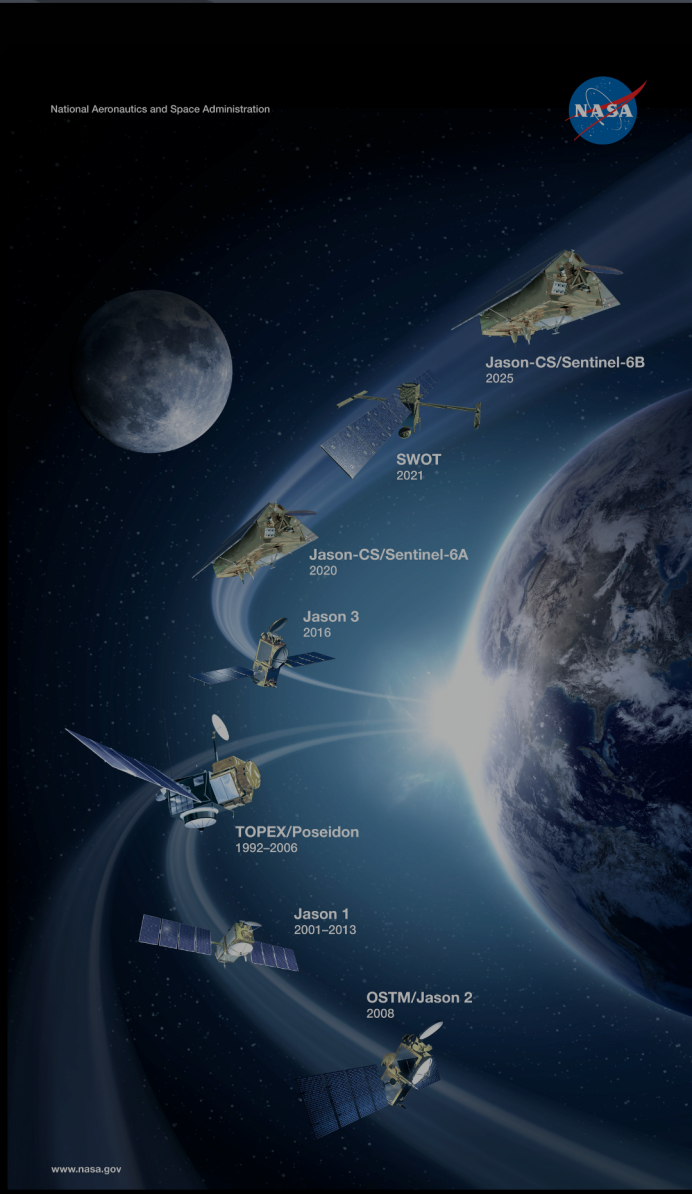
Jet Propulsion Laboratory
California Institute of Technology

podaac
Physical Oceanography Distributed Active Archive Center

The Road Towards Cloud-Based Data Access, Usability and Interoperability – a DAAC's tale

Catalina Oaida, Michael Gangl, Jessica Hausman, Michelle Gierach, Suresh Vannan

Jet Propulsion Laboratory, California Institute of Technology



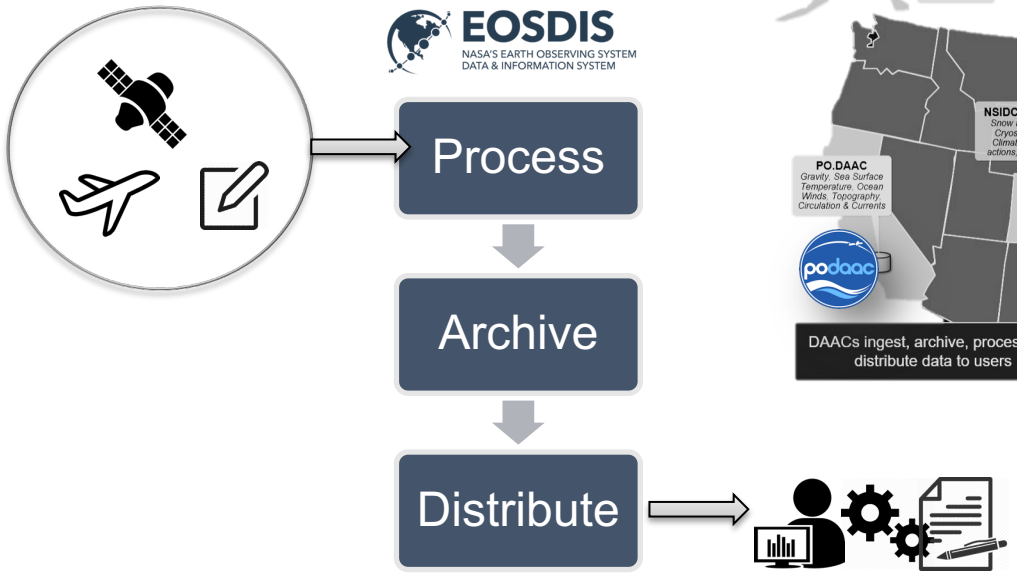
December 9, 2019
American Geophysical Union
Fall Meeting
San Francisco, California

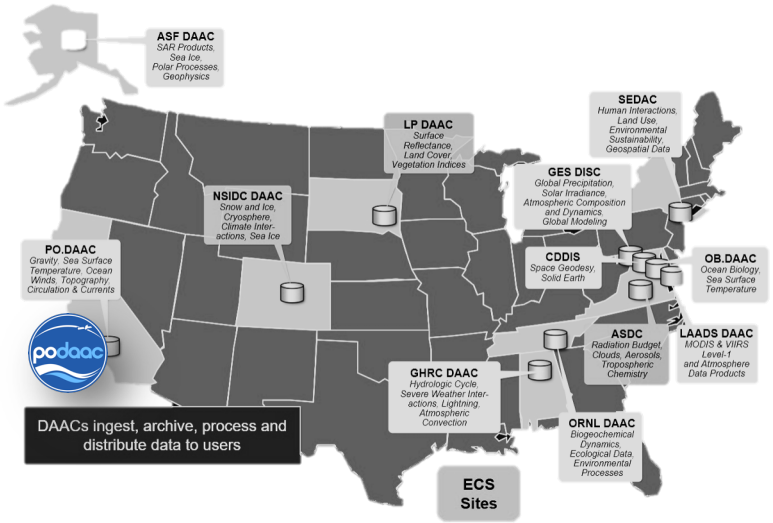
Contact:
Catalina.Oaida@jpl.nasa.gov

© 2019 California Institute of Technology. Government sponsorship acknowledged.

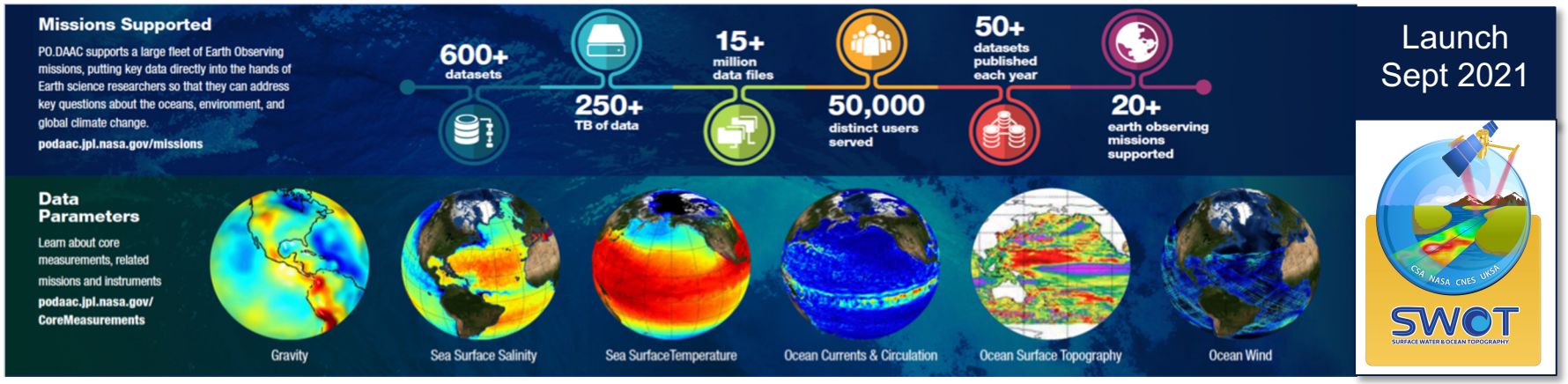
Physical Oceanography Distributed Active Archive Center (PO.DAAC)

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





Missions, Projects, Parameters



Launch Sept 2021

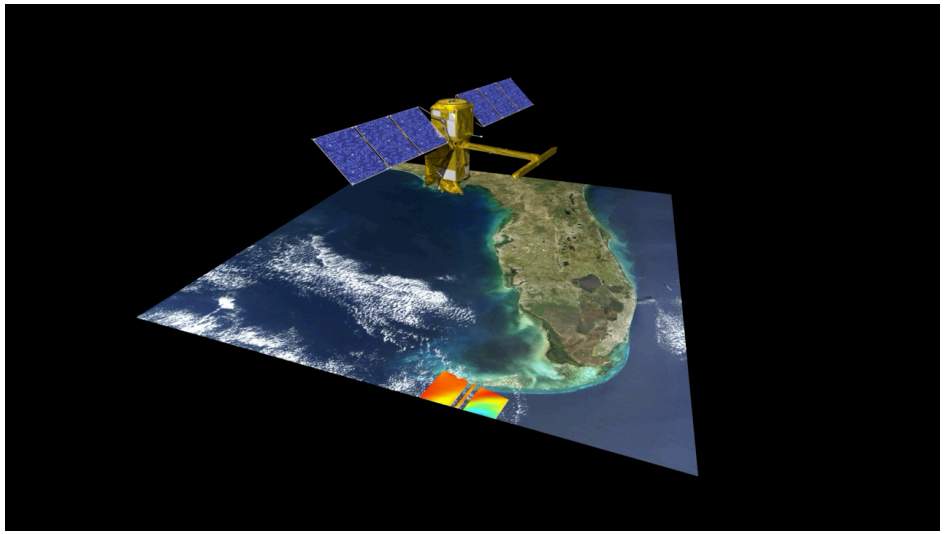
Tools

© 2019 California Institute of Technology. Government sponsorship acknowledged.

Surface Water and Ocean Topography (SWOT)

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

SWOT will measure global ocean surface topography and land surface water elevation with great accuracy using interferometry.

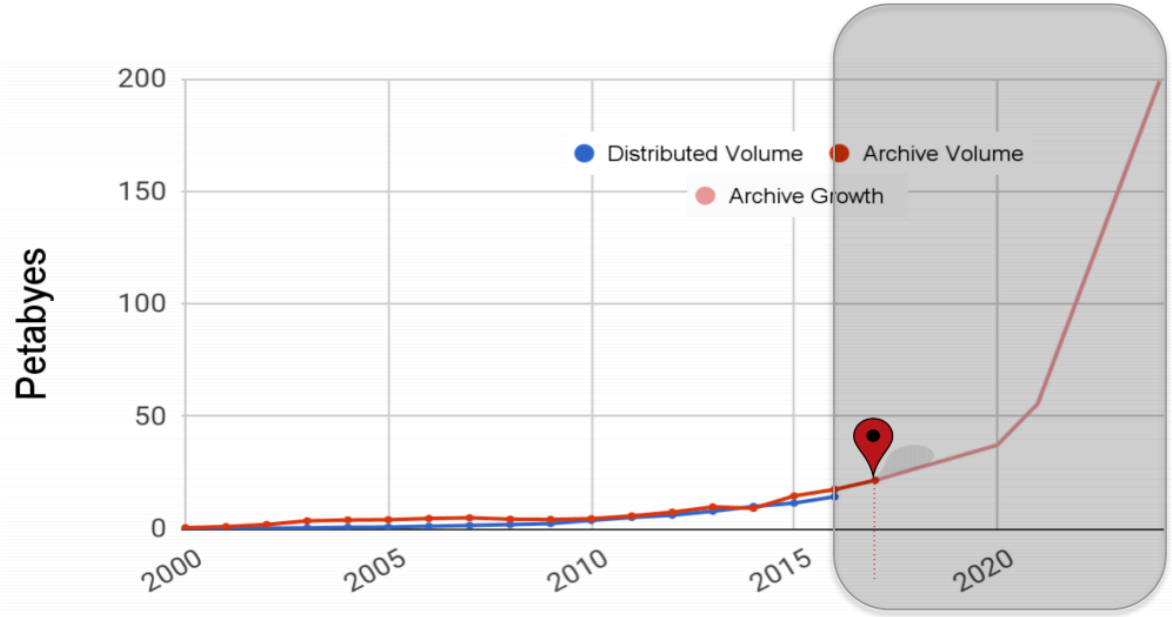


Credit: NASA/JPL-Caltech

Applications:

- Flood and Drought Monitoring
- Water Resources Management
- Inland and Coastal Commerce
- Coastal Zone Management
- Climate and Weather

Launch Date: September, 2021
26 PB **total mission volume** (3 years)
3.7 PB of **Level-2 volume**

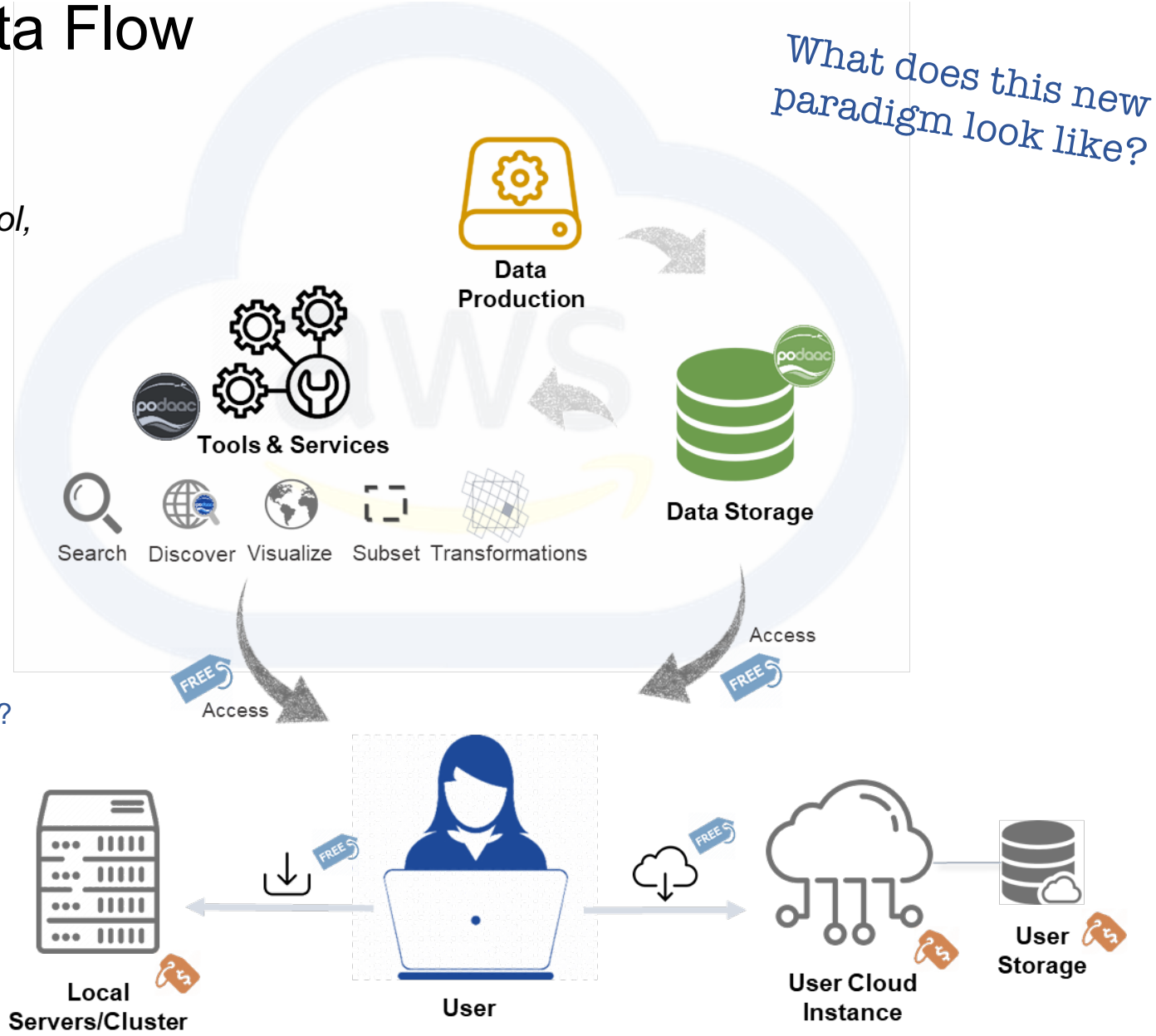


© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud Data Flow

What does this new paradigm look like?

"moving data processing and analysis, but not control, away from users and closer to the data"



Some Questions

- What services are needed?
- What is the priority for the development of cloud capabilities?
- What type of analysis will be performed?
- Are end-users ready to use the cloud?

© 2019 California Institute of Technology. Government sponsorship acknowledged.

End User-driven Cloud Services

Moving data processing and analysis, but not control, away from users and closer to the data.

Application Journeys & Survey → Services Roadmap → Implementation Plan

User Needs Assessment → **Database** → **Use Cases** → **PO.DAAC & ESDIS Cloud Services & Roadmap**

- **SWOT Survey 2.0** (n=111)
  - SWOT Science Team
  - SWOT Early Adopters
  - PO.DAAC User Working Group
  - PO.DAAC SOTO use cases
  - SWOT Hydrology wishlist
- **Application Journeys** (n=65)

→

- Application **data requirements** and **user capabilities**
- **User workflows** (use case traceability matrix)

→

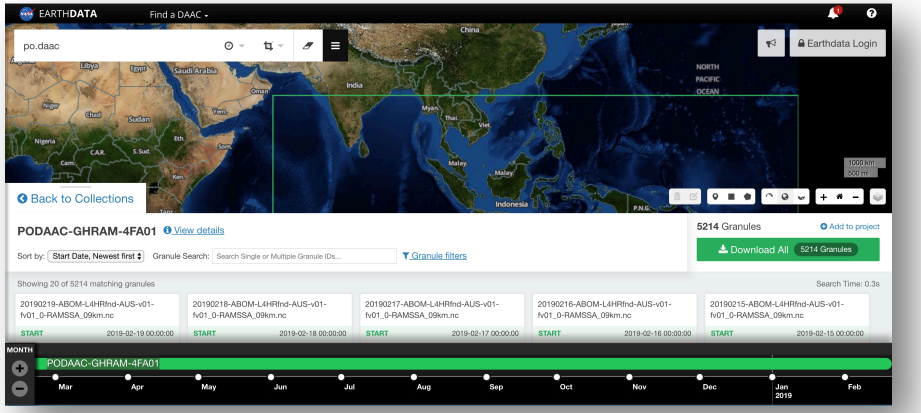
- Prioritized use cases based on **% users impacted**
- Use cases can be looked at by **User Persona** (e.g. oceans, hydrology, or coastal)
- Use cases complemented by user data preferences (e.g. data file format, projections, software & tools)

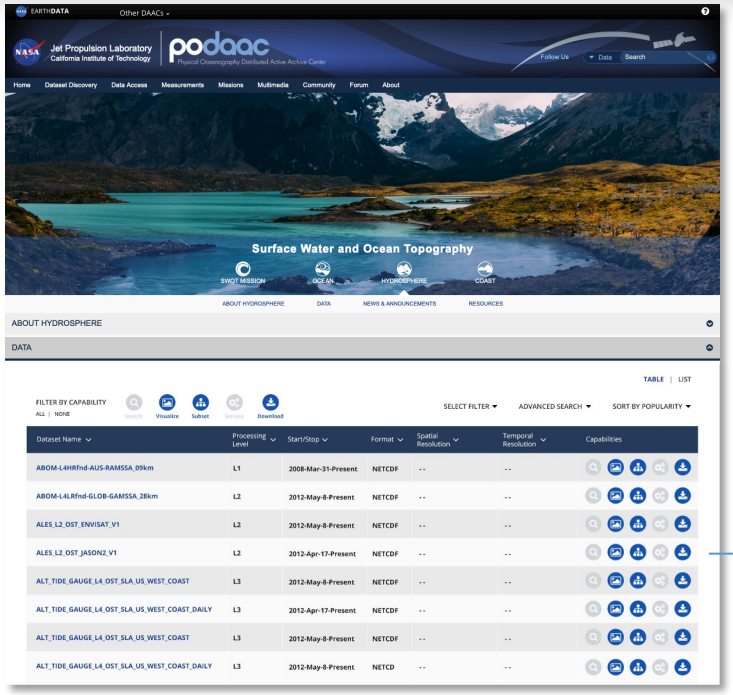
Services on the Cloud

Data processing

© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud: Hydrology User Workflow





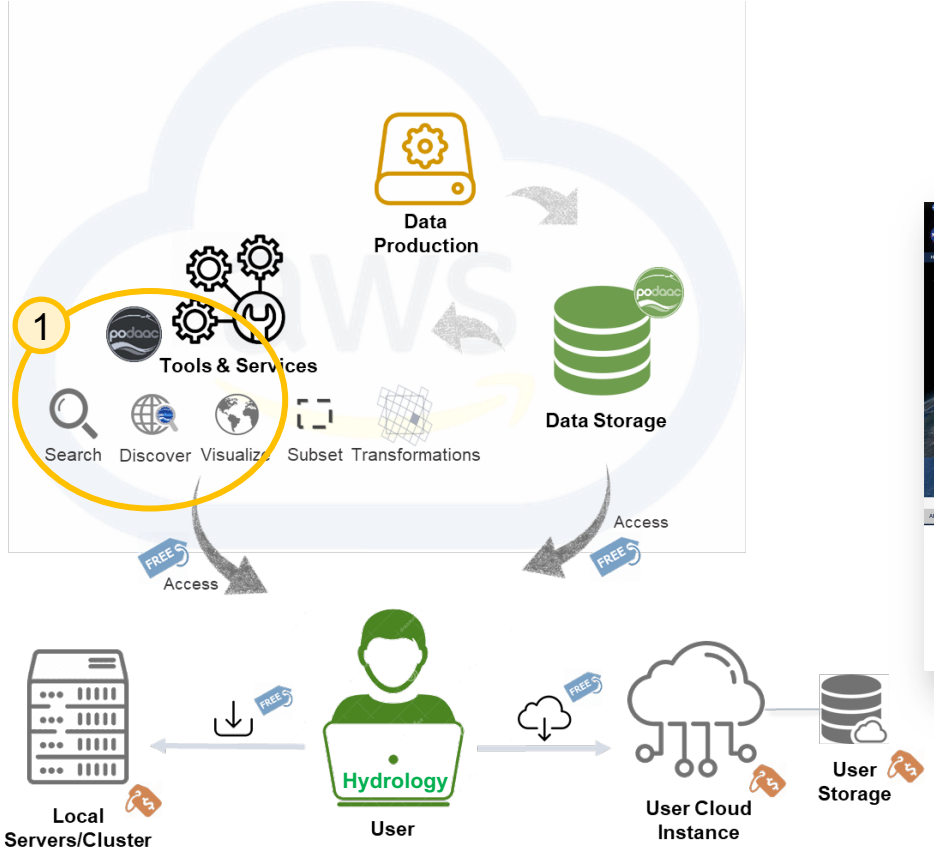
mockup

1 Data Search and Discovery

- **Earthdata Search** (available now)
  https://search.earthdata.nasa.gov/search

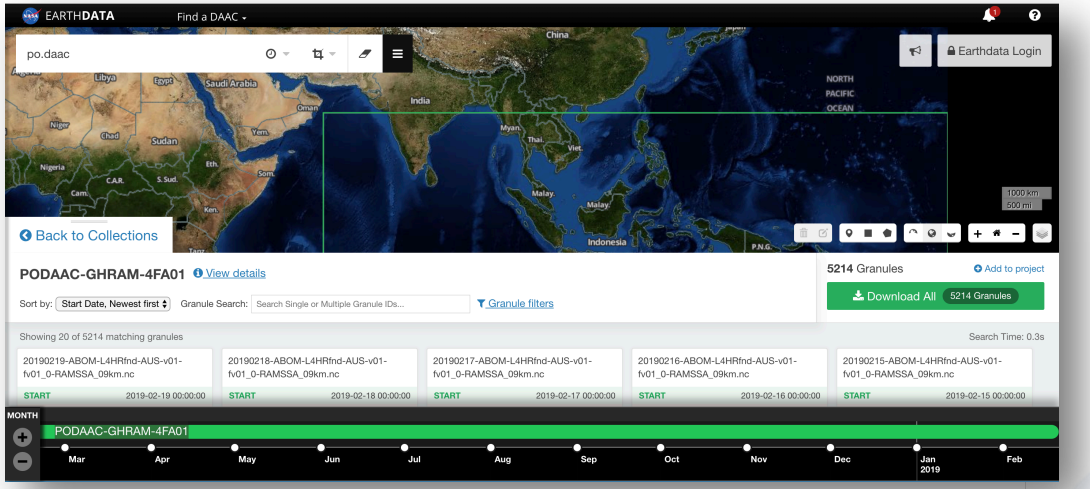
- **PO.DAAC SWOT Search portal** (available now*)
- https://podaac.jpl.nasa.gov/swot

*No SWOT data yet (launch Fall 2021). Mission data information and relevant data available.



© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud: Hydrology User Workflow



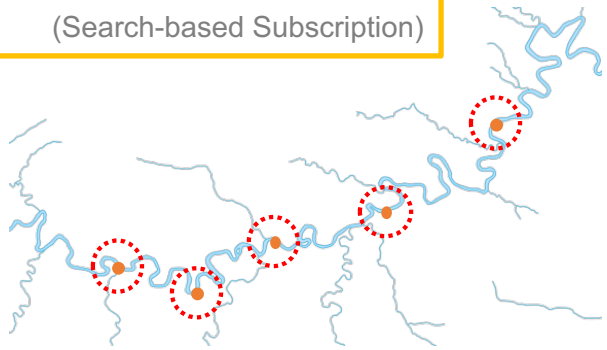
2

SWOT Feature **Search**

HUC **Search**

Subscribe by HUC

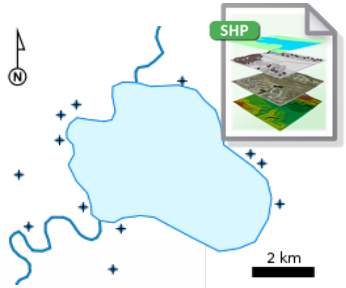
(Search-based Subscription)

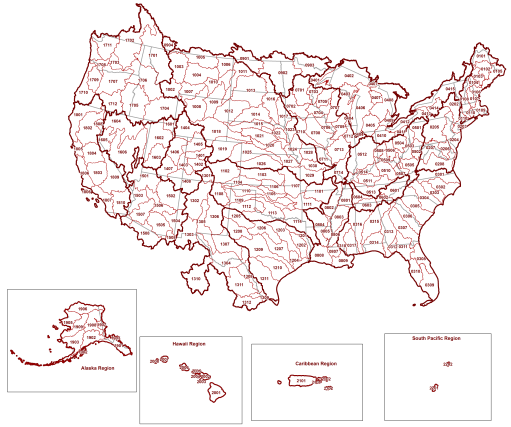


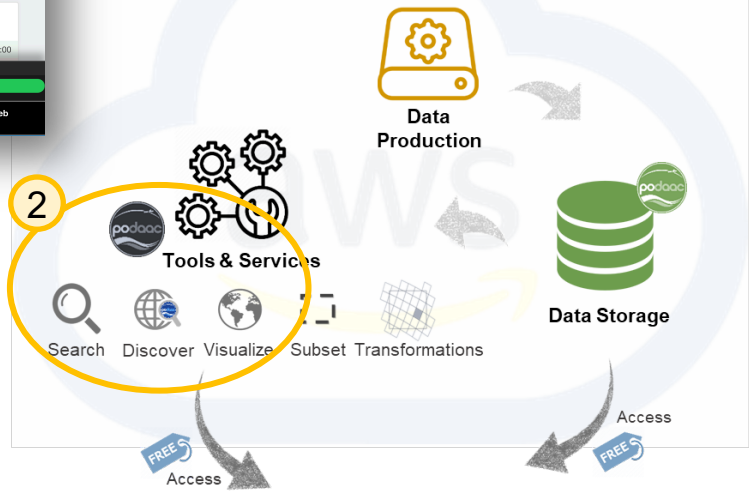
Point-Radius & Polygon Search

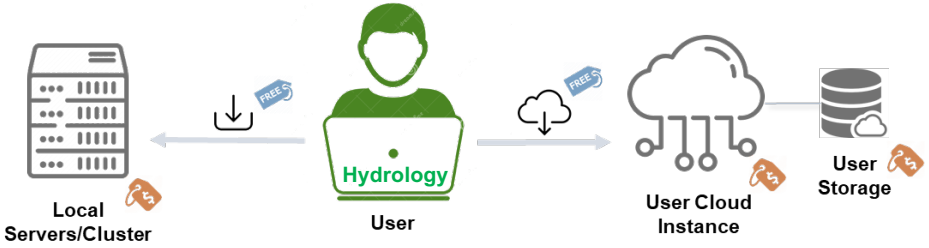
Regional Search – HUC and SWOT Feature IDs

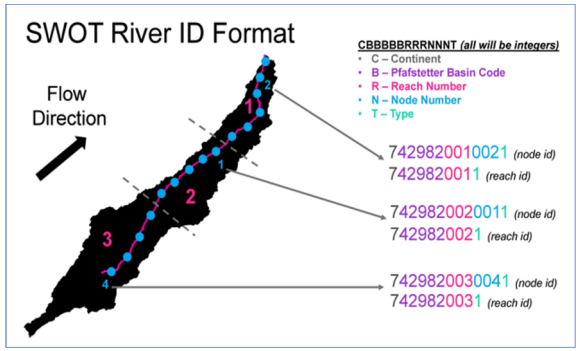
Custom Shapefile Search





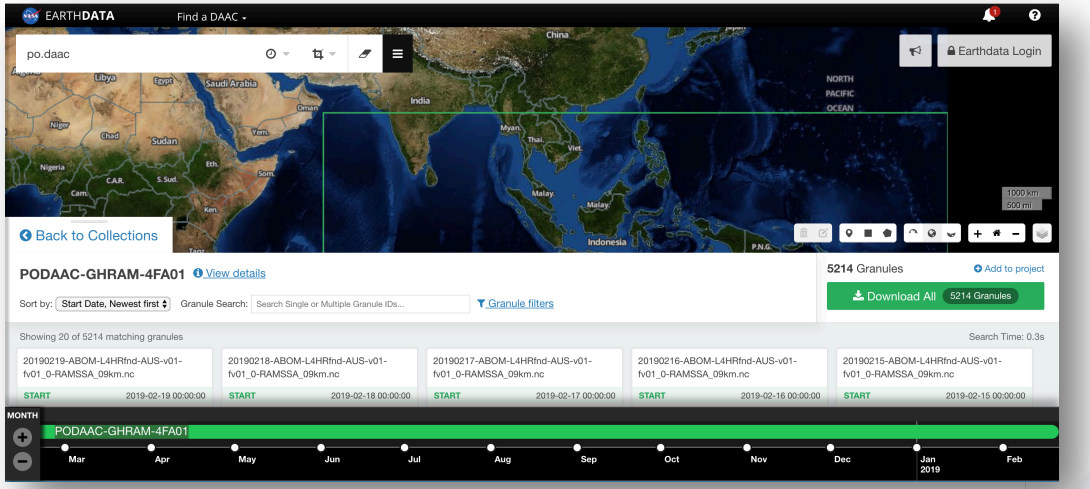






© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud: Hydrology User Workflow

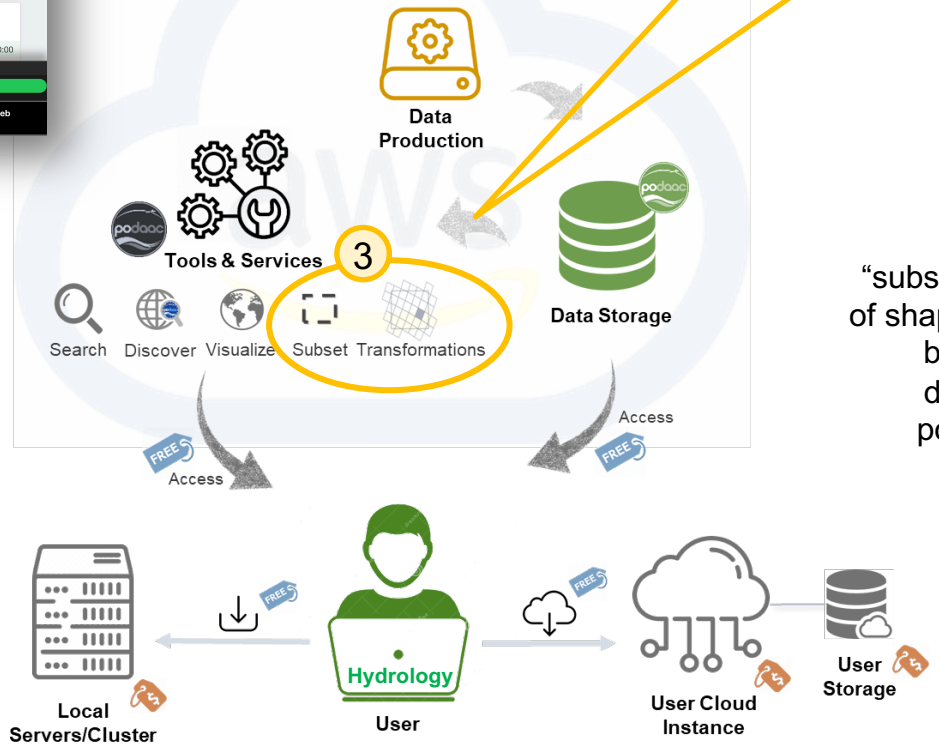


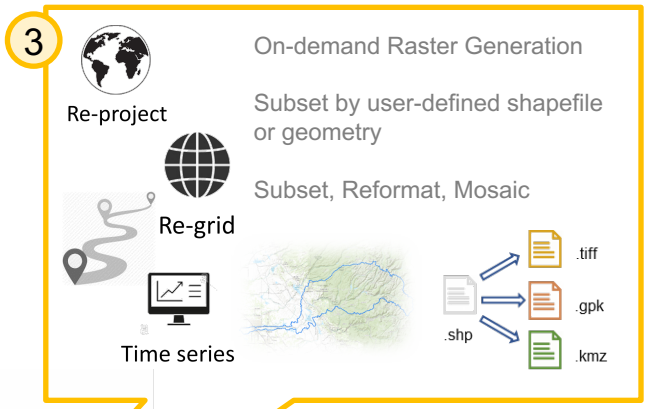
Subscribe to any search parameters available:

- Feature ID or HUC
- Points, Lines, & Geometries
- User supplied shapefiles

Digest or per-granule (search hit) notifications

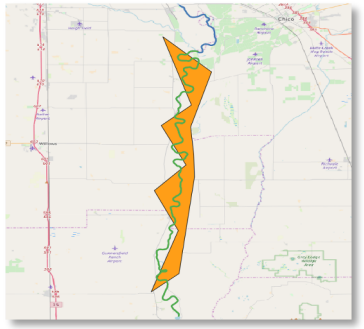
Combine with transformations



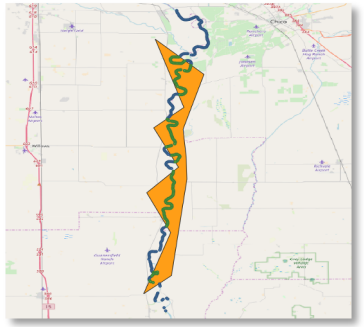


"subsetting" of shapefiles by user defined polygon

Filtering Reaches

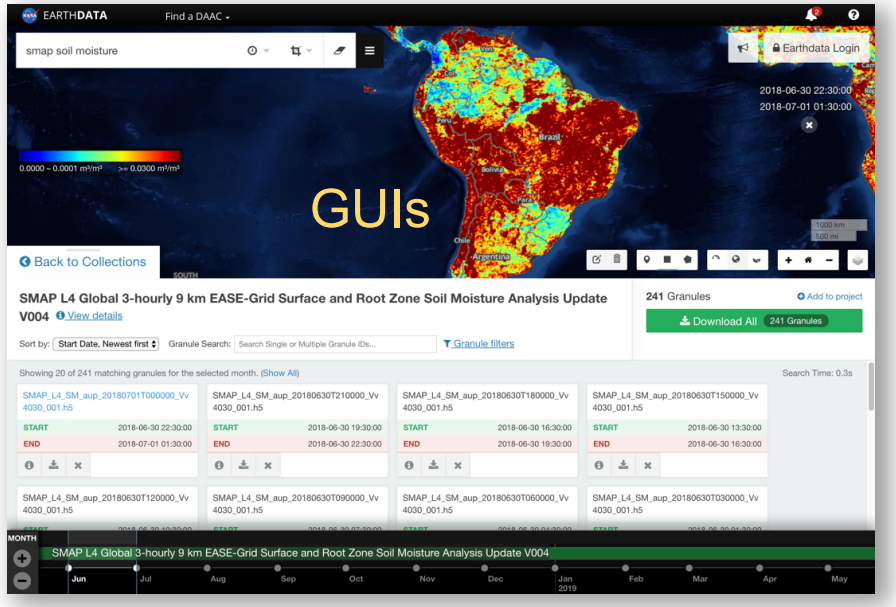


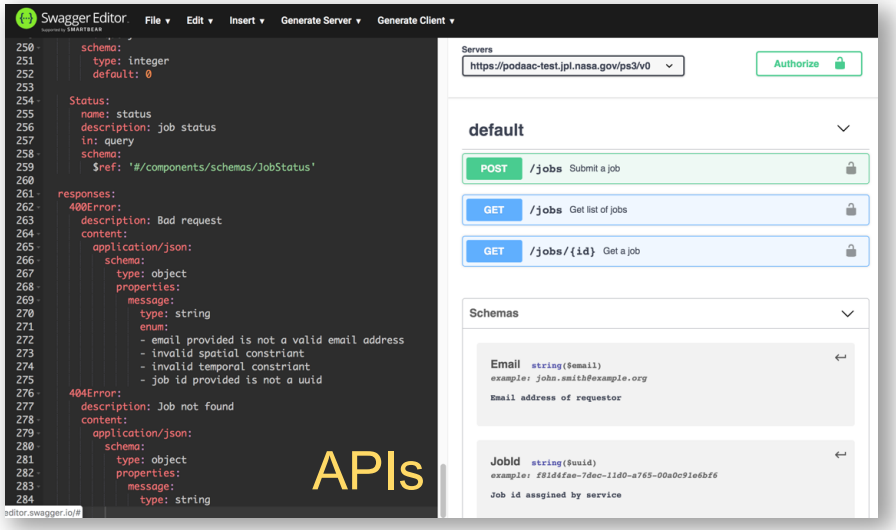
Filtering Nodes

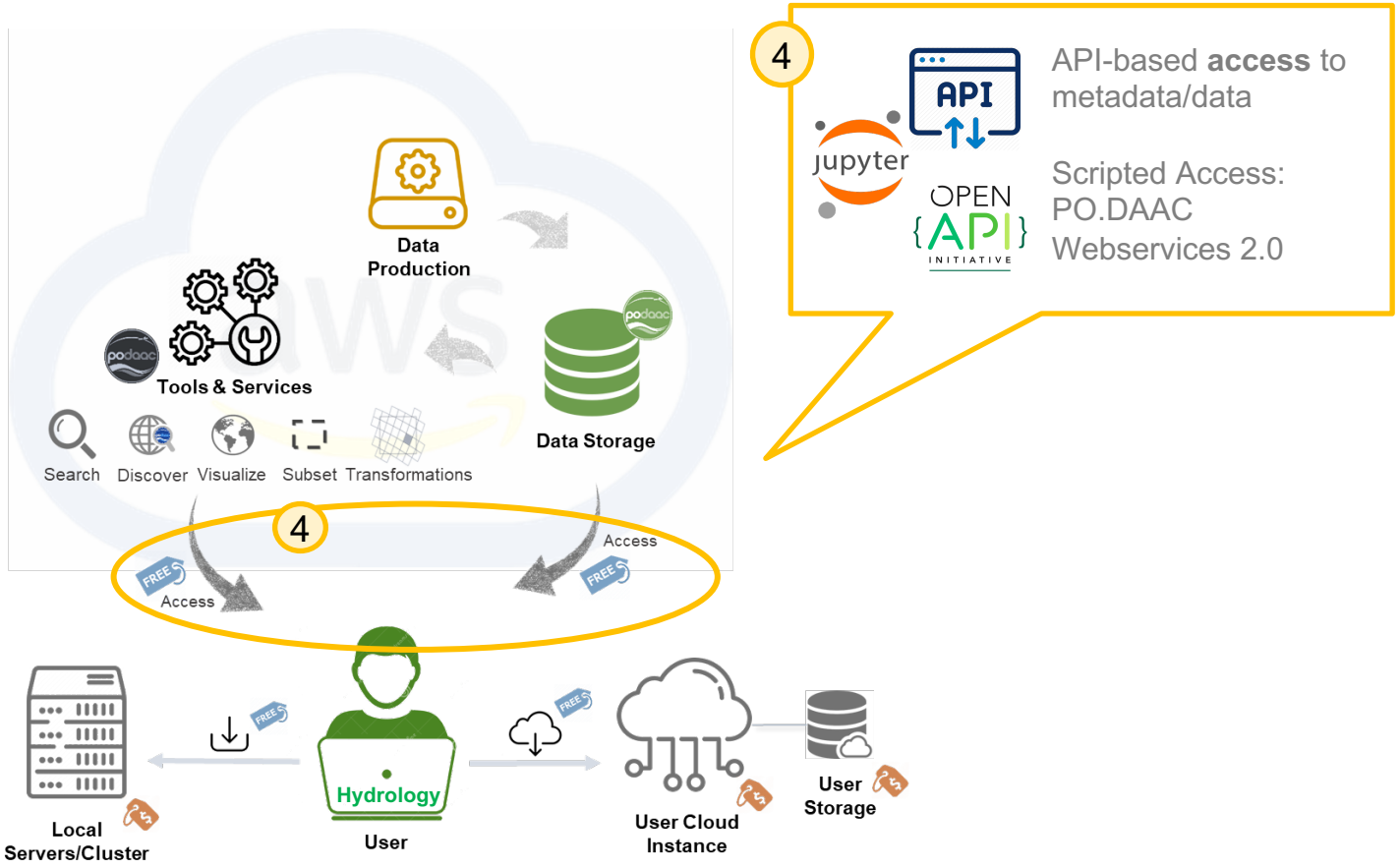


© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud: Hydrology User Workflow

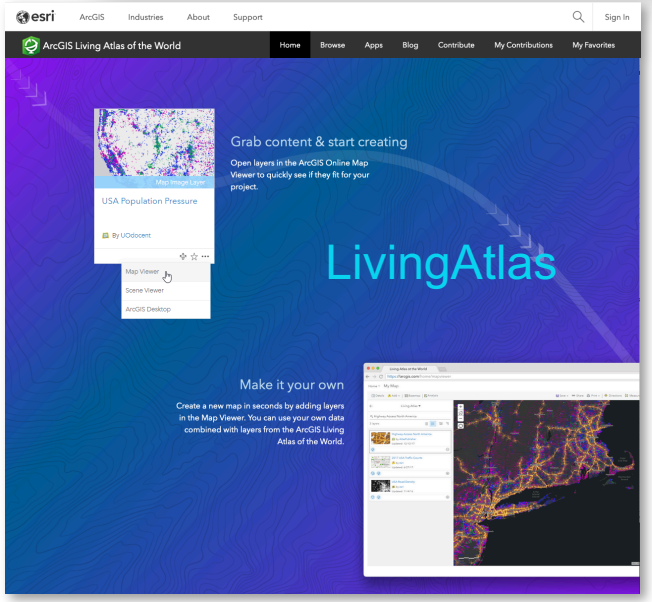


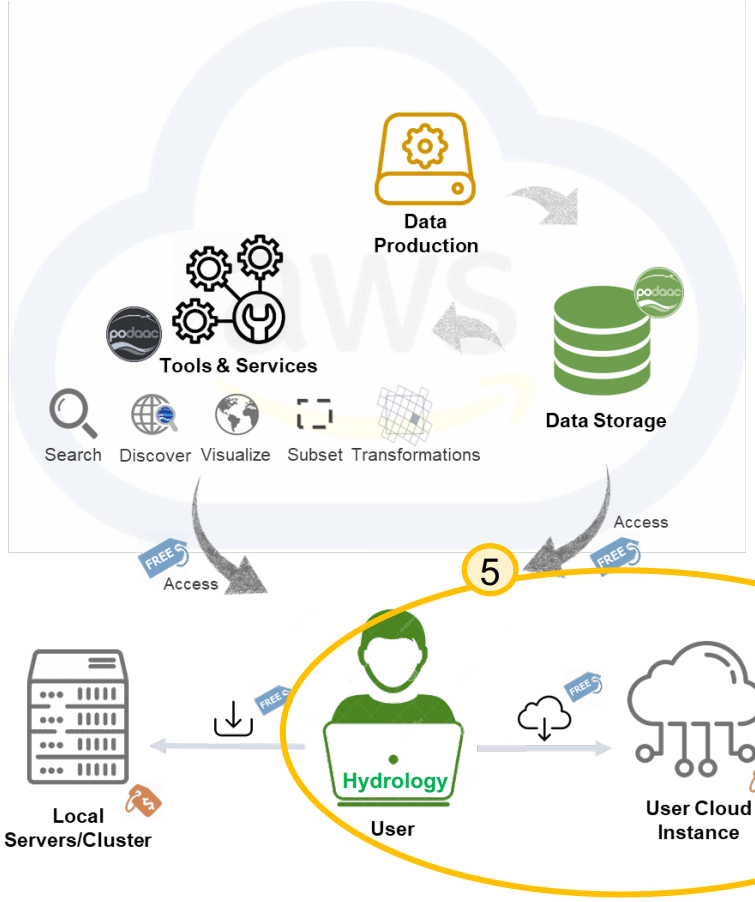




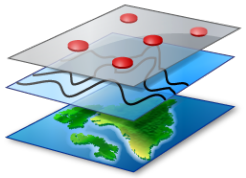
© 2019 California Institute of Technology. Government sponsorship acknowledged.

PO.DAAC/Earthdata Cloud: Hydrology User Workflow





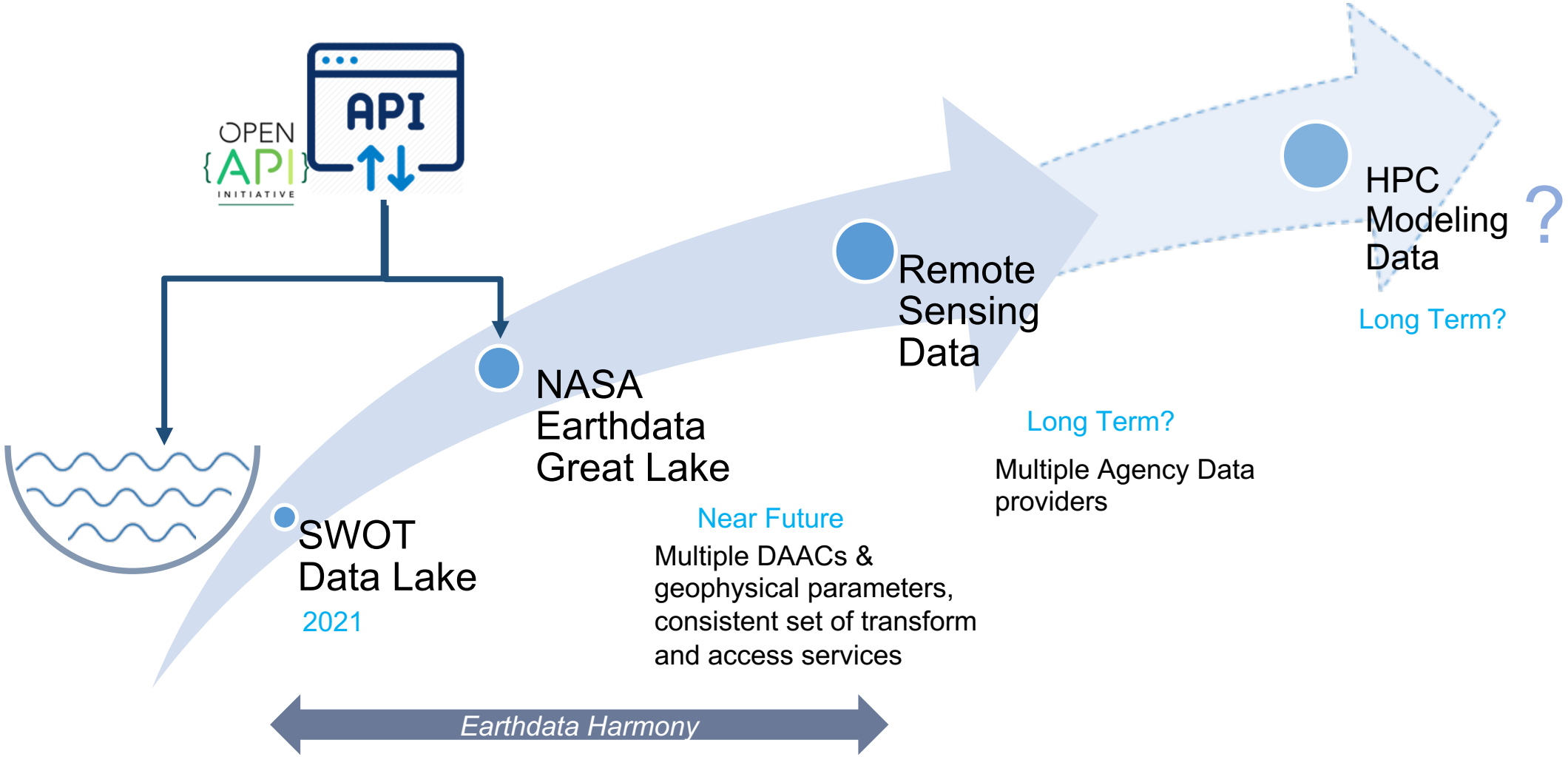
5 Access from User Cloud-based workspace, or View in GIS Client



© 2019 California Institute of Technology. Government sponsorship acknowledged.

Longer term Vision

Data Lake Inundation Timeline



© 2019 California Institute of Technology. Government sponsorship acknowledged.

Thank you

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

Catalina Oaida, Michael Gangl, Jessica Hausman, Michelle Gierach, Suresh Vannan

PO.DAAC, NASA Jet Propulsion Laboratory, California Institute of Technology

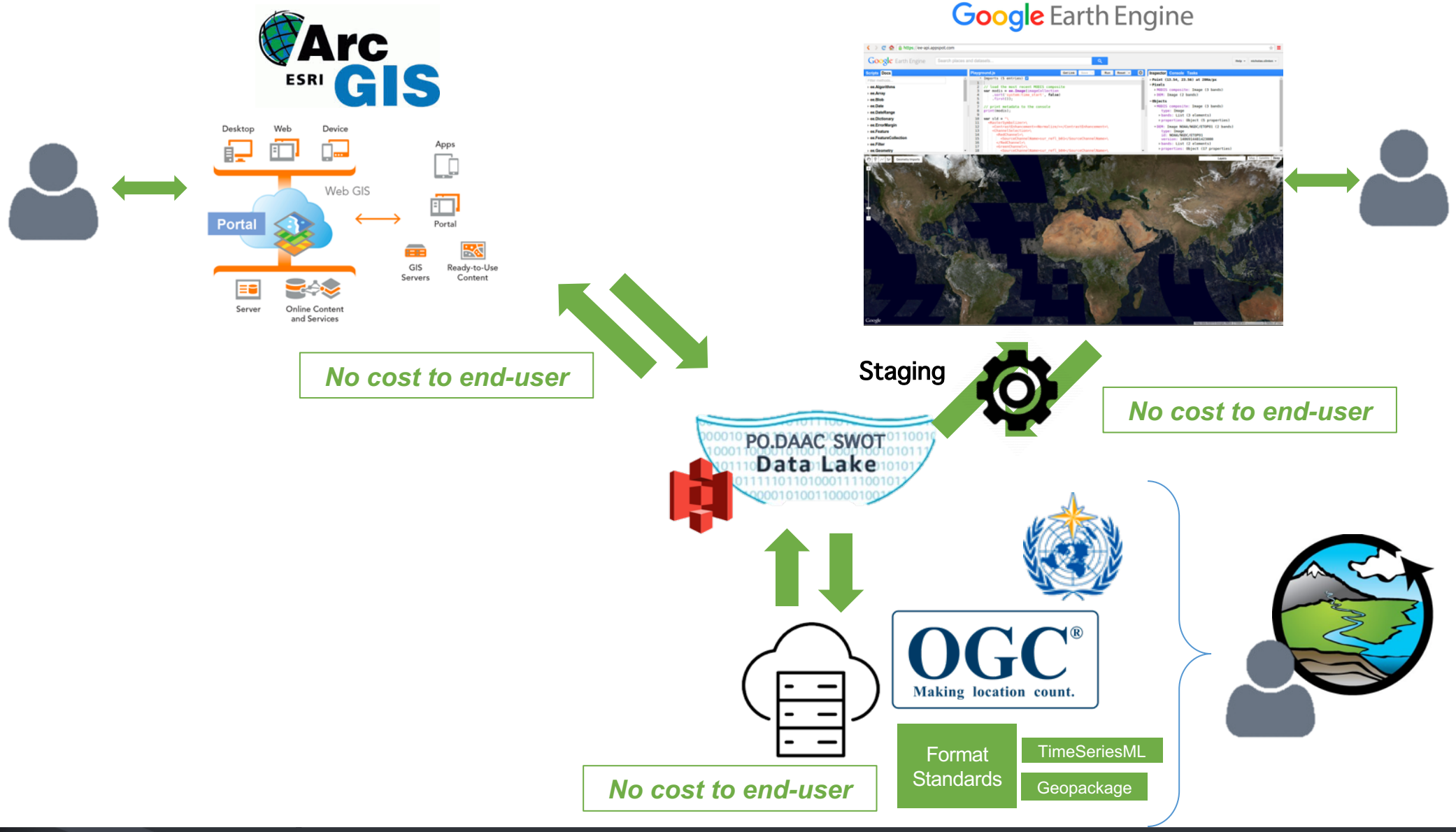
Contact: Catalina.Oaida@jpl.nasa.gov

© 2019 California Institute of Technology. Government sponsorship acknowledged.

Backup Slides

© 2019 California Institute of Technology. Government sponsorship acknowledged.

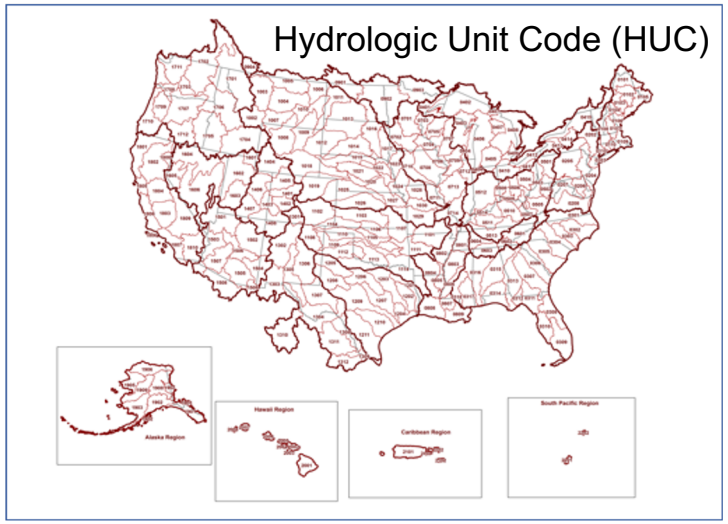
PO.DAAC/Earthdata Cloud: Interoperability

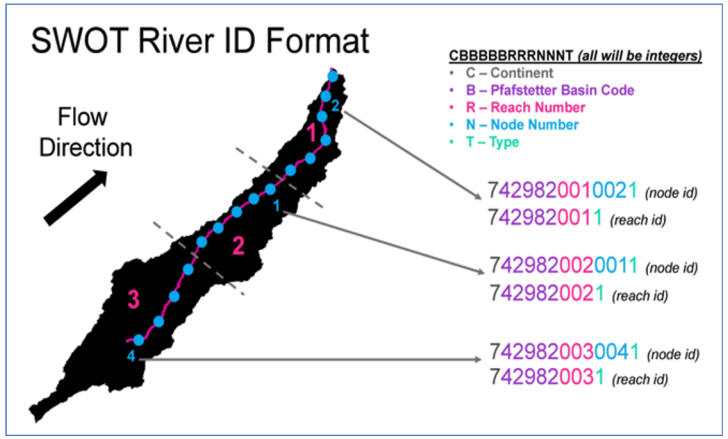


© 2019 California Institute of Technology. Government sponsorship acknowledged.

Feature Translation Service

GIS query of Features





OPEN {API} INITIATIVE

amazon web services

jupyter

ArcGIS

© 2019 California Institute of Technology. Government sponsorship acknowledged.

The Transition and Supporting the End-User

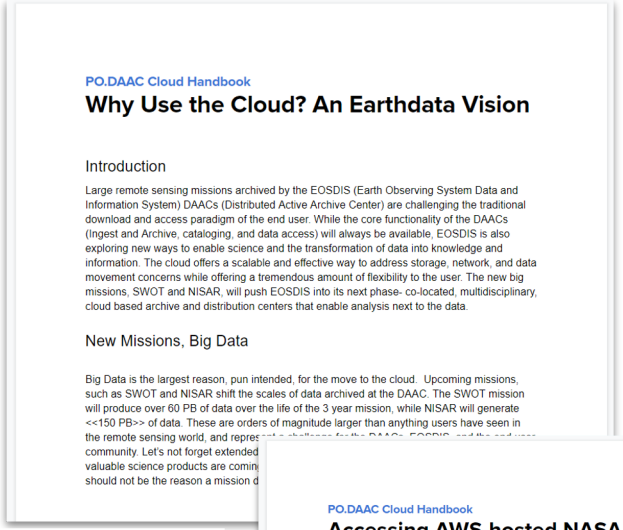
- Cloud EAP
- 5-6 datasets to be migrated to AWS cloud
- Cloud Primer
- Jupyter notebooks developed to assist with steps in workflow

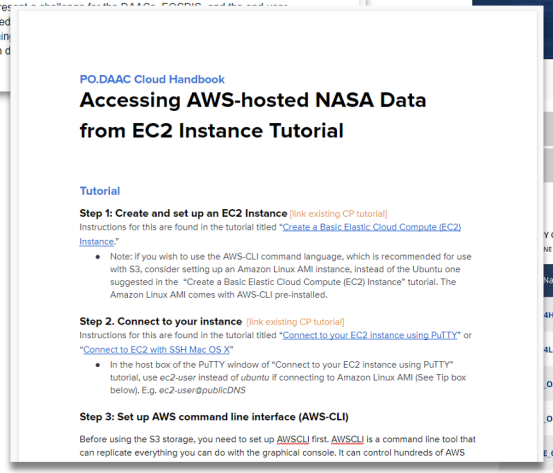
© 2019 California Institute of Technology. Government sponsorship acknowledged.

End-user Community Engagement

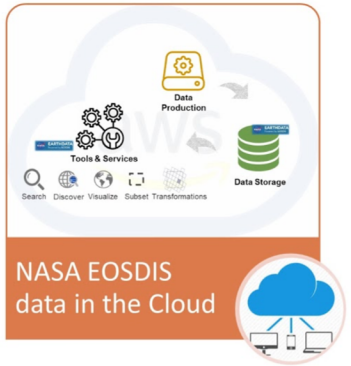
ESDIS Cloud Primer

Earthdata Dec 2019 – draft released

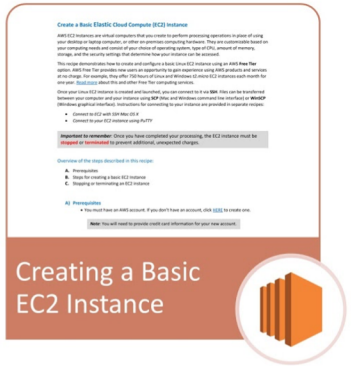




Why Use the Cloud?

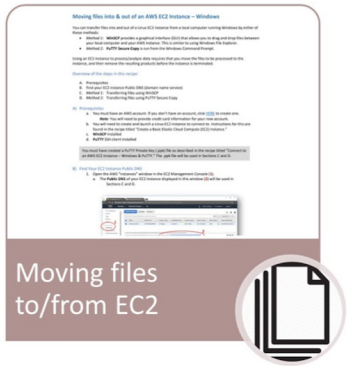


NASA EOSDIS data in the Cloud

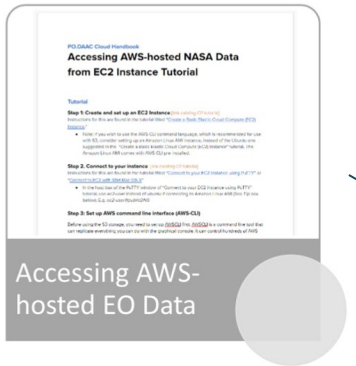


Creating a Basic EC2 Instance

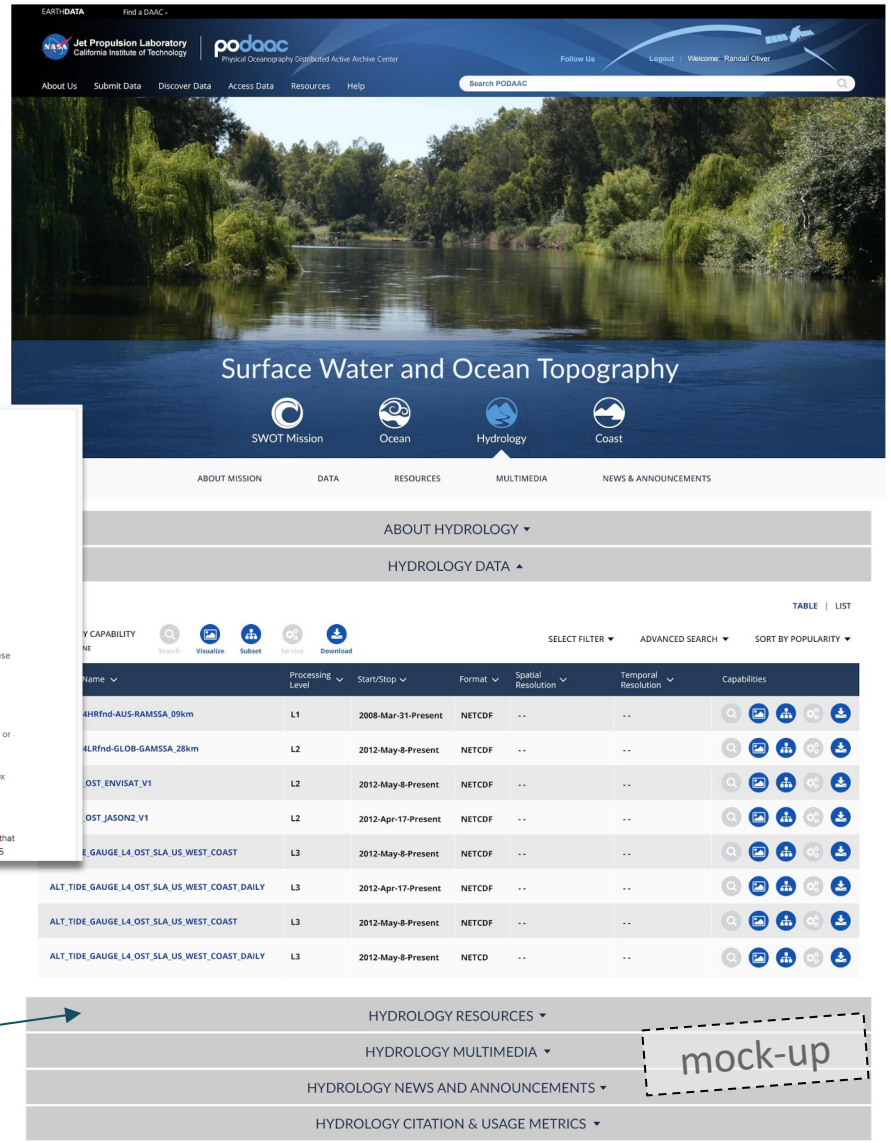
Connecting to Your Instance



Moving files to/from EC2



Accessing AWS-hosted EO Data



Webinars

Cloud Services Hackathon

Fall 2020

Resources & Services close to data

© 2019 California Institute of Technology. Government sponsorship acknowledged.

Focus on:

moving data processing and analysis, but not control, away from users and closer to the data.

* Highlight Services roadmap, including
    * Harmony
    * Access and ease of use
        * Infographic, user, data, cloud
        * New website
    * Data transformations
    * GIS transformations to enable data delivery directly in GIS**
* Current examples under dev
    * FTS
    * Access and download??
    * New L2SS in cloud??
    * Earthdata search by shp
* Examples of resources being dev
    * Jupyter notebooks
    * Data resources infographics
    * CP tutorials
* Backup slide: datasets to be available in cloud in 2020

Rough Outline

Steps we're taking to aid end-user

•Cloud EAP
•5-6 datasets to be migrated to AWS cloud
•Cloud Primer
•Jupyter notebooks developed to assist with steps in workflow
•Harmony - working with and across ESDIS to ensure cloud-based services being developed are relevant and useful for all future NASA data
•Vision: data lakes

© 2019 California Institute of Technology. Government sponsorship acknowledged.