





# PO.DAAC Cloud Early Adopters Kickoff Telecon

**April 2019** 

### Welcome!

#### Introductions

PO.DAAC Cloud - Science User Services

Catalina Oaida, PhD PO.DAAC Cloud User Services, Science User Services Engineer Raytheon Company Jet Propulsion Laboratory Pasadena, CA, USA Email: Catalina.Oaida@jpl.nasa.gov

Tel (office): (+1) 626-744-5558

Michelle Gierach, PhD

PO.DAAC Scientist. Science User Services Lead Jet Propulsion Laboratory Pasadena, CA, USA

Email: michelle.gierach@jpl.nasa.gov

Tel (office): (+1) 818-354-1933

Jessica Hausman

PO.DAAC Data Engineer Jet Propulsion Laboratory

Pasadena, CA, USA

Email: Jessica.K.Hausman@jpl.nasa.gov

Tel (office): (+1) 818-354-4588

Main point of contact

## Agenda

- Welcome & Introductions (20 min)
  - o PO.DAAC staff (3 min)
  - o EAs (17 min)
    - institution, community, level of cloud experience
    - what you hope to get from this EAP (why did you accept the invite)?
- Background (10 min)
  - PO.DAAC, ESDIS, Cloud Migration
  - SWOT mission, data products, timeline
  - PO.DAAC Cloud summary of activities
    - I&A, DSA Services + User Needs Assessment, GIS Plans
- PO.DAAC Cloud EA program overview + expectations (from TOR) (15 min)
  - Roles and Responsibilities
    - EAs
    - PO.DAAC
  - Roadmap: potential services, datasets
  - Communication/Collaboration Process
- Assign first tasks (5 min)
  - Create Earthdata login (if one does not exist)
  - Provide Earthdata login credentials to PO.DAAC for restricted Forum access
  - Fill out EA profile table
    - any current/planned work that is relevant to this EAP (Cloud, SWOT)
- **Q & A** (10 min)

#### Introductions

#### **Early Adopters**

Ocean Hydrology Coastal Informatics Novice Intermediate Advanced/dev -any/all welcomed!

homework

Name	Organization	Email	Community/ discipline (O,H,C,I)	Cloud Experience (novice, intermediate, advanced)	Current/Future work relevant to the PODAAC Cloud EAP (SWOT and/or Cloud efforts)
Ben Hamlington	Jet Propulsion Laboratory	benjamin.d.hamlington@jpl.nasa.gov			
Chelle Gentemann	Earth Space Research	cgentemann@esr.org			
Colin Gleason	University of Massachusetts Amherst	cjgleason@umass.edu			
Dan Ames	Brigham Young University	dan.ames@byu.edu			
Dave Blodgett	USGS	dblodgett@usgs.gov			
Ed Beighley	Northeastern University	r.beighley@northeastern.edu			
Faisal Hossain	University of Washington	fhossain@uw.edu			
Gregg Jacobs	NRL	gregg.jacobs@nrlssc.navy.mil			
John Wilkin	Rutgers University	jwilkin@rutgers.edu			
Mike Durand	Ohio State University	durand.8@osu.edu			
Ryan Abernathy	Columbia University, Lamont-Doherty	rpa@ldeo.columbia.edu			
Shih-Chieh Kao	DOE/ORNL	kaos@ornl.gov			
Tony Castronova	CUAHSI	acastronova@cuahsi.org			
Maikes Sonnewald	MIT	maike_s@mit.edu			

- PO.DAAC, EOSDIS, Cloud Migration
- SWOT mission, data products, timeline
- PO.DAAC Cloud summary of activities

## Background

## PO.DAAC, ESDIS, & Moving to the Cloud



- The mission of the Physical Oceanography Distributed Active Archive Center (PO.DAAC), <a href="https://podaac.jpl.nasa.gov">https://podaac.jpl.nasa.gov</a>), which is an element of the Earth Observing System Data and Information System (EOSDIS), is to preserve NASA's ocean and climate data and make these universally accessible and meaningful.
- PO.DAAC is the primary archive for the Surface Water and Ocean Topography (SWOT) mission.
  - Expected to launch in 2021, SWOT will be generating nearly 66 PB of data during the life of its mission (~20 TB/day),
     and will contribute datasets of great use to a wide variety of scientific and application communities.
- The new era of big data from Earth observations offers the opportunity to rethink the design of end-to-end data system architecture and services, particularly in a Cloud computing environment.
- PO.DAAC (and EOSDIS) is evolving to enable back- and front-end Cloud-based Archive & Data Services, including data storage, search and discovery, access, and analysis, in addition to the current, more established "download and analyze" paradigm. Performance, functionality, usability, stability, scalability, and interoperability are the foundational design elements for the PO.DAAC data tools and services.
  - This effort is leveraging the Earth Science Data Information System (ESDIS) cloud-based data management system.
  - This effort aims to enhance what the science and applications communities can do with the NASA Earth-observing data

## PO.DAAC Cloud Early Adopters - Program\* Goal

Ensure PO.DAAC-provided Cloud-based data and services meet the greater user community's expectations and needs as it adopts the Cloud as a platform for Earth data use.

\*This program is not limited to SWOT but will use the SWOT mission as a test case.

## Background: SWOT Mission

- The Surface Water and Ocean Topography (SWOT) mission aims to provide valuable data and information about the world's oceans and its terrestrial surface water such as lakes, rivers, and wetlands.
- SWOT is being developed jointly by NASA and Centre National D'Etudes Spatiales (CNES), with contributions from the Canadian Space Agency (CSA) and United Kingdom Space Agency (UKSA).
- The mission is targeted to launch September 2021.
- SWOT will measure ocean surface topography and land surface water elevation with great accuracy, using interferometry to achieve two-dimensional mapping.
- Observations from SWOT can be used to better understand ocean currents and processes
  happening at spatial scales on the order of 15-150 km, something that has not been done
  before.
- SWOT will enable high resolution (within 1 km from land) monitoring of **coastal** regions, including coastal currents, storm surges, and regional sea level change.
- On land, SWOT will provide **hydrology** measurements of water storage changes (surface water area and water depth) of major lakes, reservoirs, rivers, and wetlands, and support derived estimates of river discharge, which aid in assessing water resources.



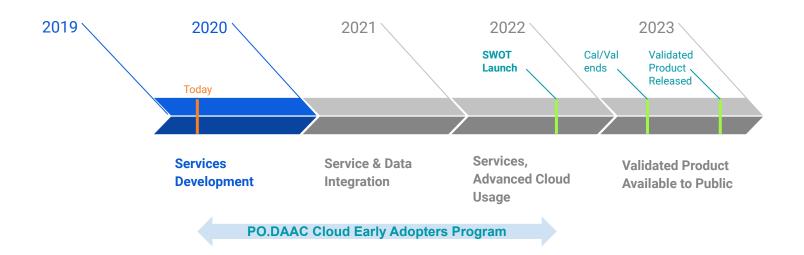
https://swot.jpl.nasa.gov/ https://podaac.ipl.nasa.gov/SWOT

## Background: SWOT Mission

#### **SWOT Data Products**

levelA	Data product	file type	pass/ cycle-based	description
L1	KaRIn SLC	netCDF4	pass	Interferomic single look complexes over land
L1	KaRIn interferogram	netCDF4	pass	Interferograms over the ocean
L2	Pixel Cloud	netCDF4	pass	water mask; inland + coastal; highest spatial res but much noisier than other hydro products
L2	Pixel Cloud River/Lake Vector	shapefile	pass	Pixel cloud vector attribute, geolocated river data
L2	Hydro River Vector	shapefile	pass	River data
L2	Hydro Lakes Vector	shapefile	pass	Lakes data
L2	Hydro Raster	netCDF4	pass	Raster product of lakes and rivers
L2	Cycle avg river vector prod	shapefile	cycle	21 day average of a specific river reach
L2	Cycle avg lake vector prod	netCDF4	cycle	21 day average of a specific lake/reservoir
L2	KaRIn oceans SSH	netCDF4	pass	Sea surface heights and significant wave heights from the KaRIn altimeter
L2	Nadir alt SSH	netCDF4	pass	Sea surface heights and significant wave heights from the nadir altimeter, similar to the Jason series datasets

## Background: PO.DAAC, SWOT, Cloud - Timeline



#### Summary of Activities:

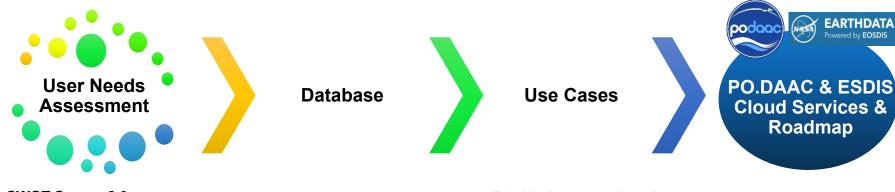
- 1. Ingest & Archive status
- 2. Discovery, Search, & Access (DSA) Services + Science User Services activities
- 3. GIS Plans

#### 1. Ingest & Archive status

- a. Ingest and Archive is proceeding well and is progressing as according to plan
- b. Ingest and Archive Risks have been identified and mitigated
- c. Performance stress testing currently support peak SWOT performance requirements
- d. Working on Disaster Recovery plan

2. DSA + User Services: a user-driven approach

Application Journeys & Survey → Services Roadmap → Implementation Plan



- **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

2. DSA + User Services: a user-driven approach

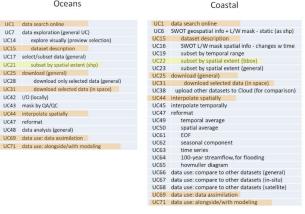
Application Journeys (user stories)

1. Identify common use cases across user personas

**Use Cases** → **Functionality** 

2. Compare and prioritize use cases, across user persona





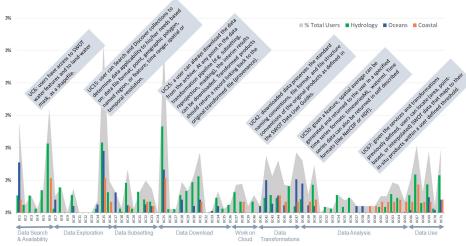


Fig 5. Percent of users (surveyed) impacted by given Use Case, by user persona (Hydrology, Oceans, Coastal). Highlights how different applications of data may have varying needs for Discovery, Search, and Access of data. Box inserts give examples of what requirements derived from the use cases would look like.

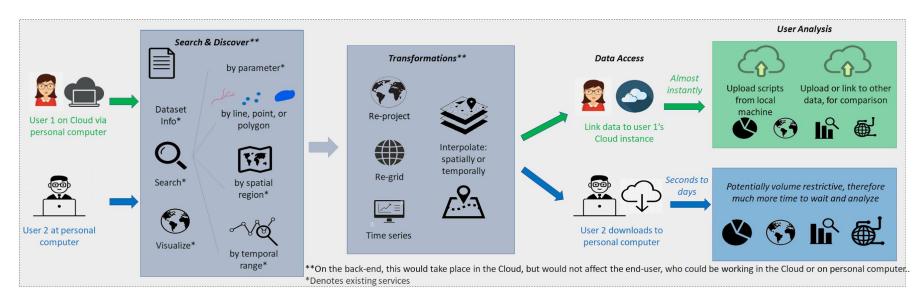
#### 2. DSA + User Services: a user-driven approach

#### Finalized Services Roadmap - May 2019

- List of functionalities
- Timeline: when beta version available for testing

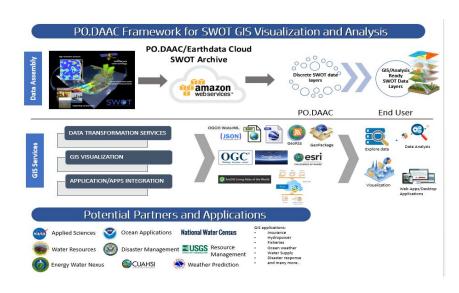
#### Preliminary Services Identified, through User Needs Assessment

- Discovery, Search, Transformations experience for the user would be the same whether user is on the Cloud or not.
- It is at the data Access point (wait time, bandwidth) where the difference arises.



#### 3. GIS Plans

- PO.DAAC will be supporting SWOT surface water datasets in GIS ready formats
- PODAAC to provide services to support Exploratory data Visualization and Analytics (EVA) of SWOT GIS datasets
- Integrate with common GIS software (ArcGIS, qGIS, etc), and partner with various communities (e.g. CUAHSI, OGC) to build services
- One archive format (.shp), many output/distribution formats (geopackage, netCDF, geoTIFF)
- On-demand hydrology focused raster generation



PO.DAAC's conceptual framework for storing, organizing, and accessing SWOT GIS datasets to enable EVA within the cloud environment based on anticipated community needs and big data approaches

# EA Program Overview

#### Why the PO.DAAC Cloud Early Adopters Program?

→ Ensure PO.DAAC-provided Cloud-based data and services meet the greater user community's expectations and needs as it adopts the Cloud as a platform for Earth data use.

#### Goals for the PO.DAAC Cloud Early Adopters Program

- 1. Support the development of appropriate PO.DAAC Cloud-based Services for use and adoption by the oceans, hydrology, and coastal communities.
- 2. Support the user community in the adoption and use of PO.DAAC Cloud-based Data holdings and Services.
- 3. Expand the PO.DAAC user community through:
  - a. Fostering a (SWOT-focused) community that is ready to adopt Cloud Computing when using Earth data;
  - b. Relationship development with the hydrology community, a new user community for PO.DAAC.

#### Roles and Responsibilities

#### Early Adopters will:

- Test and provide feedback on Cloud-based Services;
- Beta test the data pipeline into their own workflows;
- Test and provide feedback on training materials, e.g., tutorials/jupyter notebooks/recipes for discovering, searching, accessing, utilizing (SWOT) data on the Cloud;
- Co-develop application-specific training materials (e.g., jupyter notebook/recipes);
- Provide feedback on pain points regarding transitioning to the Cloud, e.g., barrier types: technical, institutional, knowledge, money, resources (e.g. IT);
- Suggest relevant datasets from their respective community of practice
  - For validation of, or comparison to, SWOT
  - Data that may complement SWOT data;
- Be ambassadors for PO.DAAC Cloud development activities.

#### Roles and Responsibilities

#### The **PO.DAAC** will provide:

- Defined boundaries for Early Adopters to operate within;
- Access to PO.DAAC Cloud-based Services being developed, e.g., search, access, download, subset;
- Access to simulated and/or complementary SWOT datasets\*\* (as made available).
- Training materials, including recipes/tutorials;
- Specifications, guidance or best practices on how to develop recipes that can be used with PO.DAAC data on the Cloud.

<sup>\*\*</sup>Note that the purpose of the datasets made available as part of the Early Adopters Program is solely to test PO.DAAC Services, and thereby may not be applicable for scientific endeavors at this time;

#### Services

Preliminary Services Identified

Finalized Services Roadmap - May 2019

- List of functionalities
- Timeline: when beta version available for testing

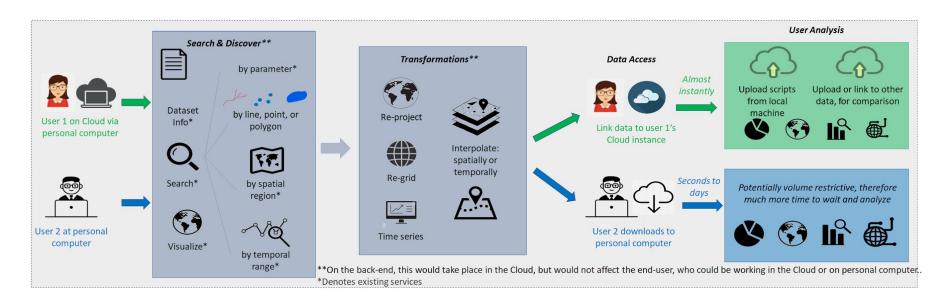
[ancillary prod] As a hydrologist, I want to have access to shapefiles for standard SWOT water features such as rivers or lakes.

[search] As a hydrologist, I want to search for SWOT inland water collections which intersect a specified SWOT water feature.

[subset] As a SWOT user, I want to subset my selected granules by user-defined shapefile.

[subset] As a SWOT user, I want to subset my selected granules by temporal range.

[on-demand analysis] As a hydrologist, I want to download spatially aggregated L2 hydrology time series data in WaterML format, so that I can easily work with existing user scripts.



#### **Datasets**

- MEaSUREs products available now
  - Hydrology [netCDF] (Dennis Lettenmaier)

Pre SWOT Hydrology Global Lake/Reservoir Water Height GREALM Pre SWOT Hydrology Global Lake/Reservoir Surface Water Area Extent Pre SWOT Hydrology GRRATS Jason-2 Virtual Station Heights Version 1

Oceans [netCDF] (Josh Willis / Brian Beckley)

Integrated Multi-Mission Ocean Altimeter Data for Climate Research Version 4.2
Integrated Multi-Mission Ocean Altimeter Data for Climate Research complete time series Version 4.2
Global Mean Sea Level Trend from Integrated Multi-Mission Ocean Altimeters Version 4.2

- Datasets in AWS available now
  - Landsat 8
  - MODIS
  - Sentinel-2
- Official SWOT Simulated data (as available)
  - Proxy Hydrology L2B pixel data [netCDF] anticipated Summer 2019 (TBC)
  - Proxy Hydrology L2B vector data [shp] anticipated Summer 2019 (TBC)
  - Proxy Hydrology L2B raster data [netCDF] anticipated Fall 2019 (TBC)
  - Proxy Oceans L2 data [netCDF] anticipated Summer 2019 (TBC)

Are there additional datasets you would like to have available (from other DAACs, modeled, etc) as part of this Program?

#### **Process**

Program Length: April 2019 - July 2021

Communication / Collaboration

- Quarterly Telecons as a group
  - o roughly April, July, October, January

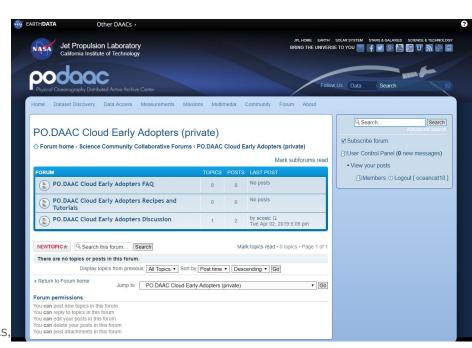
PO.DAAC Cloud Early Adopters **Forum** (restricted) <a href="https://podaac.jpl.nasa.gov/forum/">https://podaac.jpl.nasa.gov/forum/</a>

**FAQ**: quarterly telecon presentations, meeting notes, Terms of Reference, reports, etc.

**Recipes & Tutorials**: recipes, tutorials to be tested, comments specific to those recipes/tutorials

**Discussion**: communication & collaboration among EAs, and with PODAAC staff





## First 'homework' Due by next telecon "Early Aug 2019

- 1. Create an Earthdata account (optional, but might be useful regardless, down the road: ftp retirement!)
- Send PO.DAAC your Earthdata user name so you can be added to restricted Forum for collaboration/communication (hold off, collaboration process/platform being re-evaluated by PODAAC - TBD)
- 3. Fill out the Participant Table: relevant Work yes, please still fill out. Catalina will send table/doc in next couple of day after telecon
- 4. Next telecon "Early Aug 2019
  - Meanwhile, please email us should you have any questions, comments, concerns
    - Catalina Oaida <u>catalina.oaida@jpl.nasa.gov</u>



# PO.DAAC Cloud Early Adopters

Questions? Comments?

Thank you!

## Extras

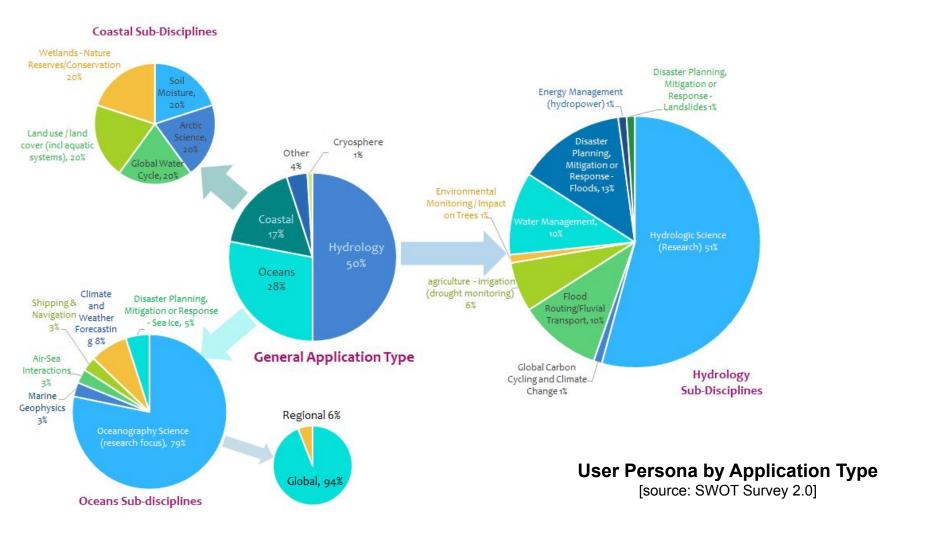
## Summary of Preliminary Services

- External search points to relevant data needed by user.
  - SWOT **portal resources** provide clear metadata, variable names, dimensions, temporal and spatial coverage info, file naming convention, file format and structure.
  - Ancillary data: SWOT water features with geospatial info (shp) available for download, and/or quick visualization (see prev point, spatial coverage)
- 2. **Query**, and return data to user based on those query parameters:
  - a. attribute (e.g. coordinated of river centerline)
  - b. SWOT Water Feature ID
  - c. HUC ID
  - d. commID
  - e. Data structure (file type)
  - f. File structure (e.g. CF)
  - g. QA/QC flag values
  - h. Spatial availability
    - i. Geometric shape: point, line, geographic polygon
    - ii. Named region or feature
  - i. Spatial resolution
  - j. Temporal availability
    - i. At time
    - ii. Time range
  - k. Temporal resolution
  - User provided
    - i. File with point-based site coordinates (e.g. stream gauges, tidal gauges, buoys)
    - ii. Shapefile (e.g. of watershed of interest)

- 3. **Transformations** here refer to any of the following:
  - subsetting (in space, time)
  - re-gridding/interpolation (in space)
  - interpolation/aggregation (in time)
  - reprojection
  - reformat
  - averaging, resulting in time series
    - over river reach(es)
    - o over area of interest

#### **Download**

- Whole granules
- Results of transformations: user can download at any point along the query/transformation pipeline
  - Downloaded data preserve the standard naming conventions, file format, and file structure conventions of the original product. Variables and metadata correspond to those in the original data products as defined in the SWOT Data User Guides.
- Offline, to local machine
- Push to user's AWS instance (down the road, to any Cloud computing platform)



ShortName	Product (1 or more files)	Shapefile metadata	Archive & Distrib'n metadata	PCM catalog metadata	Quick look image	Run Config File	Check-sum
LOA_LR_Packet	binary	-	Archive.xml	Yes	-	-	Yes
LOB_LR_Frame	netCDF	-	Archive.xml	Yes	-	Yes	Yes
L1B_LR_INTF	netCDF	-	ISO.xml	Yes	-	Yes	Yes
INT_LR_XOverCal	netCDF	-	Archive.xml	Yes	-	Yes	Yes
L2A_LR_NativeSSH	netCDF	-	ISO.xml	Yes	-	Yes	Yes
L2B_LR_FixedSSH	netCDF	-	ISO.xml	Yes	-	Yes	Yes
LOA_HR_Packet	binary		Archive.xml	Yes		•	Yes
LOB_HR_Frame	netCDF	•	Archive.xml	Yes		Yes	Yes
L1B_HR_SLC	netCDF	•	ISO.xml	Yes	Yes	Yes	Yes
L2_HR_PIXC	netCDF	-	ISO.xml	Yes	TBD	Yes	Yes
L2_HR_RiverSP	shapefile	Yes	ISO.xml	Yes	TBD	Yes	Yes
L2_HR_RiverAvg	shapefile	Yes	ISO.xml	Yes	TBD	Yes	Yes
L2_HR_LakeSP	shapefile	Yes	ISO.xml	Yes	TBD	Yes	Yes
L2_HR_PIXCVec	netCDF	-	-	Yes	TBD	-	Yes
L2_HR_LakeAvg	netCDF	Yes	ISO.xml	Yes	TBD	Yes	Yes
L2_HR_Raster	netCDF	-	ISO.xml	Yes	TBD	Yes	Yes
12 HR EPDEM	netCDF	_	ISO xml	Yes	TRD	Yes	Yes

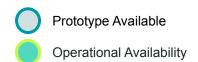
Product Short Name	Product Long Name	
L1B_LR_INTF	Level 1B KaRIn low rate nine beam interferogram product	
L2A_LR_NativeSSH	Level 2A KaRIn low rate native sampling sea surface height product	
L2B_LR_FixedSSH	Level 2B KaRIn low rate fixed sampling sea surface height product	
L1B_HR_SLC	Level 1B KaRIn high rate single look complex data product	
L2_HR_PIXC	Level 2 KaRIn high rate water mask pixel cloud product	
L2_HR_RiverSP	Level 2 KaRIn high rate river single pass vector product	
L2_HR_RiverAvg	Level 2 KaRIn high rate river average vector product	
L2_HR_LakeSP	Level 2 KaRIn high rate lake single pass vector product	
L2_HR_PIXCVec	Level 2 KaRIn high rate pixel cloud vector attribute product	
L2_HR_LakeAvg	Level 2 KaRIn high rate lake average vector product	
L2_HR_Raster	Level 2 KaRIn high rate raster product	
L2_HR_FPDEM	Level 2 KaRIn high rate floodplain DEM product	
L2_RAD_GDR	Level 2 Radiometer Geophysical Data Record	
L2_NALT_OGDR	Level 2 Nadir Altimeter Operational Geophysical Data Record	
L2_NALT_IGDR	Level 2 Nadir Altimeter Interim Geophysical Data Record	
L2_NALT_GDR	Level 2 Nadir Altimeter Geophysical Data Record	

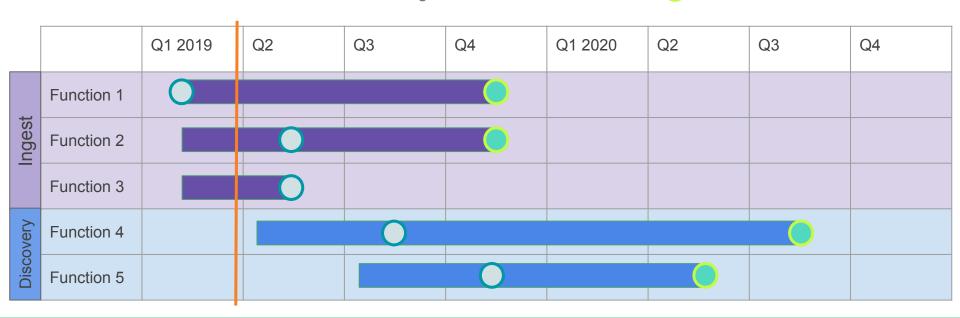
#### Services

Preliminary Services Identified

Finalized Services Roadmap - May 2019

- List of functionalities
- Timeline: when beta version available for testing





## Demo - Earthdata Search

https://earthdata.nasa.gov/

