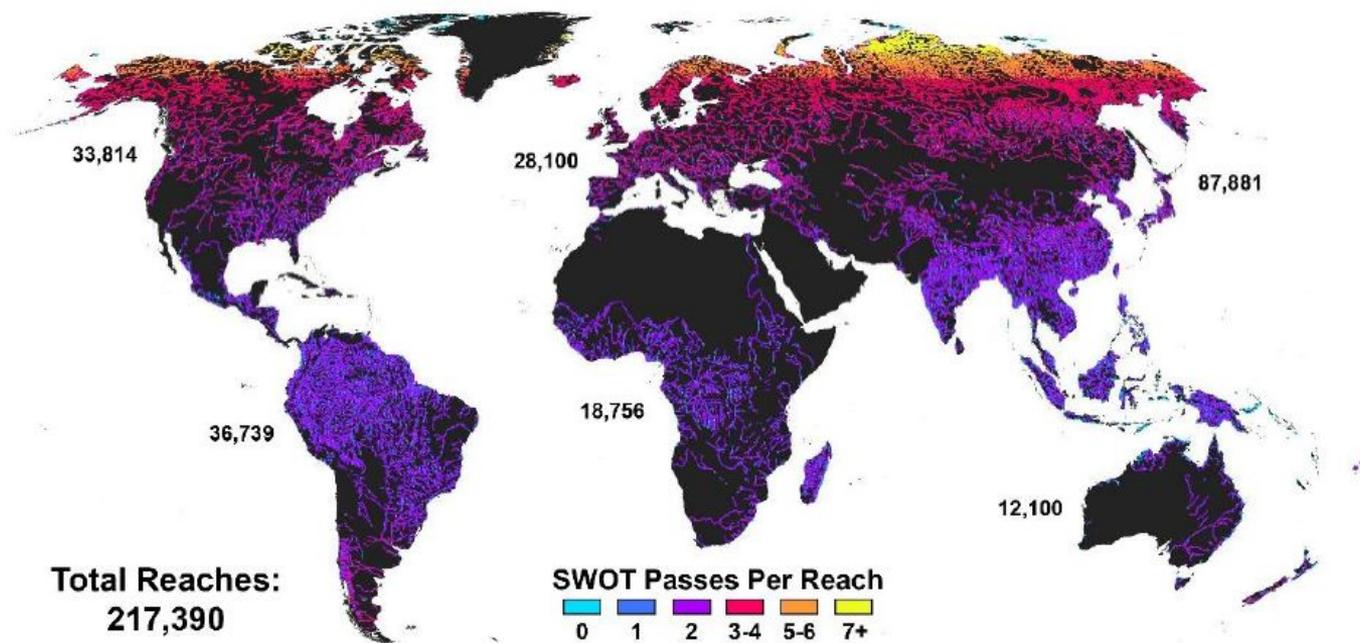
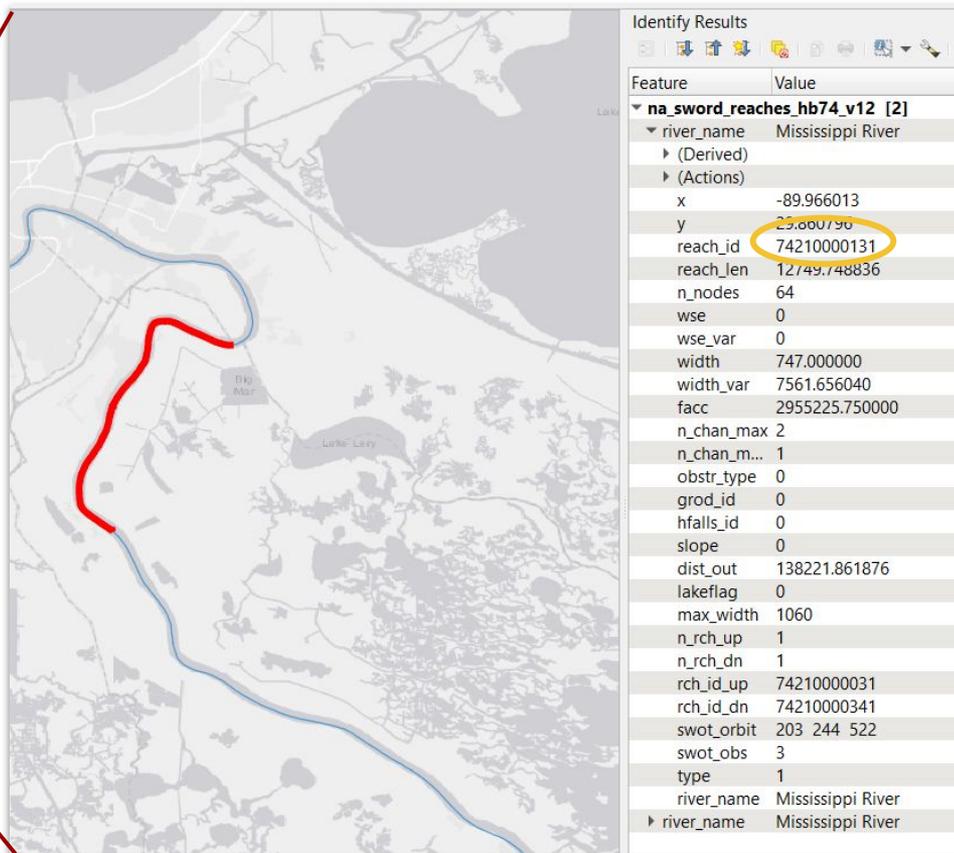
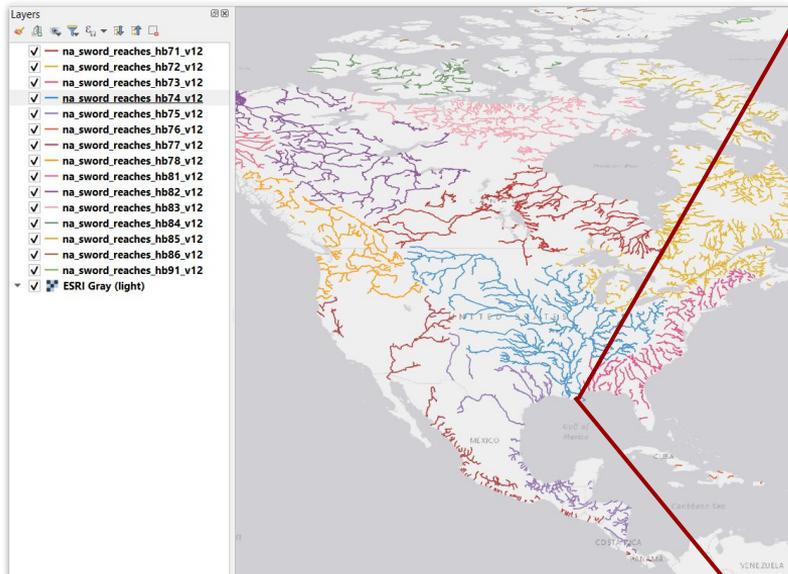


Figure 1: SWORD reach numbers per continent (not including ghost reaches). Colors display the number of SWOT passes per reach during the 21-day orbit cycle.

[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!

Advanced Search

Search by Feature

River Reach

74210000131

Exact match

Search

Advanced Search

[← Back to Feature](#)

1 results for "74210000131"

Select a region from the list below to filter your search results.

REACH 74210000131 *Mississippi River*

5,630 Matching Collections

Showing 20 of 5,630 matching collections

Export Sort View

- SENTINEL-1A_SLC**
177 Granules • 2014-04-03 ongoing • Sentinel-1A slant-range product
GEOSS SENTINEL-1A_SLC v1 - ASF
- SENTINEL-1B_SLC**
3 Granules • 2016-04-25 ongoing • Sentinel-1B slant-range product
GEOSS SENTINEL-1B_SLC v1 - ASF
- SENTINEL-1A_DUAL_POL_GRD_HIGH_RES**
177 Granules • 2014-04-03 ongoing • Sentinel-1A Dual-pol ground projected high and full resolution images
GEOSS SENTINEL-1A_DP_GRD_HIGH v1 - ASF
- SENTINEL-1B_DUAL_POL_GRD_HIGH_RES**
3 Granules • 2016-04-25 ongoing • Sentinel-1B Dual-pol ground projected high and full resolution images
GEOSS SENTINEL-1B_DP_GRD_HIGH v1 - ASF
- HLS Landsat Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m v2.0**
378 Granules • 2013-04-11 ongoing • The Harmonized Landsat Sentinel-2 (HLS) project provides consistent surface reflectance (SR) and top of atmosphere (TOA) brightness data from a virtual constellation of satellite sensors.



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

```
{'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 reach



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 (visualizing) the queried region of interest, and using FTS results to query data through NASA's Common Metadata Repository (CMR).

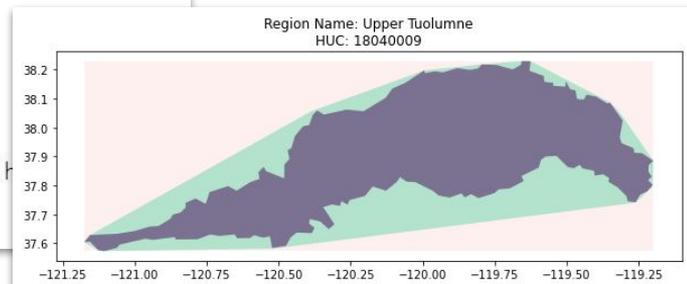
Example Use Case: Check if data is available over my region of interest using HUCs. In this example we are using the FTS-HUC API (<https://fts.podaac.earthdata.nasa.gov/>) to geospatially define our region of interest, namely the Upper Tuolumne River Basin in the San Joaquin River Basin in California's 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 identify region of interest (e.g. HUC value or name) can be found at <https://water.usgs.gov/GIS/regions.html>

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



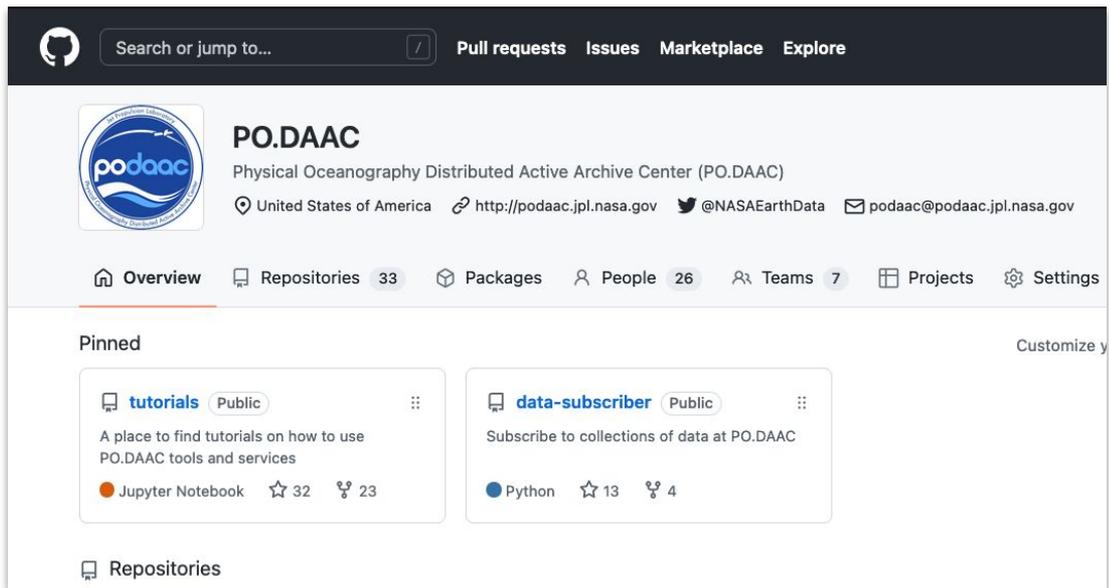
<https://github.com/podaac/tutorials/blob/master/notebooks/HUC%20Feature%20Translation%20Service%20Examples-updated-20210804.ipynb>

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



The screenshot shows the GitHub profile for PO.DAAC. At the top, there is a search bar and navigation links for Pull requests, Issues, Marketplace, and Explore. The profile header includes the PO.DAAC logo, the name "PO.DAAC", and the full name "Physical Oceanography Distributed Active Archive Center (PO.DAAC)". Below this, it lists the location "United States of America", the website "http://podaac.jpl.nasa.gov", the Twitter handle "@NASAEarthData", and the email "podaac@podaac.jpl.nasa.gov". A navigation bar shows "Overview" as the active tab, along with "Repositories 33", "Packages", "People 26", "Teams 7", "Projects", and "Settings". The "Pinned" section features two repositories: "tutorials" (Public) described as "A place to find tutorials on how to use PO.DAAC tools and services" with 32 stars and 23 forks, and "data-subscriber" (Public) described as "Subscribe to collections of data at PO.DAAC" with 13 stars and 4 forks. A "Repositories" section is partially visible at the bottom.

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.

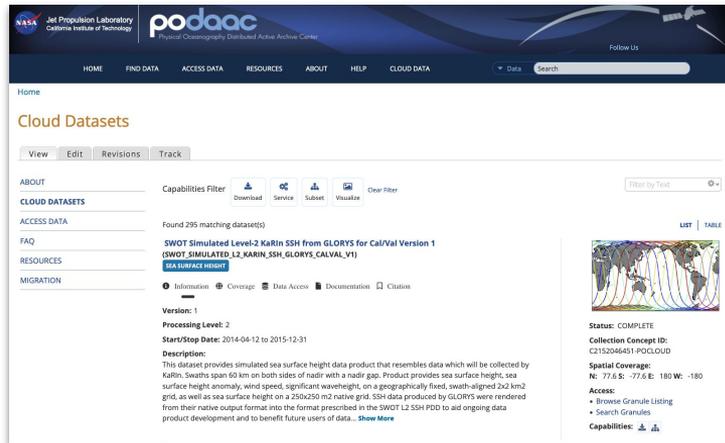
Water Resources Research, 57, e2021WR030054.

<https://doi.org/10.1029/2021WR030054>

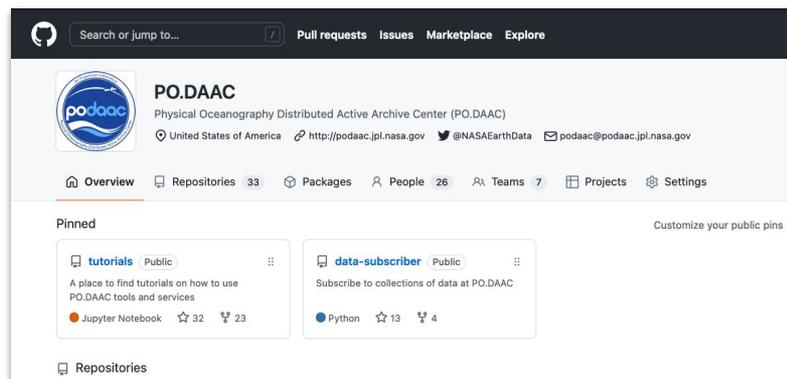
Extra Slides

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)

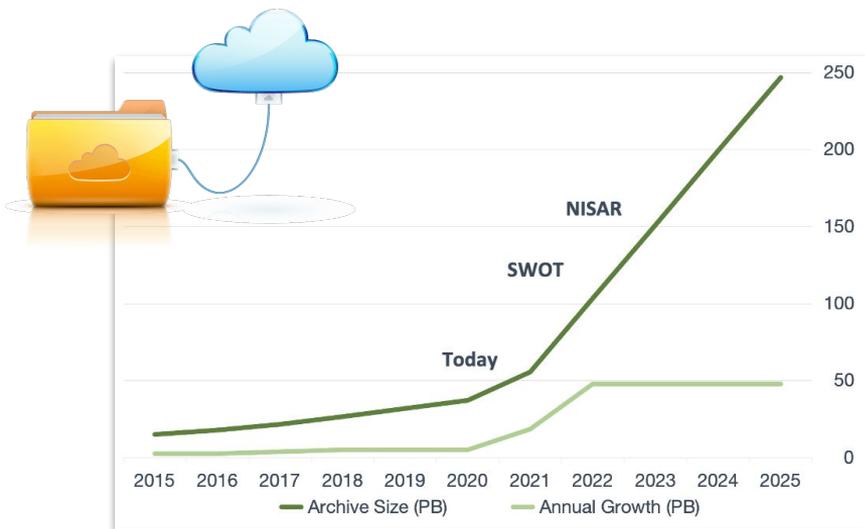


<https://podaac.jpl.nasa.gov/cloud-datasets>



<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>

- CLOUD DATASETS
- ABOUT
 - Earthdata Harmony
 - Earthdata CMR
 - Earthdata Search
- ACCESS DATA
- FAQ
- RESOURCES
- MIGRATION**

CLOUD DATA - ABOUT

Introduction: Access to PO.DAAC datasets in the cloud

PO.DAAC is in the process of moving its data holdings to the cloud. The **Cloud Data** page at PO.DAAC offers access cloud-based datasets as well as resources to help guide data users in discovering, accessing, and utilizing cloud data.

The **Cloud Datasets** section provides a listing page for cloud-archived datasets, with more tools/services integration. The **Resources** section shares information, updates, data recipes, and other materials that help support the user in discovering, accessing and using datasets from and within the Earthdata Cloud. The **Migration** section offers information on the transition timeline and datasets, what to expect, and migration-specific FAQs and tutorials. For questions on what this transition means, please see the **FAQ** section.

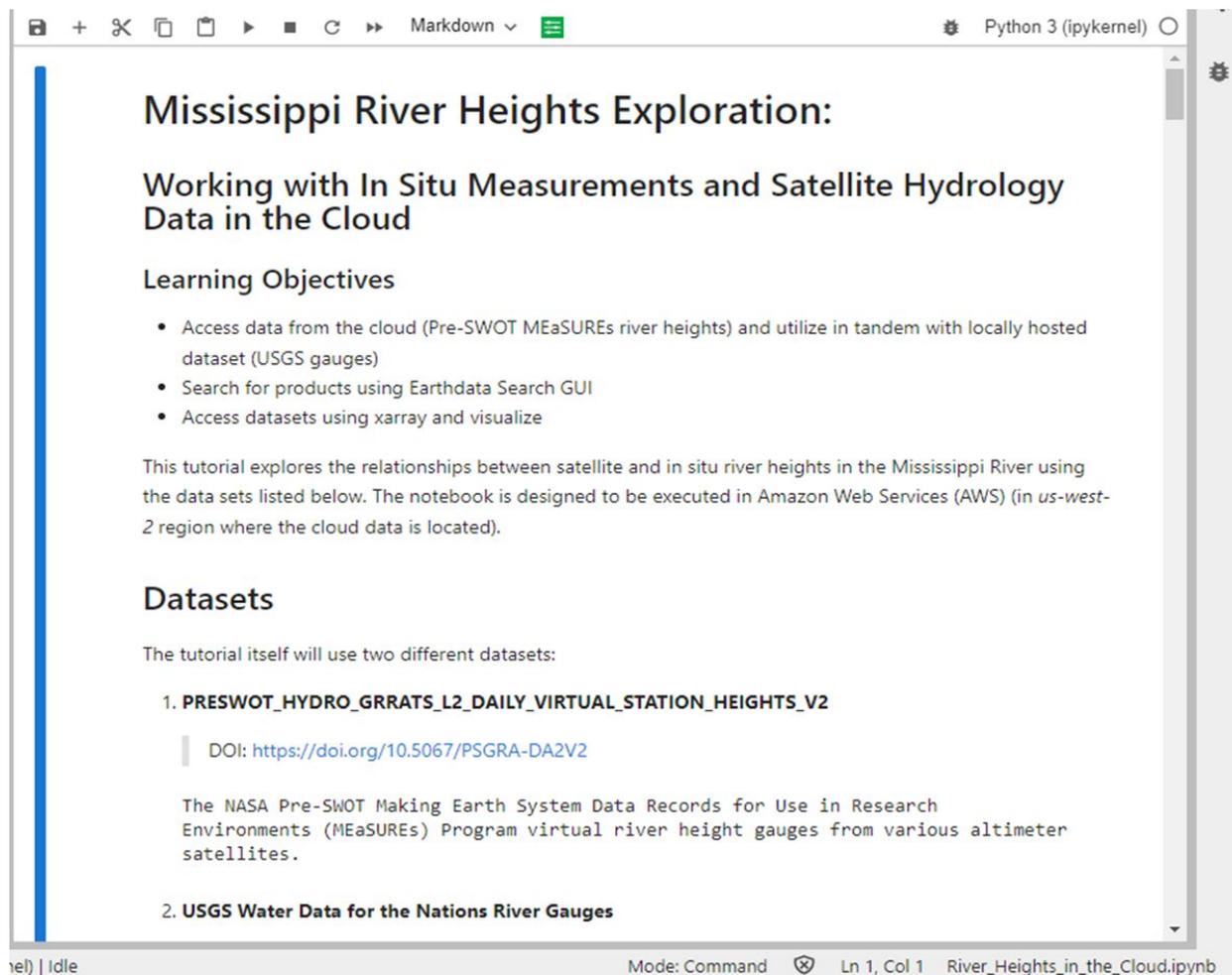
During this transition to the cloud, this Cloud Data page will be evolving and continuously updated with new content and data - please check back regularly.

What does the new cloud paradigm look like?

The diagram shows a central cloud icon containing 'Data Production' and 'Data Storage' icons. Arrows indicate data flow between these components and external 'Data Storage' icons. Below the cloud, a user is shown interacting with a laptop, with arrows indicating 'Access' to the cloud data. A 'Tools & Services' box includes icons for 'Search', 'Discover', 'Visualize', and 'Subset/Transformations'.

Tutorial introducing working in the AWS cloud

Potential workflow:
River Heights [Jupyter
Notebook](#)



Python 3 (ipykernel)

Mississippi River Heights Exploration: Working with In Situ Measurements and Satellite Hydrology Data in the Cloud

Learning Objectives

- Access data from the cloud (Pre-SWOT MEaSURES river heights) and utilize in tandem with locally hosted dataset (USGS gauges)
- Search for products using Earthdata Search GUI
- Access datasets using xarray and visualize

This tutorial explores the relationships between satellite and in situ river heights in the Mississippi River using the data sets listed below. The notebook is designed to be executed in Amazon Web Services (AWS) (in *us-west-2* region where the cloud data is located).

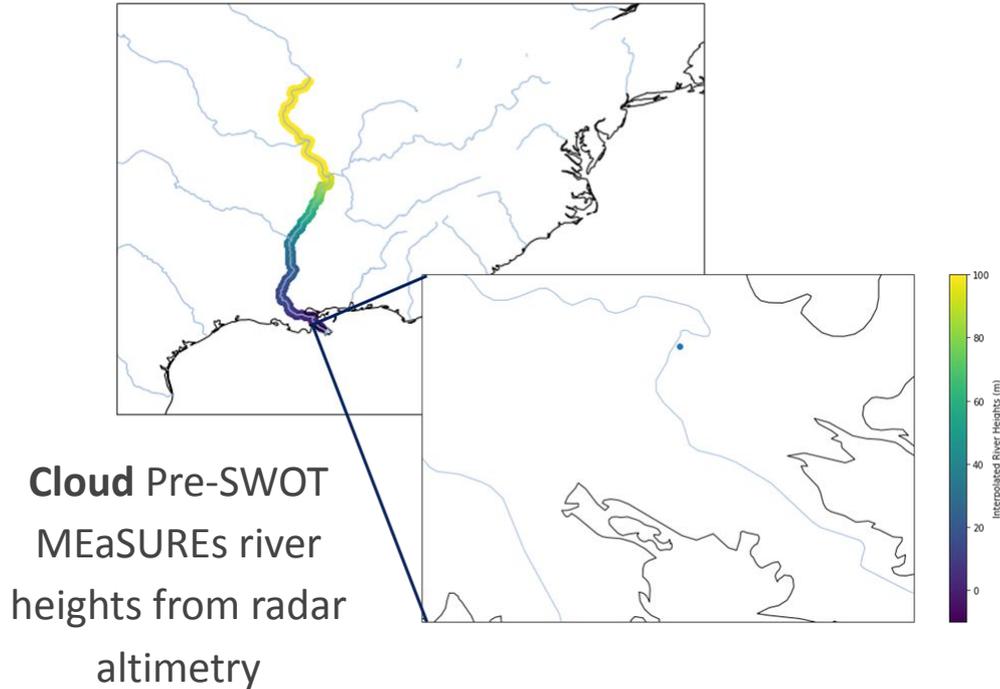
Datasets

The tutorial itself will use two different datasets:

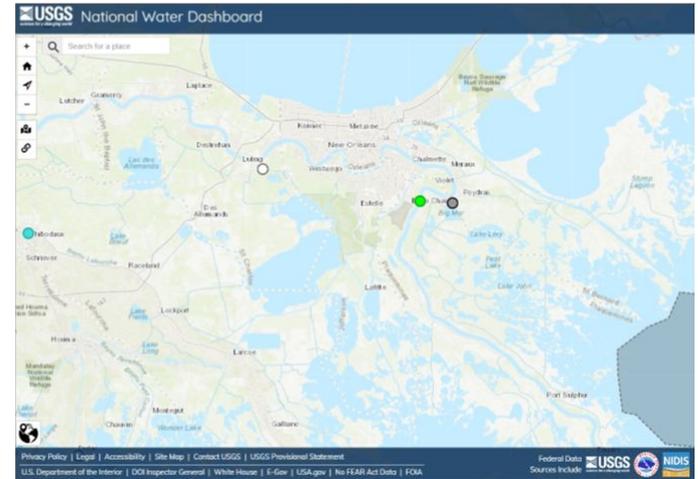
1. **PRESWOT_HYDRO_GRRATS_L2_DAILY_VIRTUAL_STATION_HEIGHTS_V2**
DOI: <https://doi.org/10.5067/PSGRA-DA2V2>
The NASA Pre-SWOT Making Earth System Data Records for Use in Research Environments (MEaSURES) Program virtual river height gauges from various altimeter satellites.
2. **USGS Water Data for the Nations River Gauges**

re) | Idle Mode: Command Ln 1, Col 1 River_Heights_in_the_Cloud.ipynb

Cloud Use Case: Dataset Validation



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



CHOOSE YOUR CLOUD ADVENTURE

