maap-docs

Jul 20, 2023

Contents

CHAPTER 1

Getting Started

1.1 About the Multi-Mission Algorithm and Analysis Platform (MAAP)

The MAAP platform is designed to combine data, algorithms, and computational abilities for the processing and sharing of data related to NASA's GEDI, ESA's BIOMASS, and NASA/ISRO's NISAR missions. These missions generate vastly greater amounts of data than previous Earth observation missions. There are unique challenges to processing, storing, and sharing the relevant data due to the high data volume as well as with the data being collected from varied satellites, aircraft, and ground stations with different resolutions, coverages, and processing levels.

MAAP aims to address unique challenges by making it easier to discover and use biomass relevant data, integrating the data for comparison, analysis, evaluation, and generation. An algorithm development environment (ADE) is used to create repeatable and sharable science tools for the research community. The software is open source and adheres to ESA's and NASA's commitment to open data.

NASA and ESA are collaborating to further the interoperability of biomass relevant data and metadata. Tools have been developed to support a new approach to data stewardship and there is a data publication workflow for organizing and storing data and generating metadata to be discoverable in a cloud-based centralized location. The platform and data stewardship approaches are designed to ease barriers and promote collaboration between researchers, providers, curators, and experts across NASA and ESA.

This guide aims to help users get started with using the platform for searching, visualizing, accessing, processing, querying, and sharing biomass relevant data to the MAAP. These data, collected from satellites, aircraft, and ground stations, are organized into collections and granules. Collections are a grouping of files that share the same product specification. Granules are the individual files which are independently described, inventoried, and retrieved within a collection. Granules inherit additional attributes from their containing collection. Explanations of the various functions available in MAAP to use in the ADE will also be explored.

1.2 An overview of the MAAP platform

The MAAP is a cloud-based system to write science-analysis code and then run it at scale. This lets you keep all of the input and output data "in the cloud". It is composed of a few parts:


Jupyter
notebook
ADE (Algorit
Developme
Environmer

- The **Algorithm Development Environment (ADE)** is a tool that helps with the development of algorithms in a consistent, standardized environment that helps with the development and testing of algorithms and facilitates large scale data processing. MAAP's primary user interface is Jupyterlab, where code is written and tested before pushed to the large scale data processing system. Code is stored and checked out from Git-based repositories, including Github and MAAP's own code repository subsystem.
- The **Data Processing System (DPS)** is where registered algorithms (see Algorithm Catalog) can be run at scale in the cloud. The MAAP system provides a Jupyter GUI to run Jobs, or the maap.py library can be used to run a batch of Jobs in a loop using Python. The DPS also has monitoring capabilities, and again the MAAP system provides a Jupyter GUI to help monitor Jobs. This can also be done using maap.py in Python.
- The **Algorithm Catalog**, where your algorithms from the ADE can be registered and compiled for use by the DPS. The MAAP system provides API and GUI tools to help you register and view your algorithms.
- The **Code Repository** is a git-based repository to store user code. It is also used to store the configuration files necessary for building algorithms to store in the algorithm catalog and for execution in the DPS.
- Input data comes from a few **Data Catalogs**. Currently there is a MAAP STAC Catalog and the NASA CMR Catalog. More information can be found in the search tutorials section.

1.3 Setting up your account and workspace

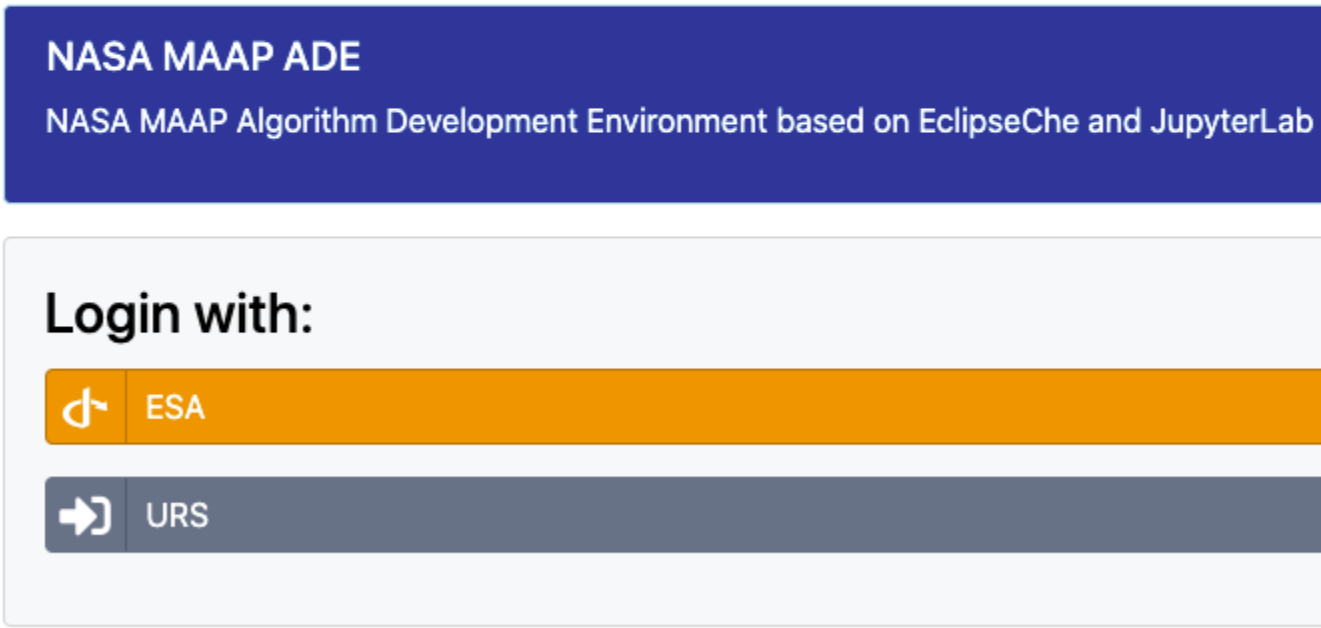
Learn how to sign up for an account and access the MAAP.

1.3.1 Signing up for an Earthdata Login account

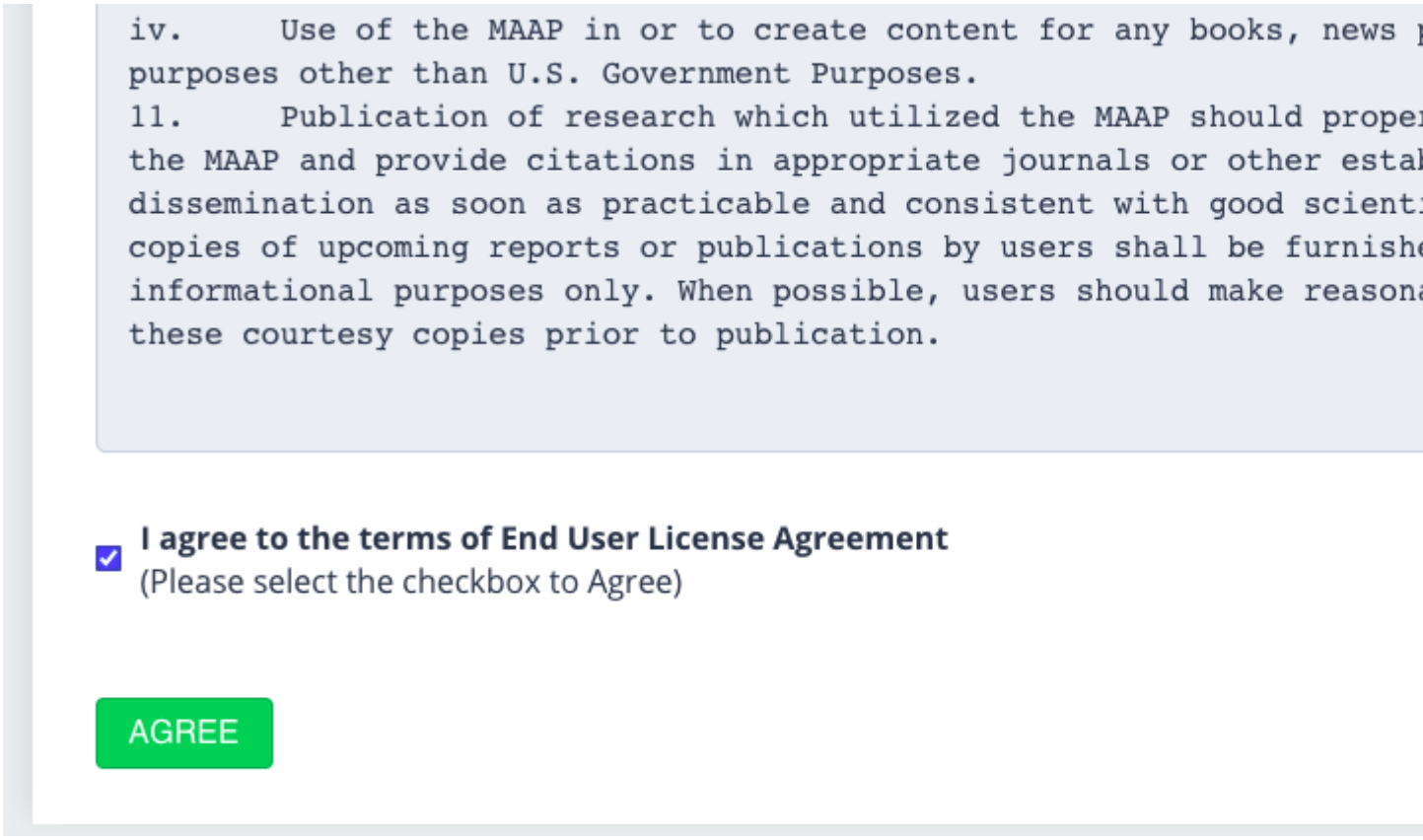
The MAAP offers accounts for NASA users through Earthdata Login. Before accessing the MAAP as a NASA user, you will need to create an Earthdata Login account. Anyone can register for an Earthdata Login profile here: https://urs.earthdata.nasa.gov/users/new.

1.3.2 Signing up for a new MAAP account

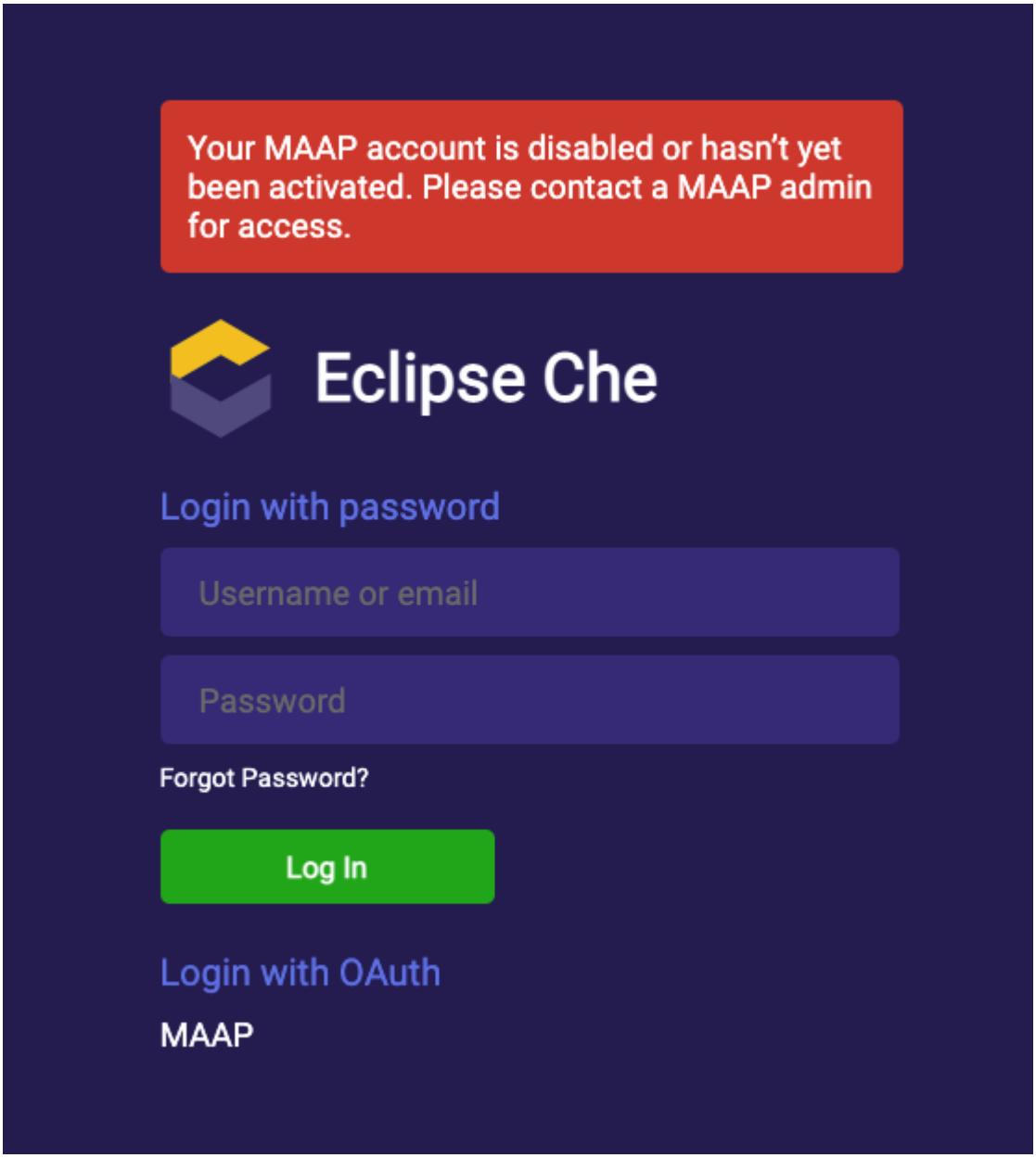
Once registered, you can register for a MAAP account by navigating to the MAAP ADE at http://ade.maap-project.org. On your first visit, select the "URS" login button shown here:



If this is your first visit to the MAAP, you will be asked to agree to the MAAP Terms of Use:



Once registered, you should be redirected back to the MAAP ADE showing a disabled account message similar to this:

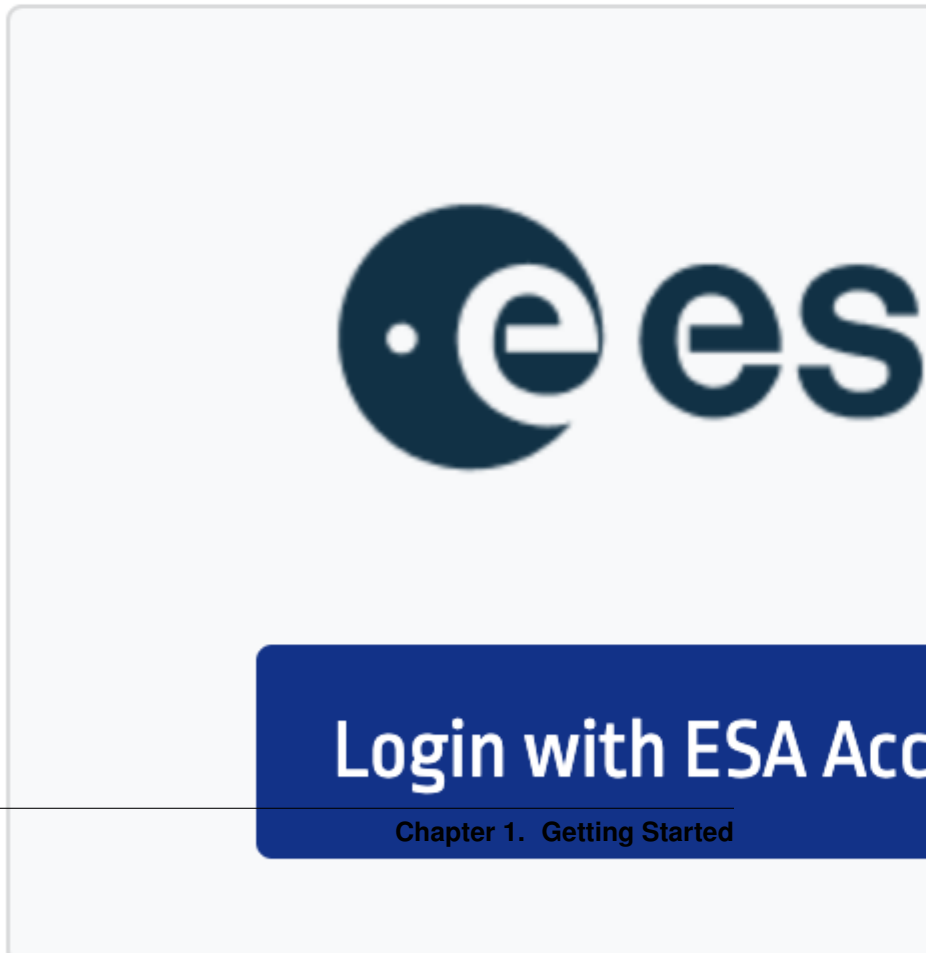


At this point, a MAAP administrator will approve your account, which will grant you access to the MAAP ADE. Access is only granted to known users in the biomass science community and other projects directly related to MAAP. To check on the status of your pending account, contact the MAAP team at support@maap-project.org.

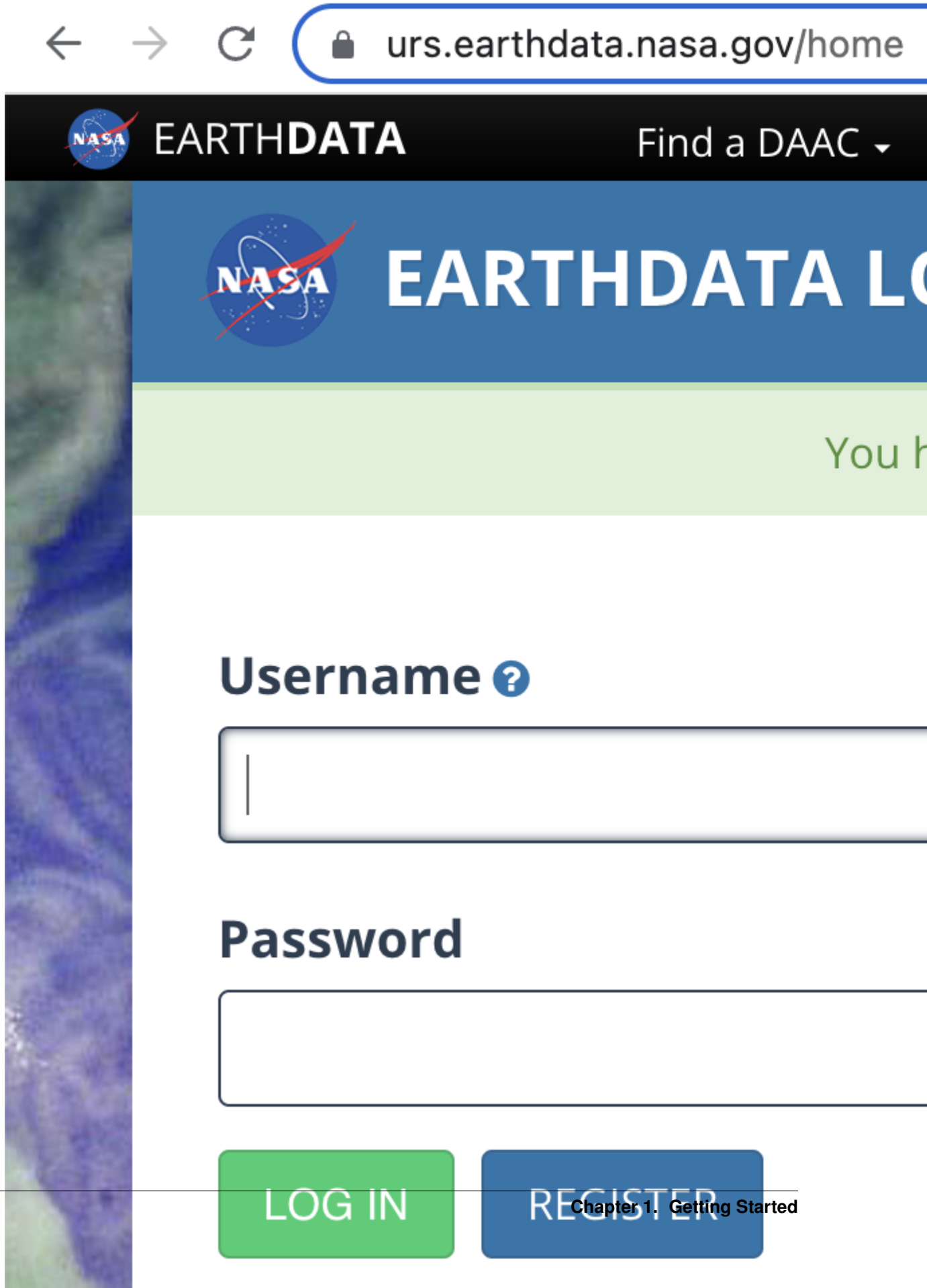
Note: Once your MAAP account is approved, you will receive an email notification using the address of your Earthdata Login account to let you know that your access is enabled.

1.3.3 Logging in

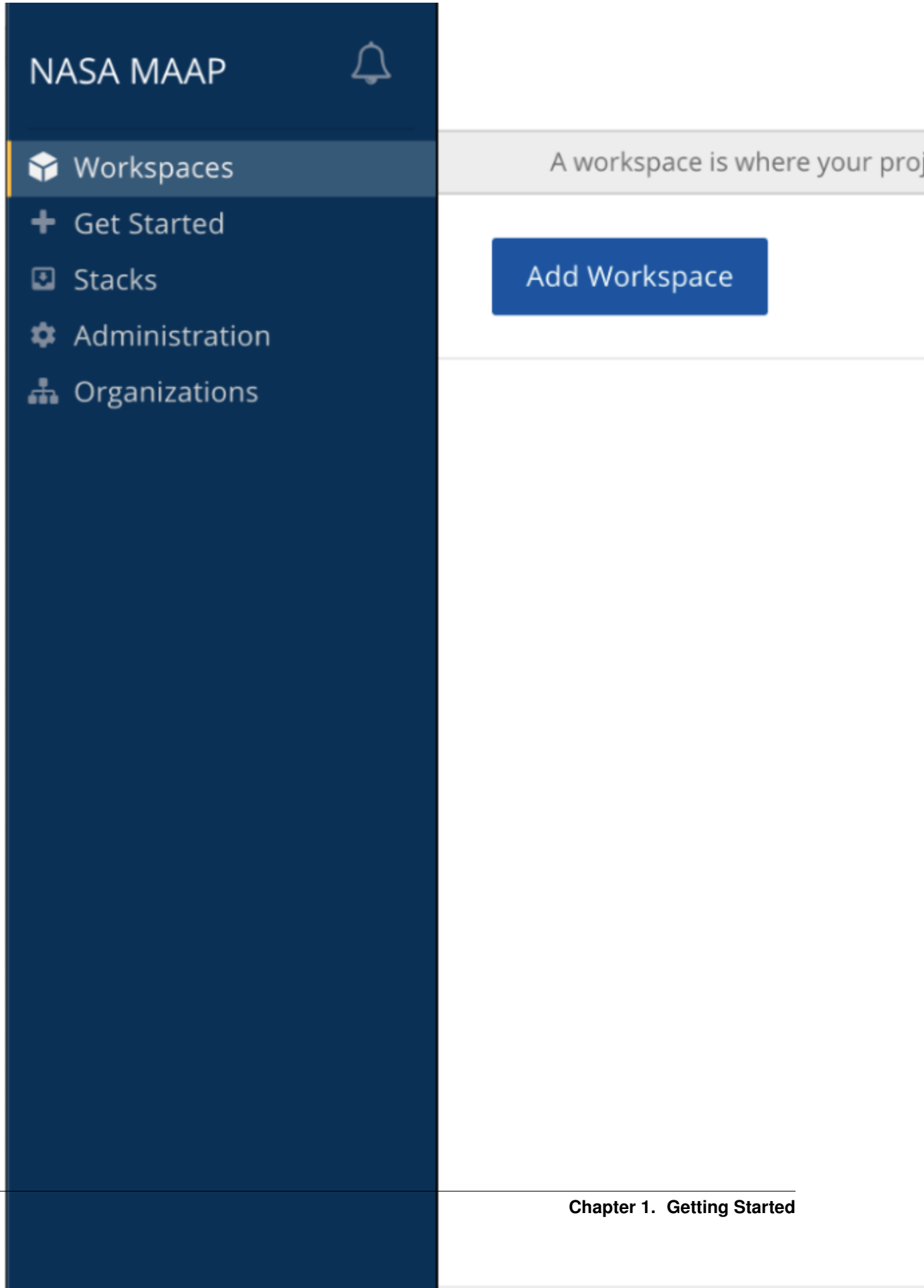
1. Navigate to https://ade.maap-project.org/ in Chrome or Firefox. You should be redirected to a page that looks



2. Click the "Login with EarthData Account" button. If this is your first time logging in, you should be redirected to an EarthData Login page that looks like this:


urs.earthdata.nasa.gov/home
EARTHDATA
Find a DAAC
EARTHDATA LO
You h
Username
Password
LOG IN
REGISTER

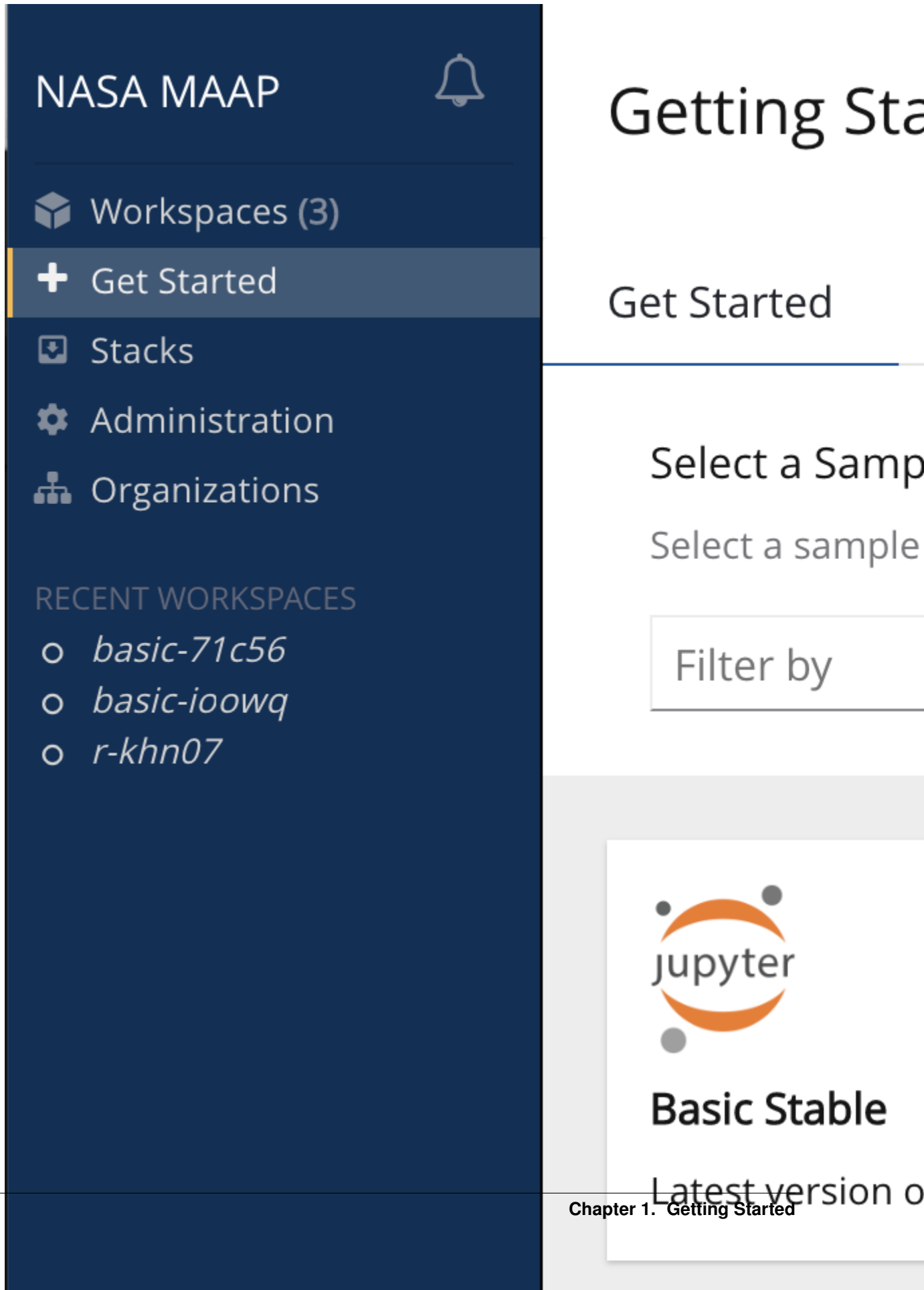
3. Enter your “EarthData Login” account credentials here and click “Log in”. You should see a temporary page that says “Redirecting”, followed by the MAAP showing your Workspaces (which will be empty to start):


NASA MAAP
Workspaces
Get Started
Stacks
Administration
Organizations
A workspace is where your proj
Add Workspace

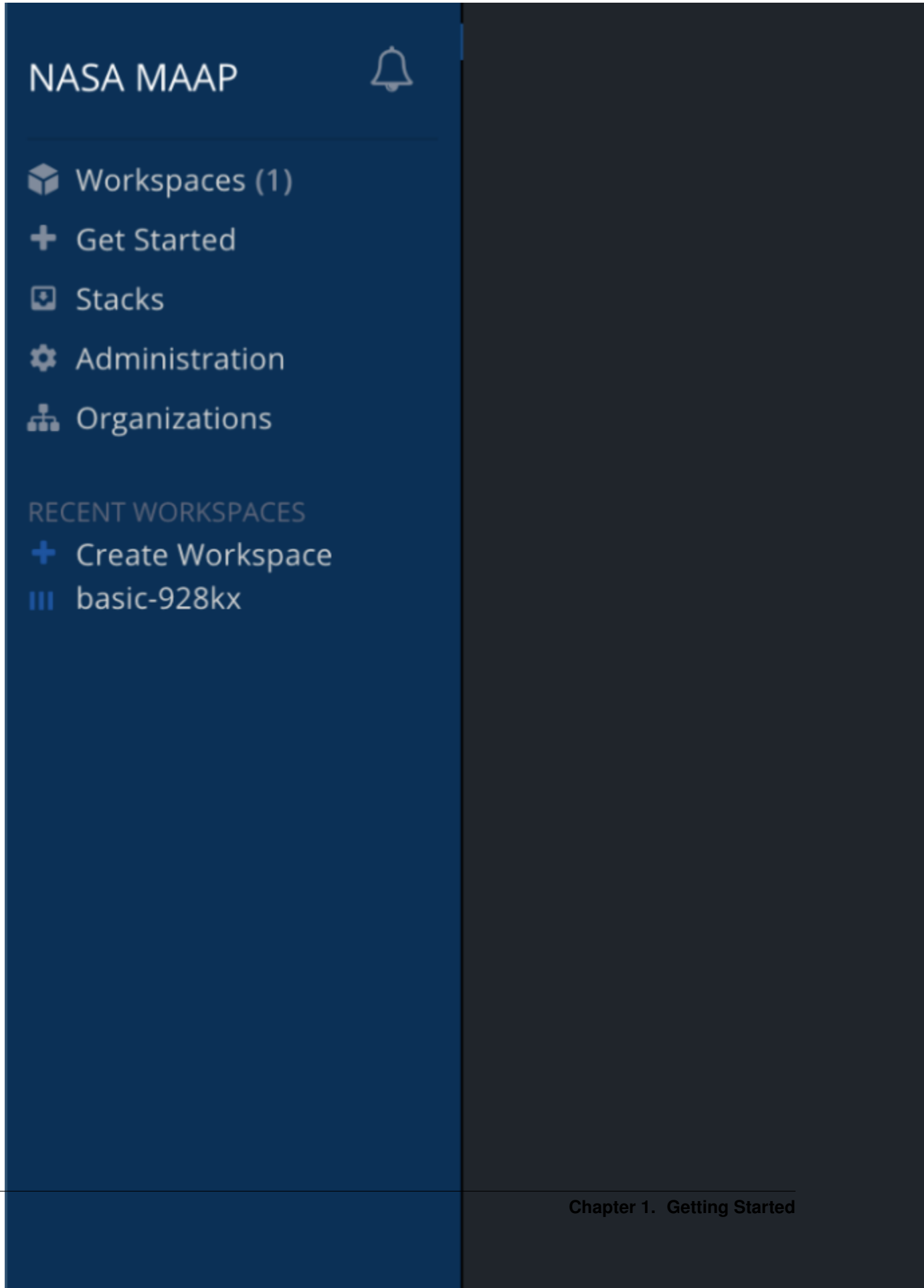
1.3.4 Creating a workspace

Workspaces are effectively a JupyterLab "computer in the cloud". To get started with Jupyter you need to create a workspace.

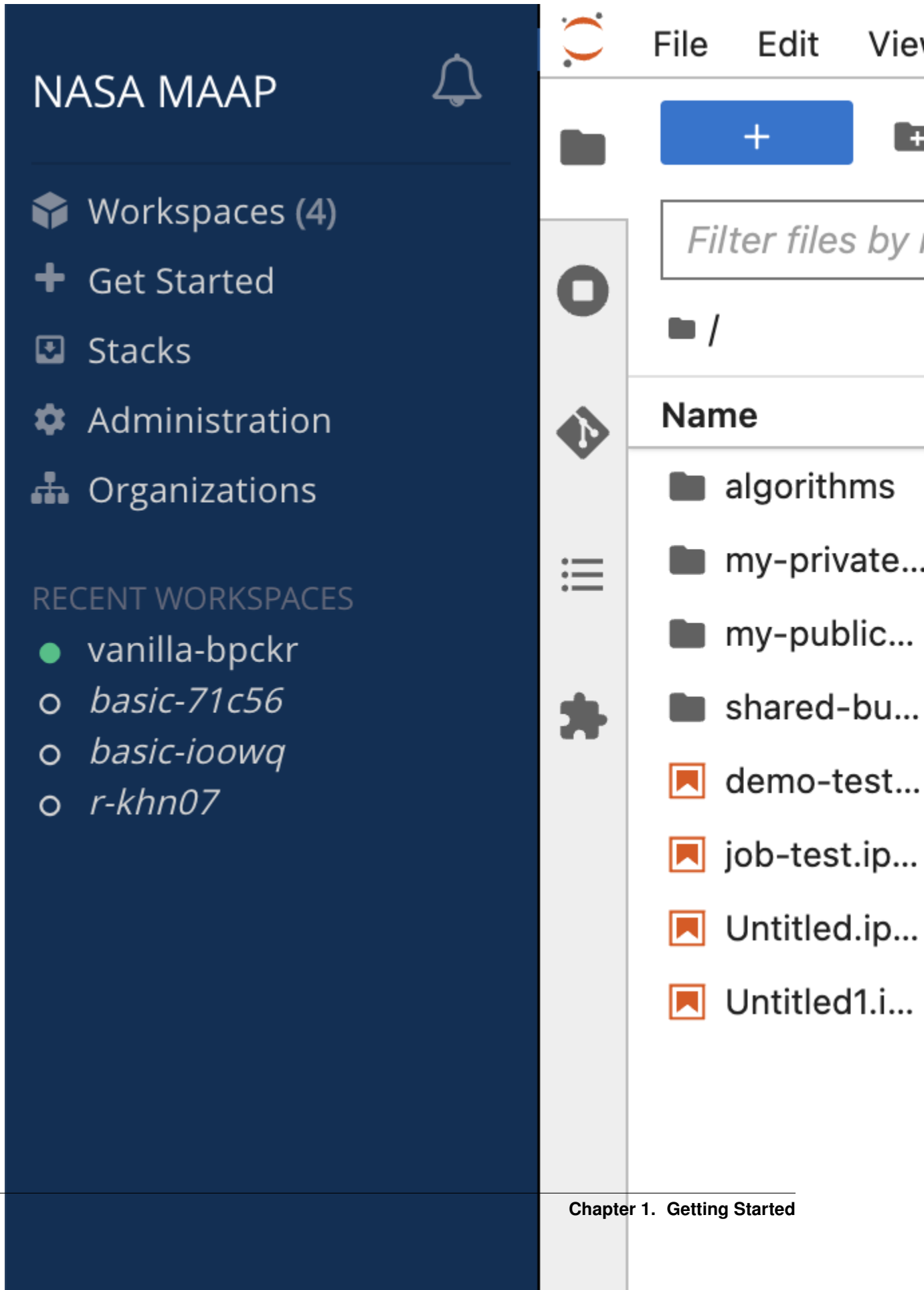
1. In the top-left corner of the MAAP dashboard, under "NASA MAAP", click "Get Started". You should see a menu that looks like this:


NASA MAAP
Workspaces (3)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
basic-71c56
basic-ioowq
r-khn07
Getting Sta
Get Started
Select a Samp
Select a sample
Filter by
jupyter
Basic Stable
Latest version o

2. Select “Basic Stable”. This is called a “Stack” and represents a type of cloud compute environment that will be set up. If you are interested in seeing more about each Stack, the adjacent link in the left-hand area is where you can see the configuration of each Stack in detail. After choosing “Basic Stable”, you will see a loading screen that looks like this – wait for it to finish loading.

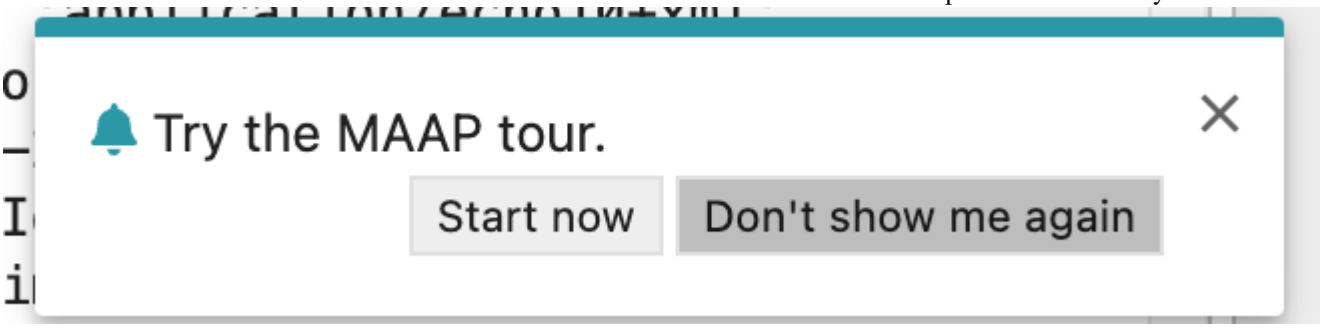

NASA MAAP
Workspaces (1)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
Create Workspace
basic-928kx

3. Once the workspace has loaded, you should see a Jupyter interface that looks like this (note: You will see fewer environments and items in your root directory — this is normal! You may also see some notifications in the bottom right that look like errors about SSH Keys and other things; that is normal as well. You will also see one asking if you would like to take a guided tour.).


NASA MAAP
Workspaces (4)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
vanilla-bpckr
basic-71c56
basic-ioowq
r-khn07
File
Edit
Filter files by
/
Name
algorithms
my-private...
my-public...
shared-bu...
demo-test...
job-test.ip...
Untitled.ip...
Untitled1.i...

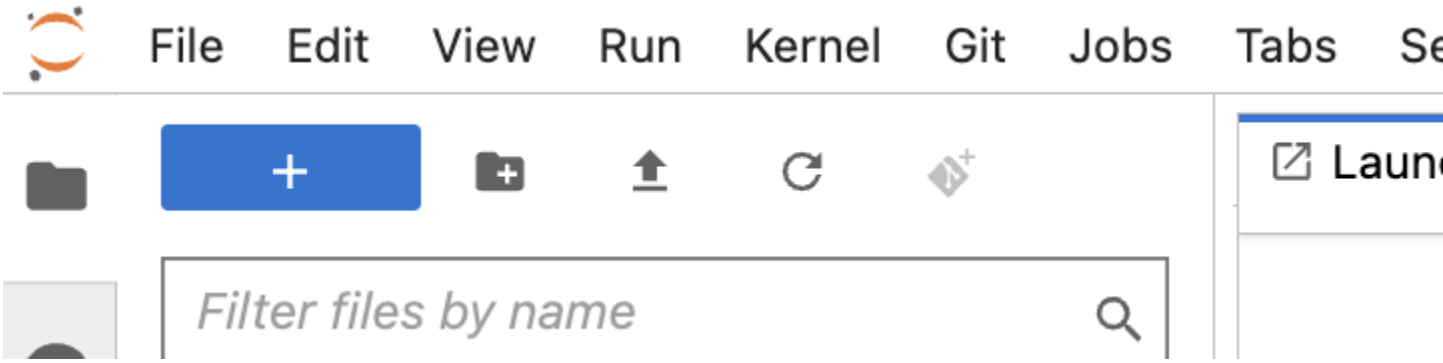
1.3.5 Jupyter Interface overview

When you first log in you may see a notification in the bottom right about a guided tour. Feel free to view the tour, which will give you a quick overview of the Jupyter user interface. You can also find the MAAP Tour in the Help menu at any time.



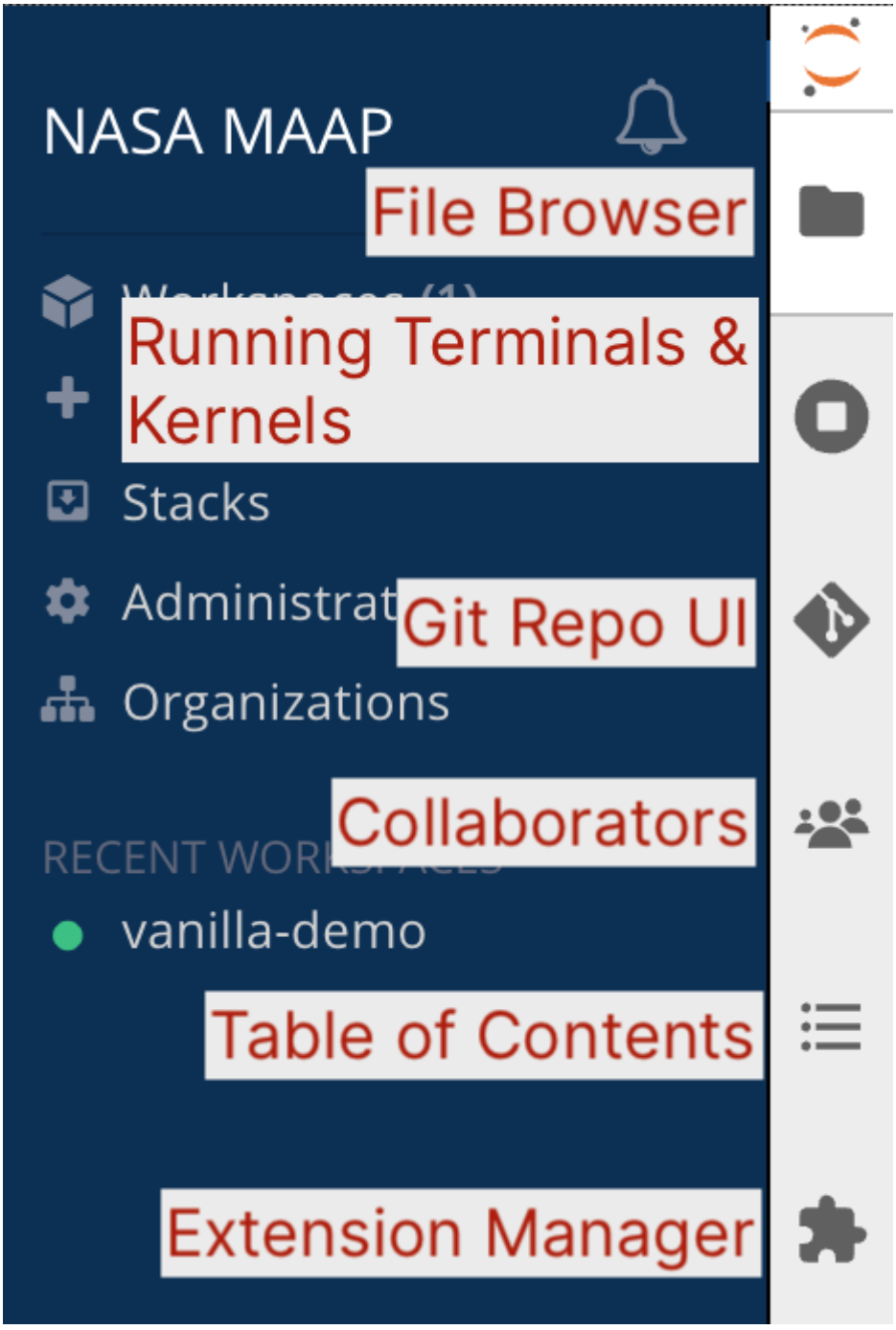
In addition to typical JupyterLab menu bar and sidebar configuration:

MAAP Jupyter Menus



- Git: Open repo in terminal, init, or clone repo.
- Jobs: Users may submit jobs through the submit tab and view their jobs through the view tab.
- Help: The help menu has several customized extensions and references to the MAAP documentation.

MAAP Jupyter Sidebar



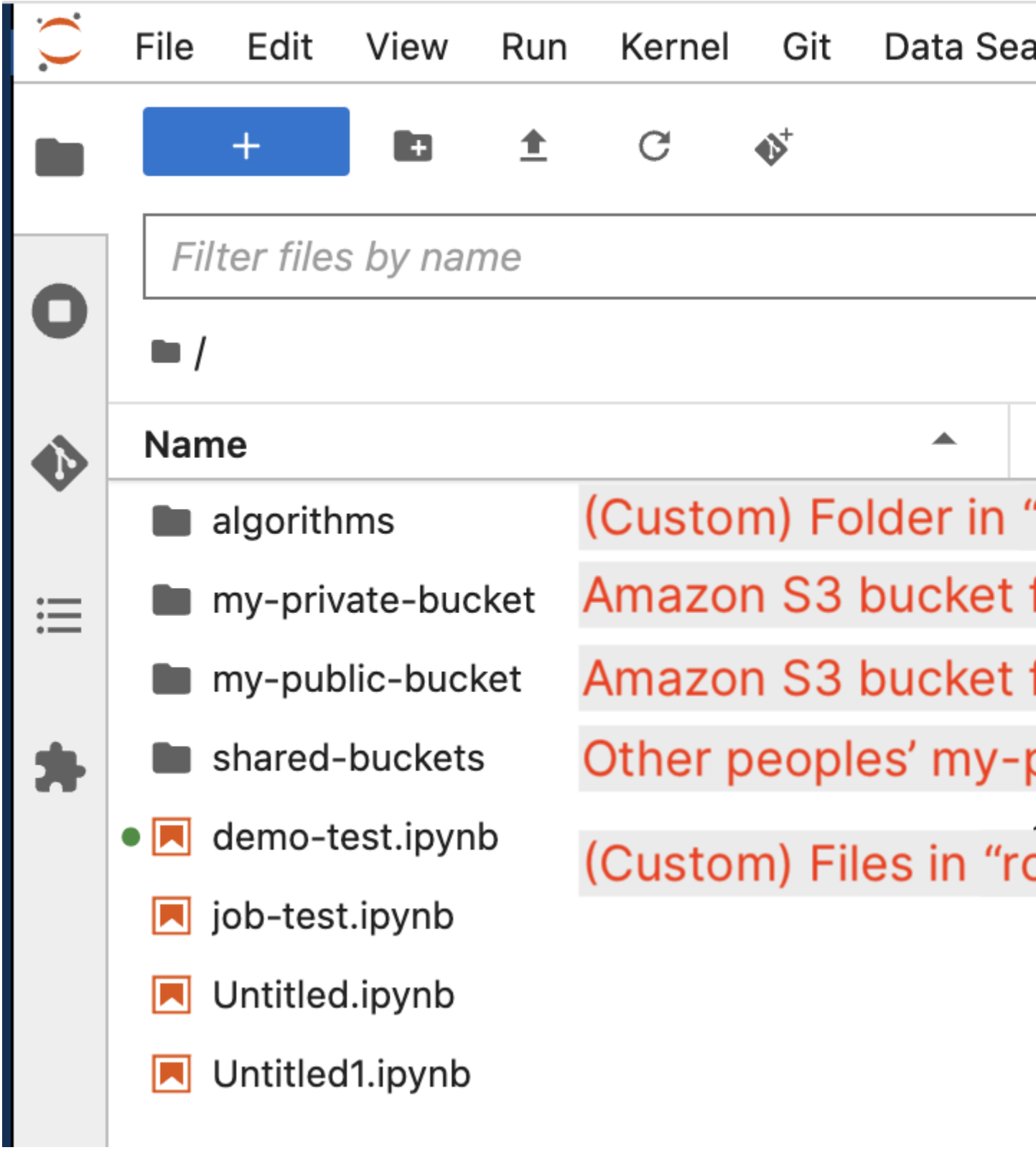
- File Browser
- Running Terminals & Kernels
- Git Repo Interface (if this folder is a Git repo)
- Collaborators
- Table of Contents

- Extension Manager

MAAP Blue Sidebar

- Workspaces: See workspaces, share them, as well as configure settings
- Stacks: See available platforms for workspaces & required memory
- Administration: Control the configuration & policies for your installation.
- Organizations: allow groups of developers to collaborate with private & shared workspaces. Resources & permissions are controlled & allocated within the organization by admin.
- Profile (bottom, labeled with your name): See account info, logout

1.3.6 MAAP Storage Options



My root folder (fast cloud storage)

- Your Jupyter home directory (`~`) is mounted to `/projects`. Files in here persist across sessions and exist across your workspaces.
  - Use this for code-related items, smaller data storage
  - Git is more likely to behave predictably here compared to other storage
  - This is also the place to make persistent conda environments (covered in another section), but make sure to not make a conda env inside a git-tracked folder, or if you do add it to the .gitignore. If git is tracking an env, it could cause your workspace to crash.
- Uses local (to Jupyter) file system; generally faster and more reliable for "normal" file operations, but expensive

Large file storage: my-private-bucket

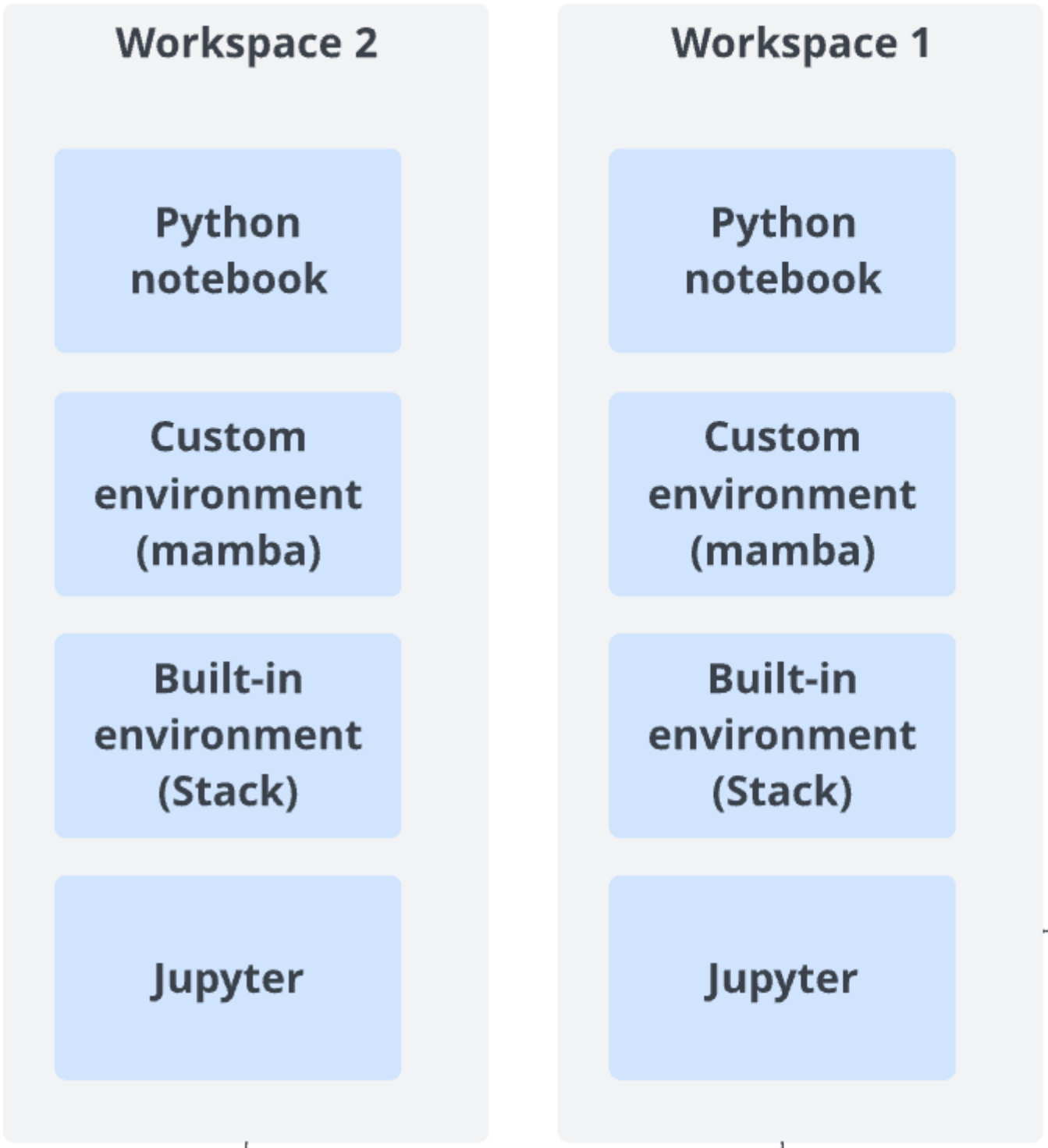
`~/my-private-bucket` is an S3 bucket with persistent storage, but accessible only to you and others in a shared workspace.

- Use for large data storage
- It will be slower than the root folder to copy and move files, which is why it is not ideal for storing smaller files that need to be read or written quickly

Sharing files: my-public-bucket and shared-buckets

`~/my-public-bucket` is an S3 bucket with persistent storage. It is the same as `~/shared-buckets/<my_username>/` — anything you put in here will be accessible to other users via `~/shared-buckets/<my_username>` as a read-only file. Likewise, to find shared files from another user, look in `~/shared-buckets/<their_username>`.

- Use for large data storage for files that you want to share across workspaces


Workspace 2
Python notebook
Custom environment (mamba)
Built-in environment (Stack)
Jupyter
Workspace 1
Python notebook
Custom environment (mamba)
Built-in environment (Stack)
Jupyter
read-write
read-write
"Root" files in

Mounting your MAAP workspace on your local computer

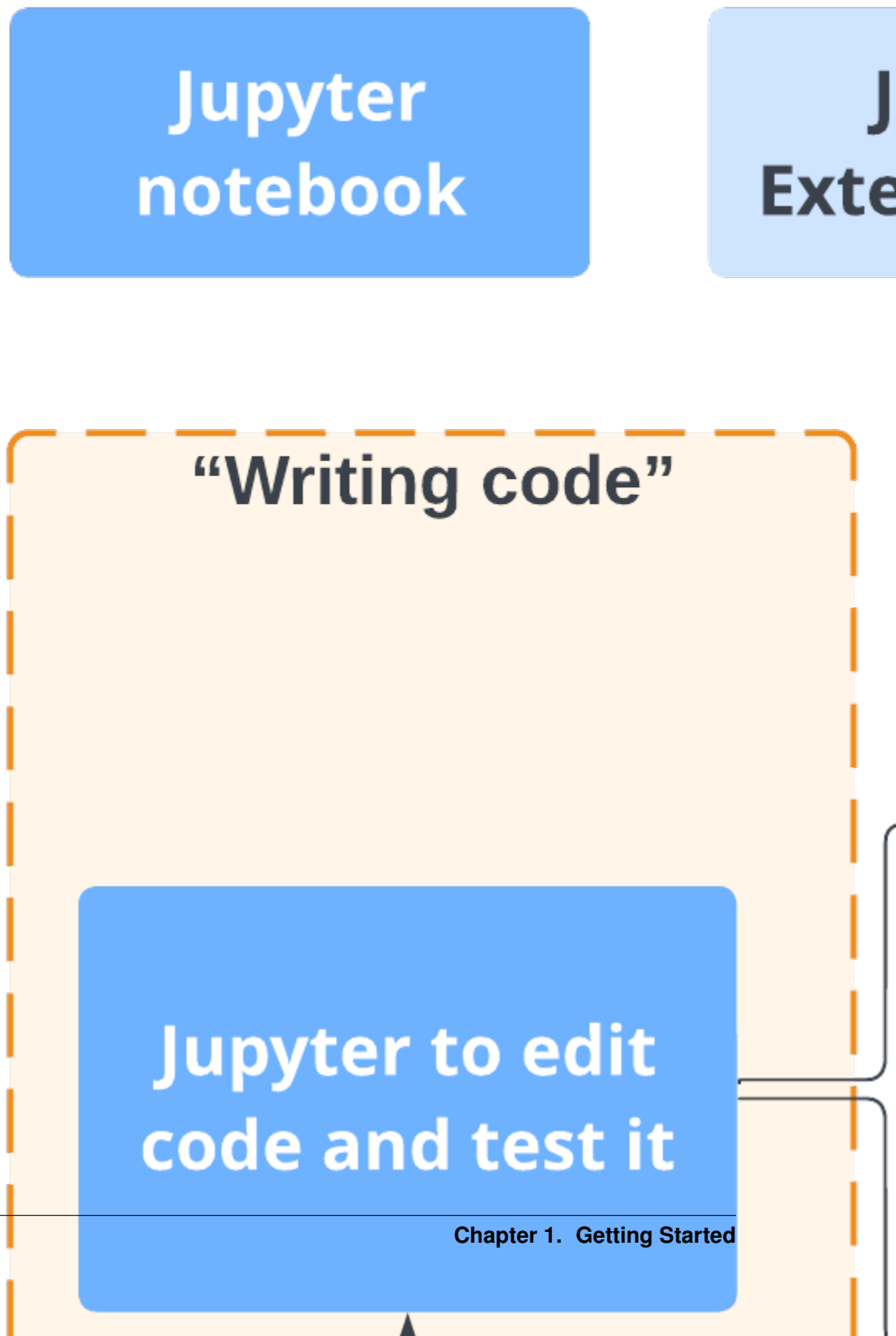
If you prefer to work on your local computer, or to drag-and-drop copy files from your computer to/from MAAP, you access the workspace via SSH. The *process for doing this* is in the system guide.

1.4 Writing code with MAAP

Writing and editing code in the MAAP is done in a Jupyter workspace. To assist connections to the MAAP system from a Jupyter notebook, a helper library called `maap.py` provides Python-native calls to the underlying MAAP API.

Code is version-controlled using git, which may be GitHub or MAAP GitLab. Jupyter also provides a GUI widget to help with code push/pull as a sidebar tool. Git is intended to help with collaborative code development; Algorithms run at scale in the MAAP must first be *pushed to the MAAP GitLab* in order to facilitate registration.

If you have not used Github or git before, it is highly recommended that you get acquainted with it. For a quick reference to git commands there is a Git Cheat Sheet in a variety of languages.


Jupyter notebook
J
Exte
“Writing code”
Jupyter to edit code and test it

1.4.1 Using maap.py to access MAAP functionality from Python notebooks

The MAAP platform offers a variety of functionality. Access to the functionality is gained via the underlying MAAP API. In a Python notebook, you will typically use this API via a helper library called `maap.py`, which will make using MAAP platform features easy, using Python syntax. For example, registering algorithms, running batches of jobs, monitoring jobs, or accessing data.

Much of the `maap.py` functionality is documented in the *Technical Tutorials section* and in-context in the *Science Tutorials*. The maap-py Github page has additional usage documentation.

1.4.2 Helpful Templates while developing Algorithms in MAAP

- This algorithm repository example is a good starting point for a new algorithm, as it contains the various accessory files that facilitate running the algorithm at scale
- Which templates will help you? Let the development or documentation team know!
- For example: conda.yml with some default packages, run_script.sh

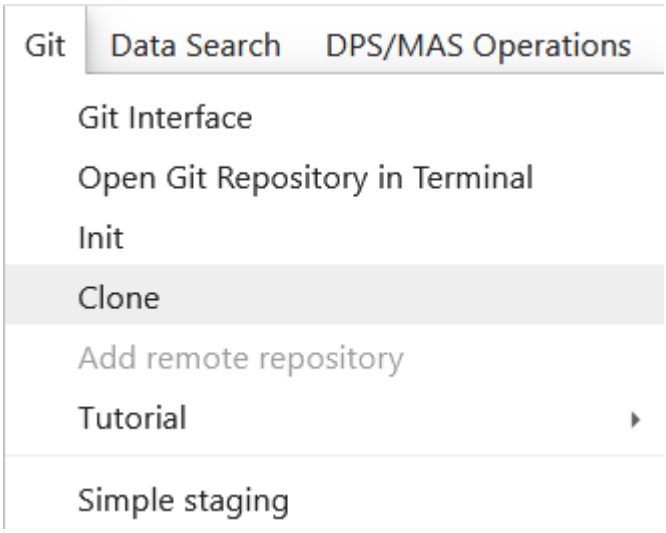
1.4.3 Working with code repositories like GitHub and GitLab

The MAAP GitLab Code repository

After creating your MAAP account, you can create a code repository by navigating to the MAAP GitLab account at https://repo.maap-project.org. This GitLab account is connected to your ADE workspaces automatically when signing into the ADE.

This example walks through cloning a repository into the ADE. Cloning a repository allows you to open, edit, and run files contained within the cloned repository. In this example, we look at cloning the "MAAP-Project/maap-documentation" Git repository, so that you are able to experiment with the code examples contained within this user documentation.

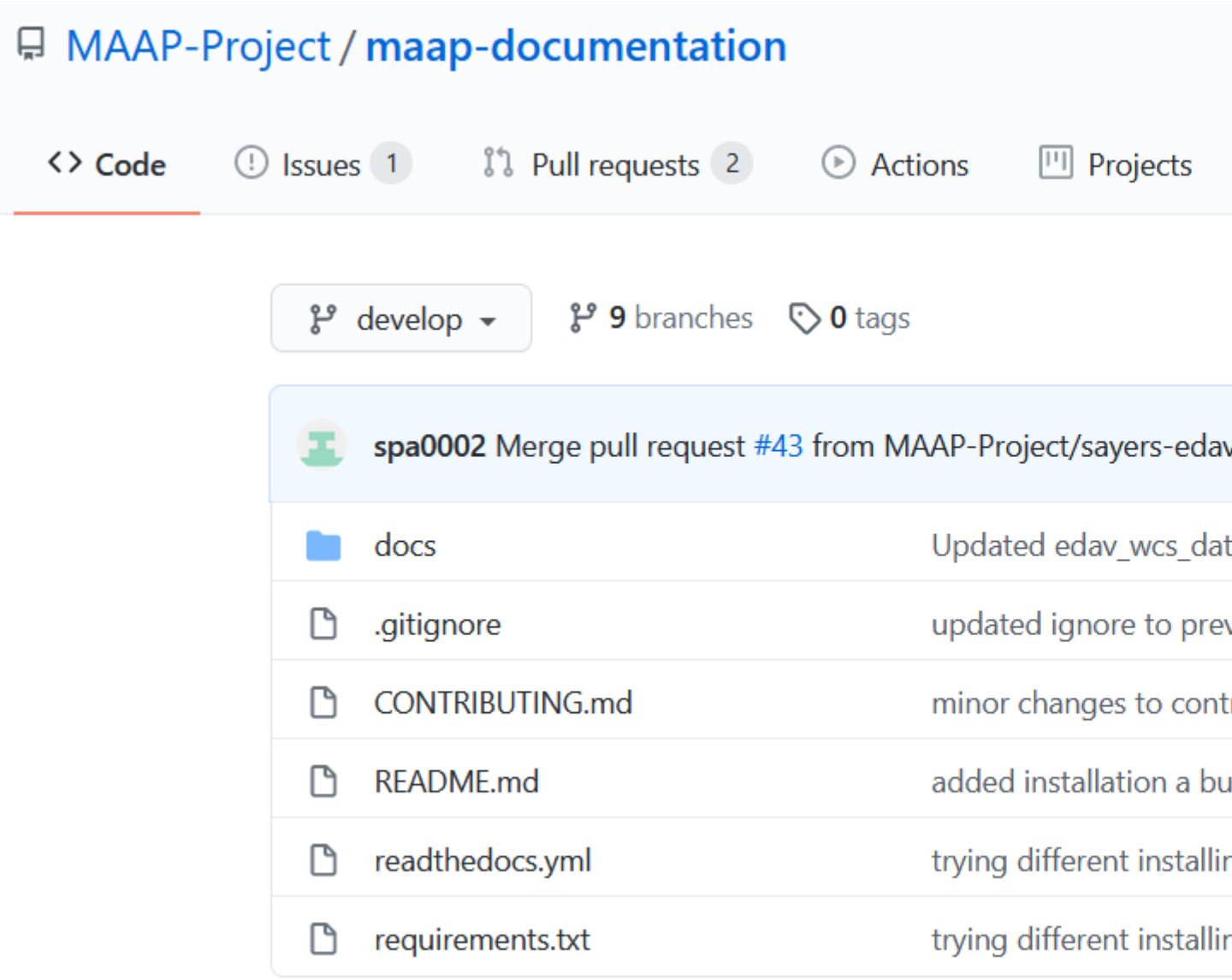
When inside of a workspace, navigate to **Git** tab at the top of the Jupyter window. Click it to see the option to **Clone**.



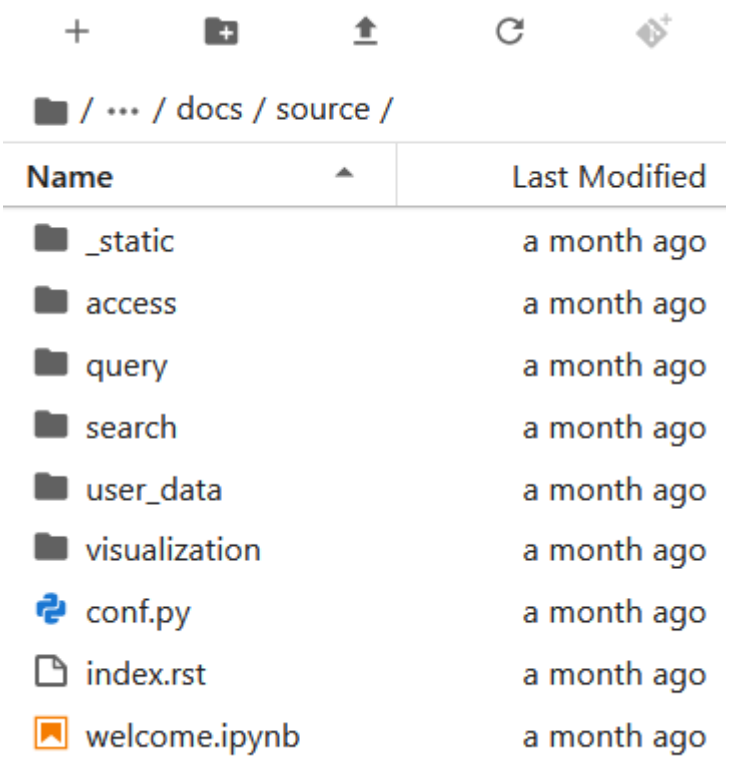
We can also access the "Clone a repository" dialogue box by selecting the **File Browser** tab on the JupyterLab sidebar, browsing to the location where we want our Git repository, and using the **Git** button located near the File

Browser icon (also to the left of the file list) and choosing "Clone a repository". The dialogue box prompts you to enter the URI of the repository you wish to clone. For this example we enter "https://github.com/MAAP-Project/maap-documentation.git".

For future reference, this URI can be found by visiting the GitHub site for the "MAAP-Project/maap-documentation" Git repository and clicking the **Code** button.



With the **File Browser** tab on the JupyterLab sidebar selected, a folder named "maap-documentation" should now appear at the location where you did the Git Clone operation. Folders for the various sections of the guide can be found in the "docs/source/" directory.



To open the IPython Notebook for an example, go to a section directory and double-click on appropriate ".ipynb" file. For more information about the using Git in Jupyterlab, see https://github.com/jupyterlab/jupyterlab-git .

Connecting to Github

Set personal access token: https://docs.github.com/en/authentication/keeping-your-account-and-data-secure/creating-a-personal-access-token

Working with the MAAP GitLab

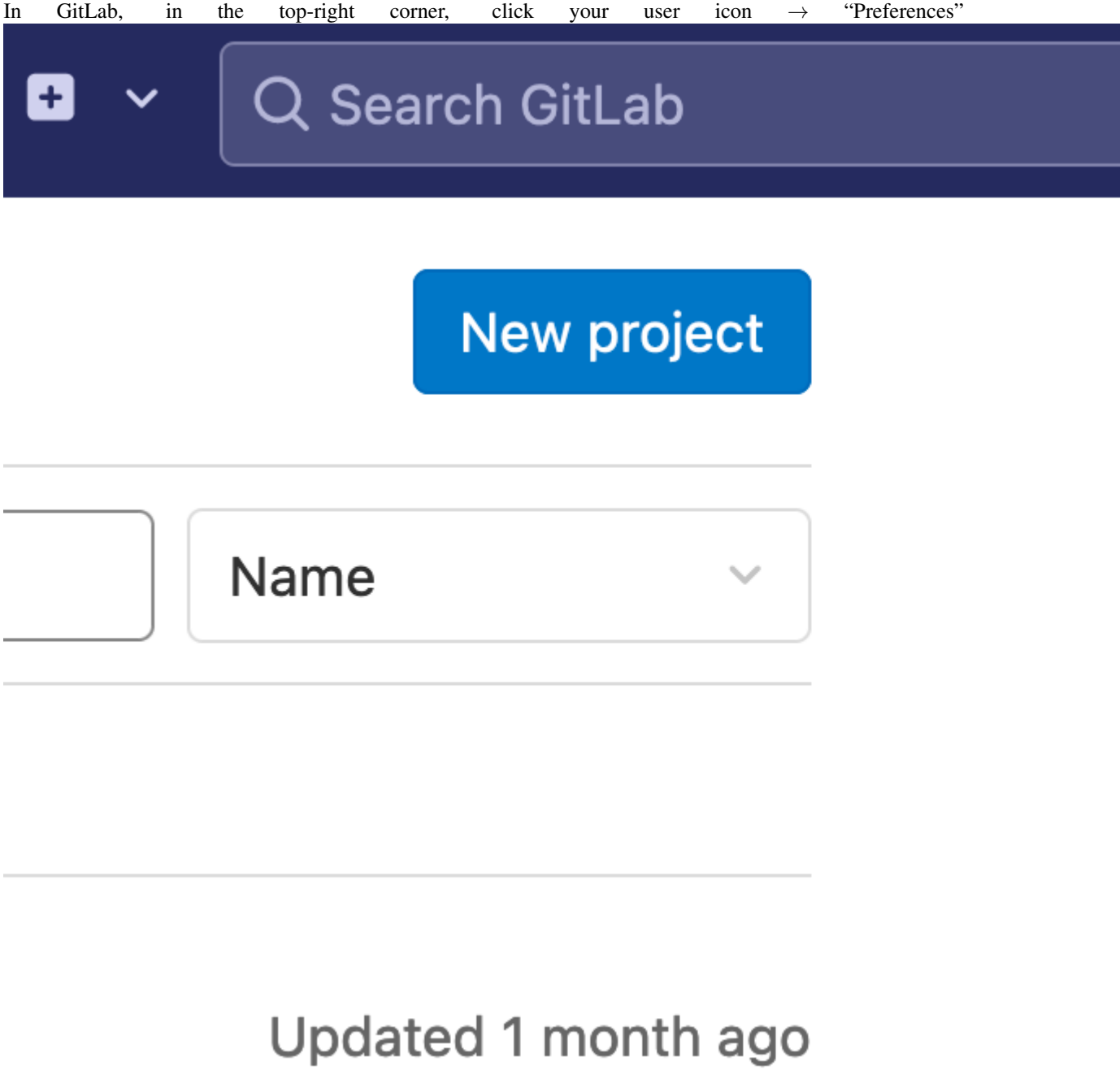
Note: In order to register algorithms, the code must be in the MAAP GitLab.

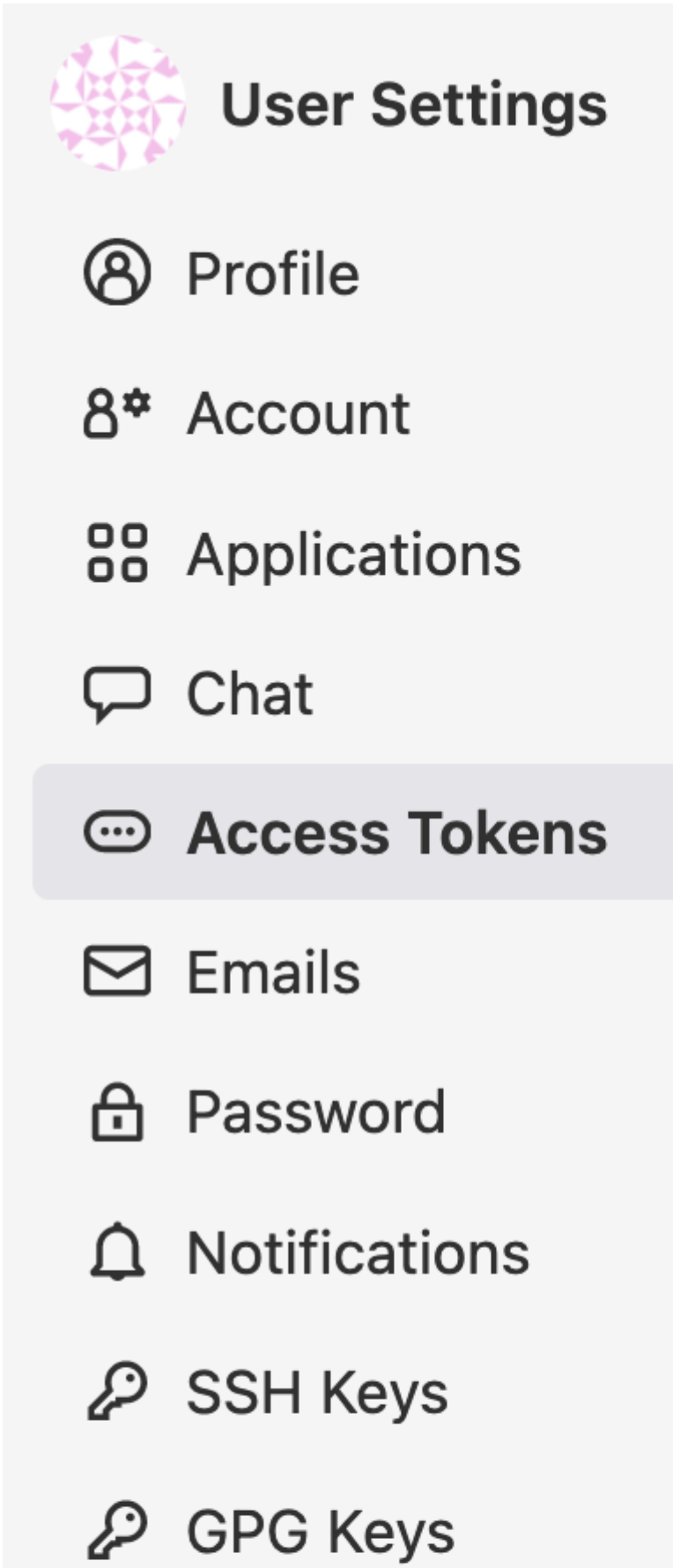
Note: git can behave slowly and strangely over s3 bucket-based storage (i.e., `my-private-bucket` and `my-public-bucket`). It is recommended to set up your git-tracked repos on the root (somewhere inside of ~ or `/projects`).

The MAAP GitLab instance is located at https://repo.maap-project.org/ . Make sure you can access this from the browser using your MAAP (EarthData Login) credentials.

For NASA security reasons, MAAP cannot communicate with its GitLab instance over SSH. There also isn't a username-password authentication option. Therefore, the recommended way to access MAAP repositories is to use GitLab Personal Access Tokens.

1. In GitLab, in the top-right corner, click your user icon → “Preferences”




User Settings
Profile
Account
Applications
Chat
Access Tokens
Emails
Password
Notifications
SSH Keys
GPG Keys
Preferences

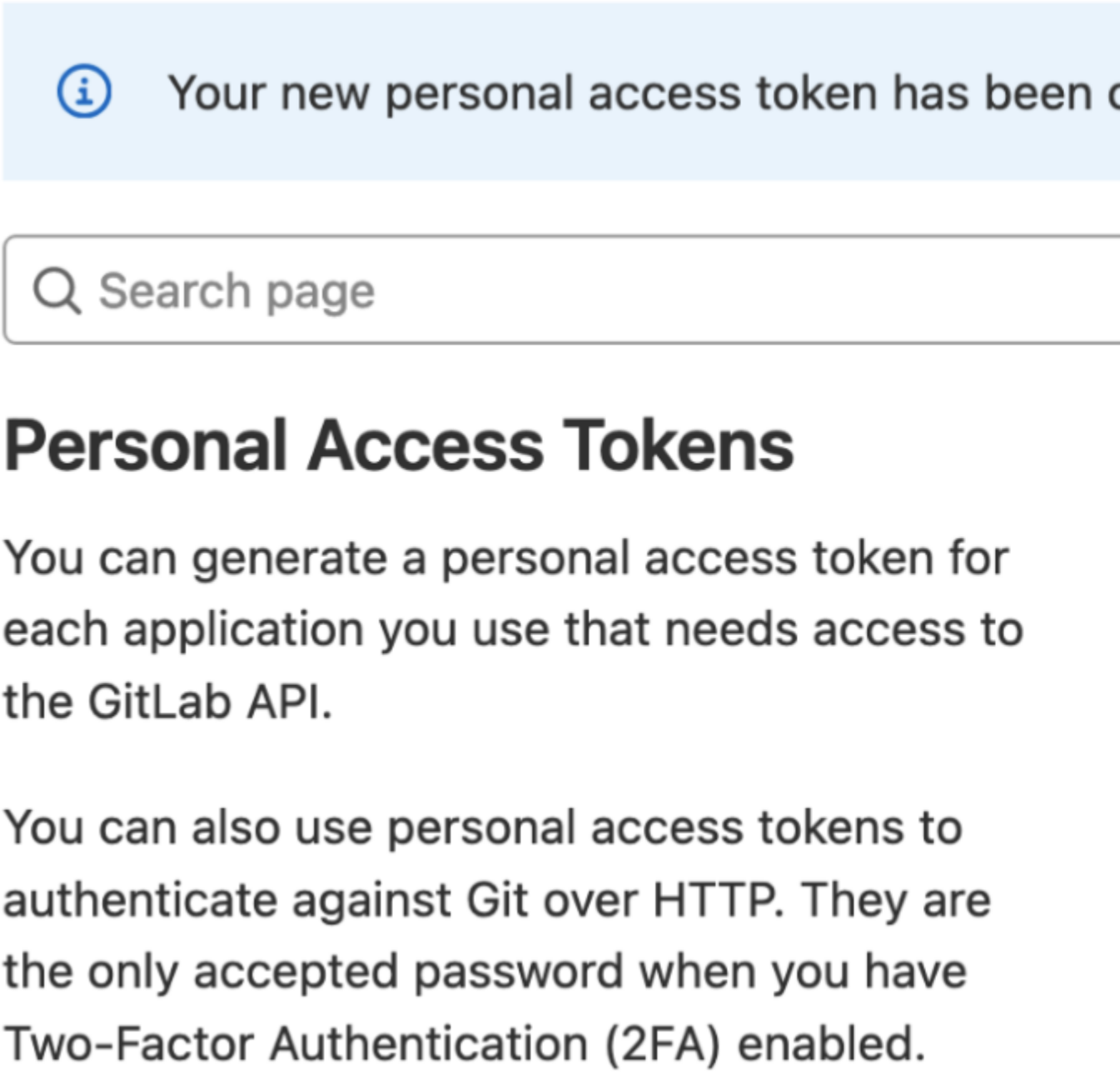
3. Create a new token with at least “read_repository” and “write_repository” permissions.

Personal Access Tokens

You can generate a personal access token for each application you use that needs access to the GitLab API.

You can also use personal access tokens to authenticate against Git over HTTP. They are the only accepted password when you have Two-Factor Authentication (2FA) enabled.

4. After clicking "create personal access token", you'll see a message like this pop up. Make sure you copy this token into a text file — you will not be able to access it again.



5. In the MAAP ADE, include this access token as part of the remote URL; e.g.,

```
git clone https://username:AccessToken@repo.maap-project.org/username/repo_name
```

For example:

```
git clone https://ashiklom:JJVimxhV8nmRNDqcCNr7@repo.maap-project.org/ashiklom/
↪fireatlas
```

For security reasons, MAAP DPS jobs can only use code that is stored on the MAAP GitLab. Therefore, if you are using the MAAP DPS, you will want to make sure that you are pushing at least a copy of your code to the MAAP GitLab as well as any other code repository (e.g., GitHub, SMCE GitLab). Note that it's possible to configure a repository to have multiple remotes — e.g.,

To add the `maap` remote and set the URL, use this:

```
git remote add maap https://username:AccessToken@repo.maap-project.org/username/repo_
↪name
```

If you already have the remote called `maap` set up, you can set the remote URL using this instead:

```
git remote set-url maap https://username:AccessToken@repo.maap-project.org/username/
↪repo_name
```

Then, you can use these commands to push your code and effectively synchronize between Github and MAAP GitLab (for algorithm registration):

Push to Github:

```
git push origin <branch name>
```

Push to MAAP GitLab:

```
git push maap <branch name>
```

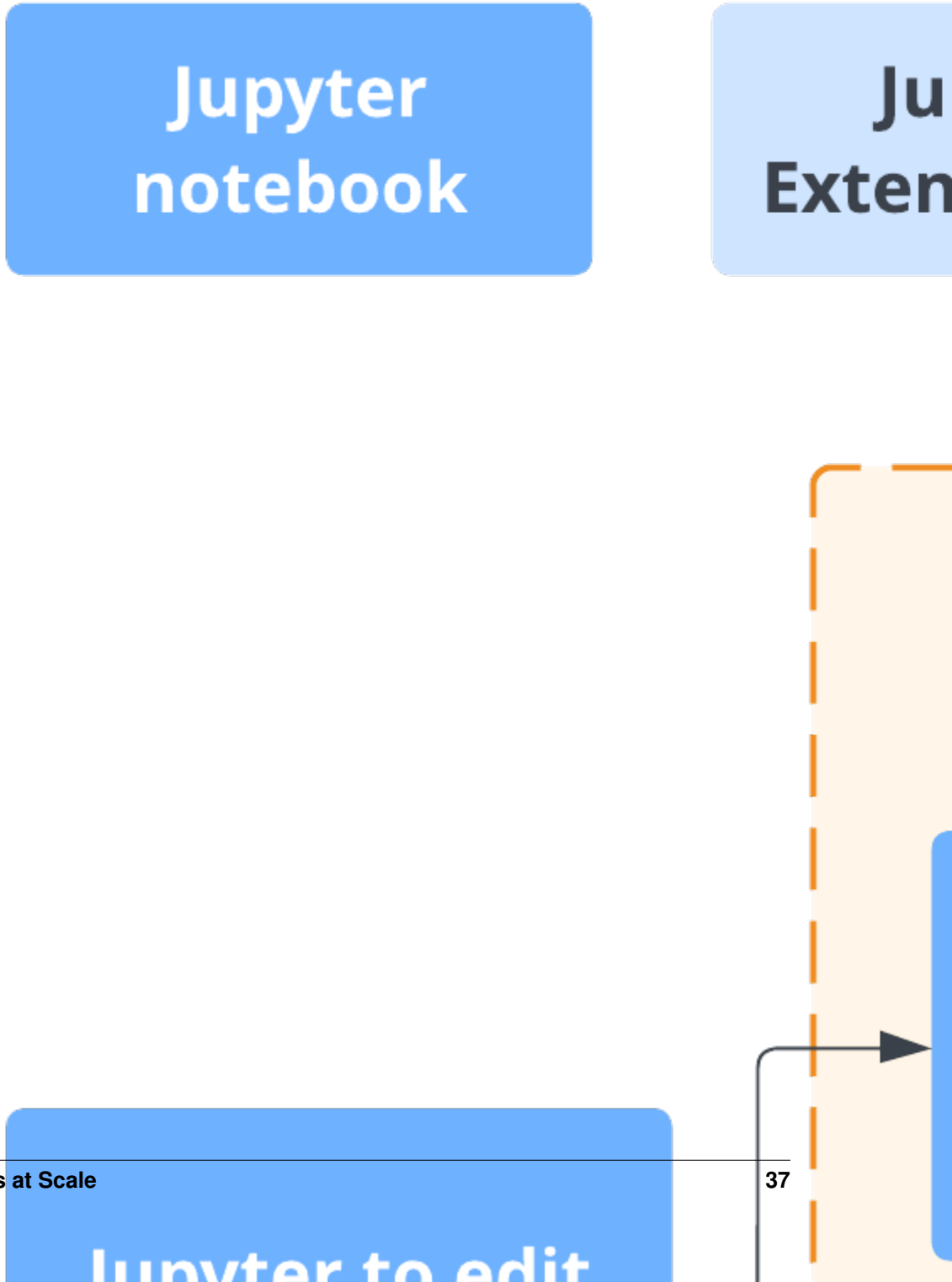
1.4.4 Customizing your workspace environment

Your Jupyter workspace has a set of pre-installed libraries, depending on *which Stack you selected*. If you need libraries that are not pre-installed, we suggest using an environment manager; `mamba` is pre-installed to help with this.

Full documentation on configuring mamba (or conda) may be found in the *System Reference Guide*.

1.5 Running Algorithms at Scale

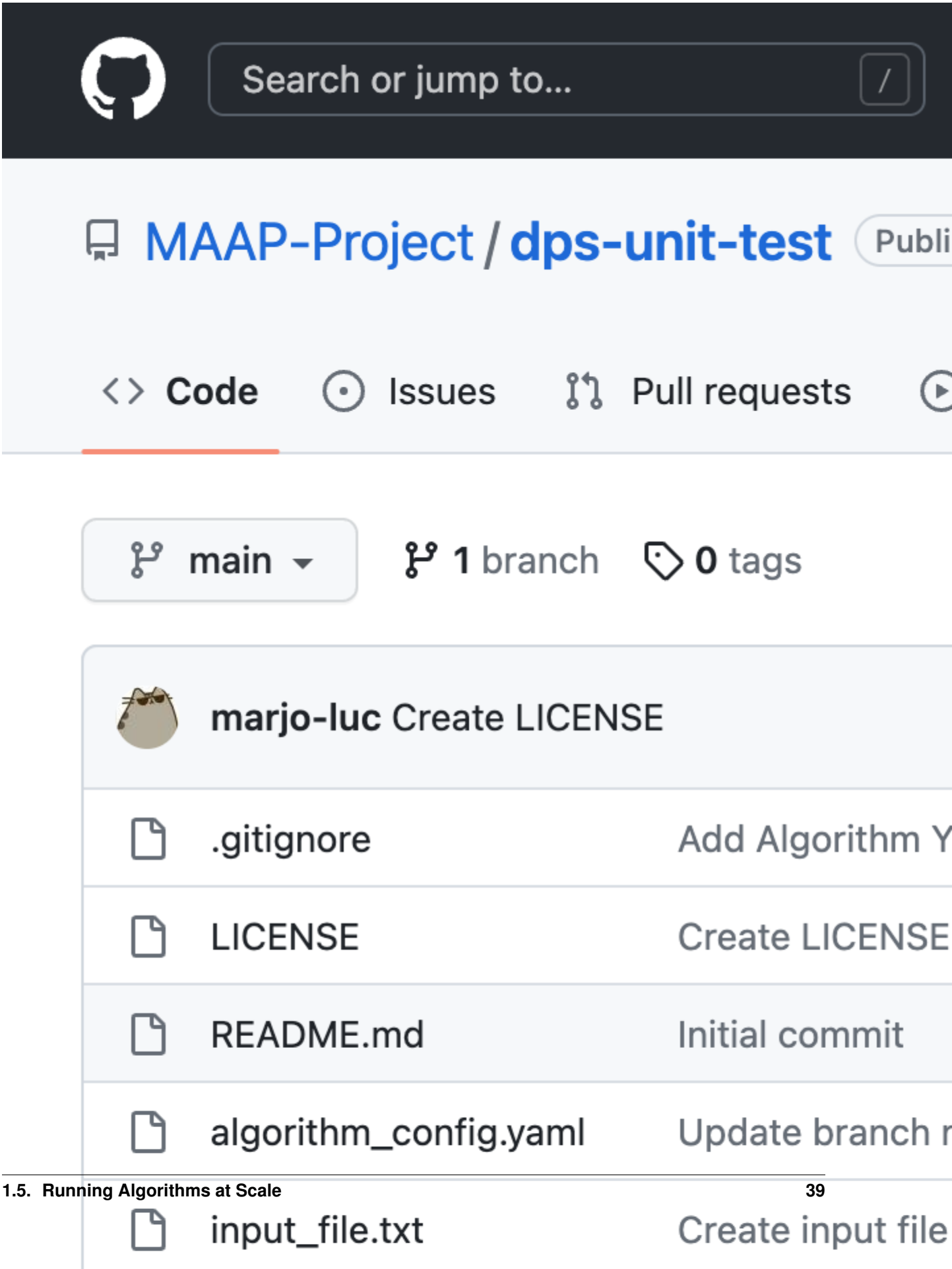
In order to run algorithms in the scaled-up cloud compute environment, they must first be "registered" in the Algorithm Catalog. This will make them available to other MAAP users, clearly define their inputs and outputs, and prepare them to be run easily in the Data Processing System (DPS).


Jupyter notebook
Ju
Exten

1.5.1 Register an Algorithm

Clone a test algorithm

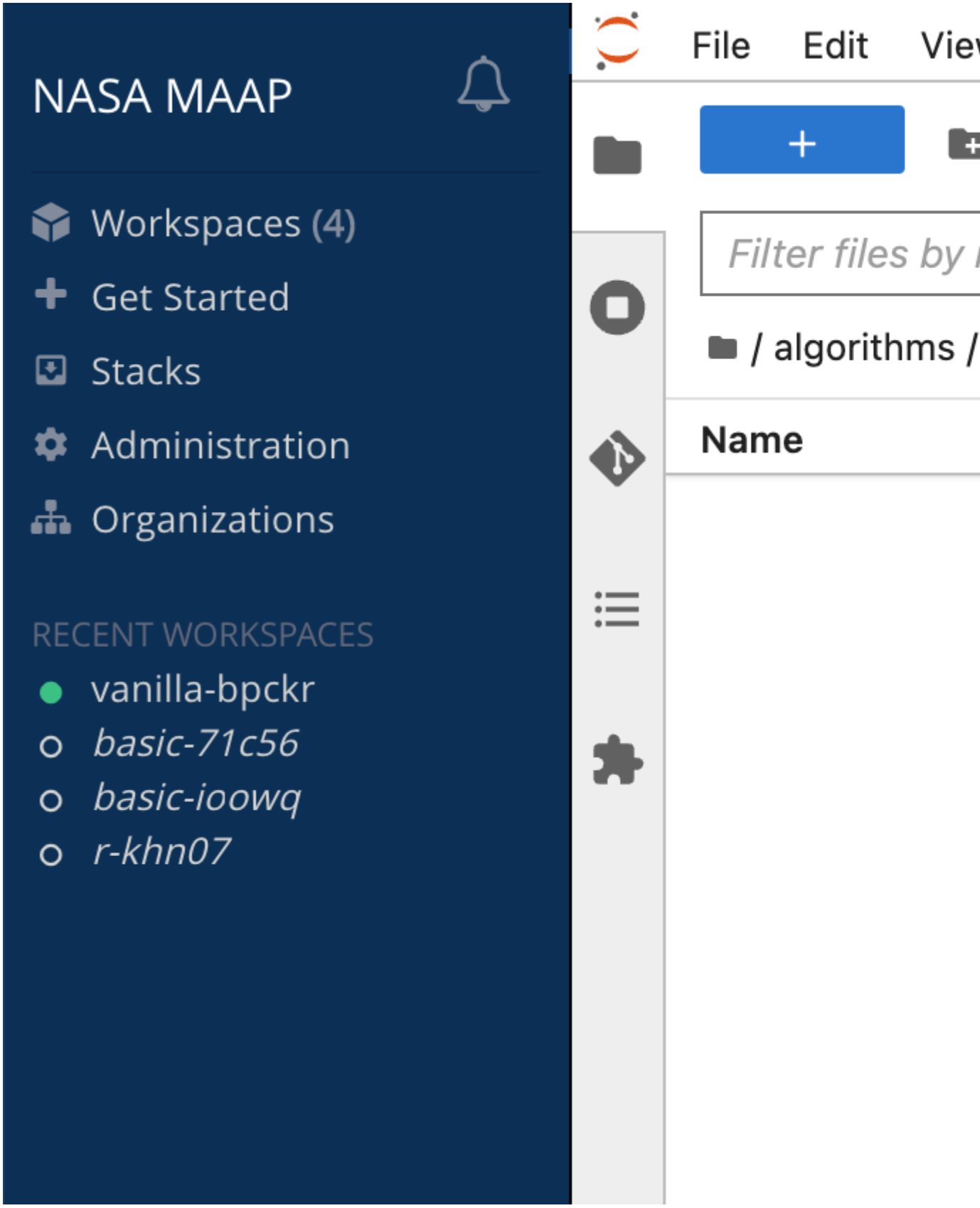
This is an example algorithm you can use for this getting started guide: https://github.com/MAAP-Project/dps-unit-test


Search or jump to...
/
MAAP-Project / dps-unit-test
Publi
Code
Issues
Pull requests
main
1 branch
0 tags
marjo-luc Create LICENSE
.gitignore
Add Algorithm Y
LICENSE
Create LICENSE
README.md
Initial commit
algorithm_config.yaml
Update branch
input_file.txt
Create input file

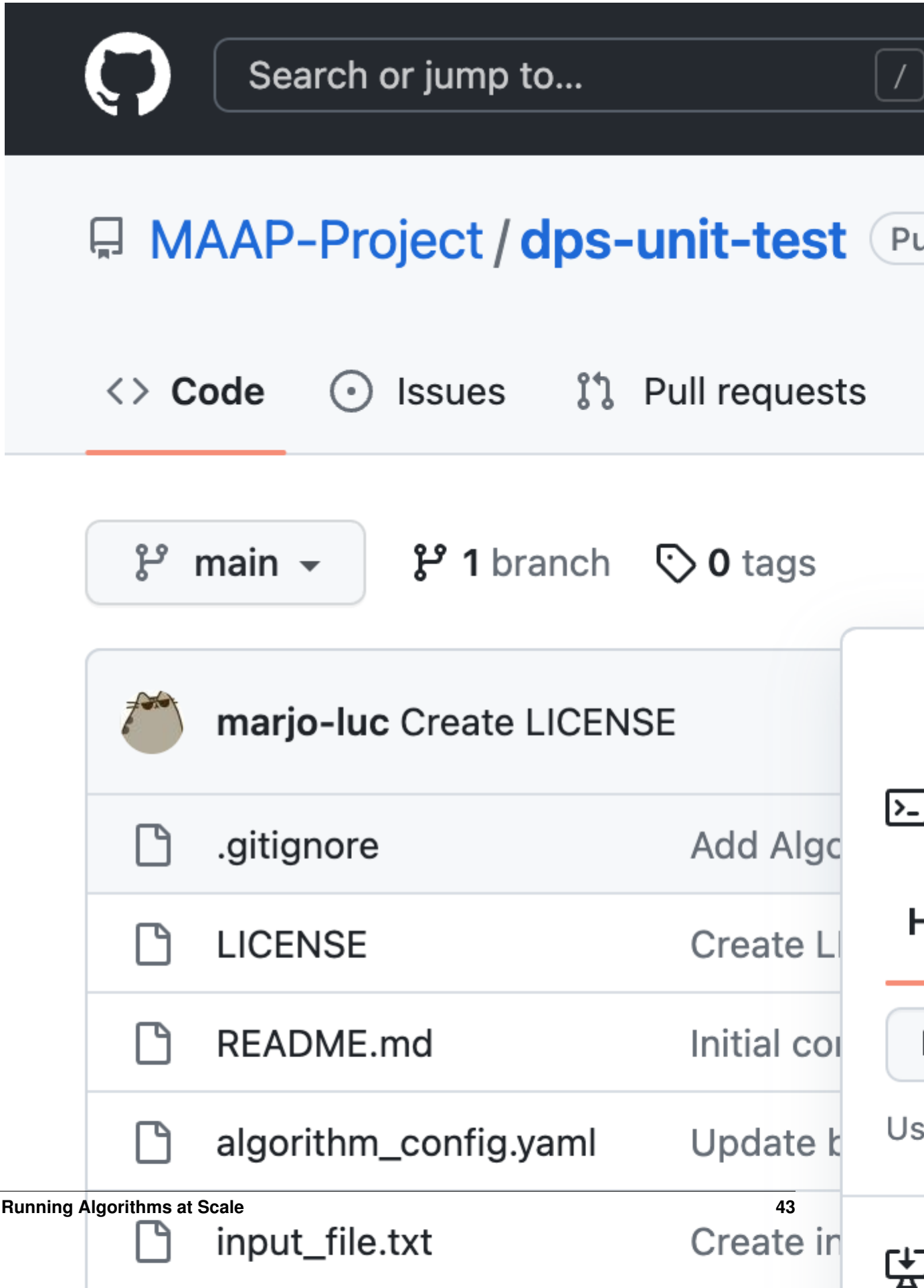
In the repo there are a few files that you will typically have, or which are required:

- `algorithm_config.yml` is a required file that has a description of the inputs and outputs of the algorithm along with other parameters like the run command.
- `run_test.sh` is the run command for this algorithm. It is typical to have a shell script to tell the system how to run the algorithm and set some environmental variables.

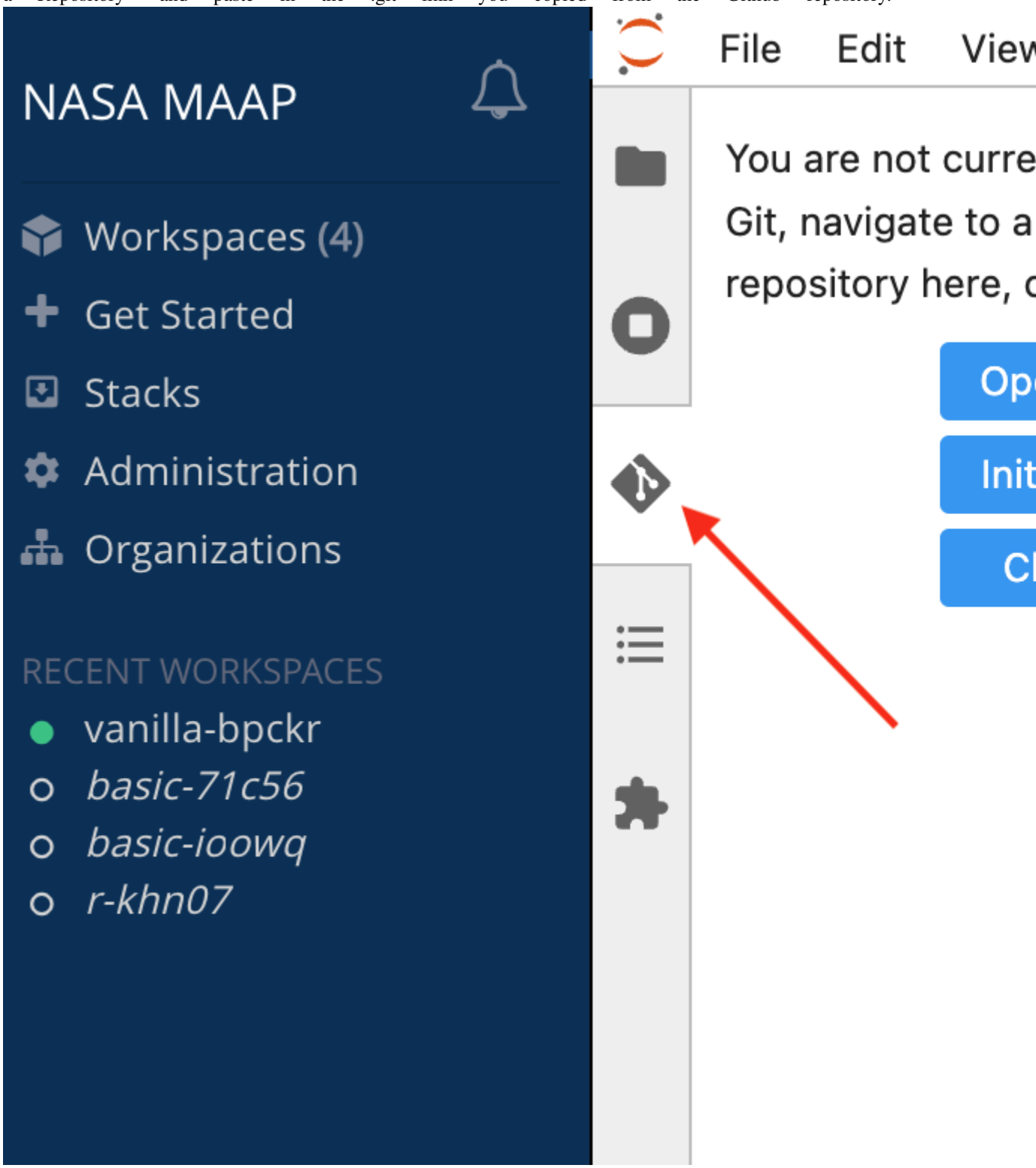
1. Make a new folder for your test algorithm. Open a terminal here (File > New > Terminal or use the blue "+" button above the Jupyter file browser).

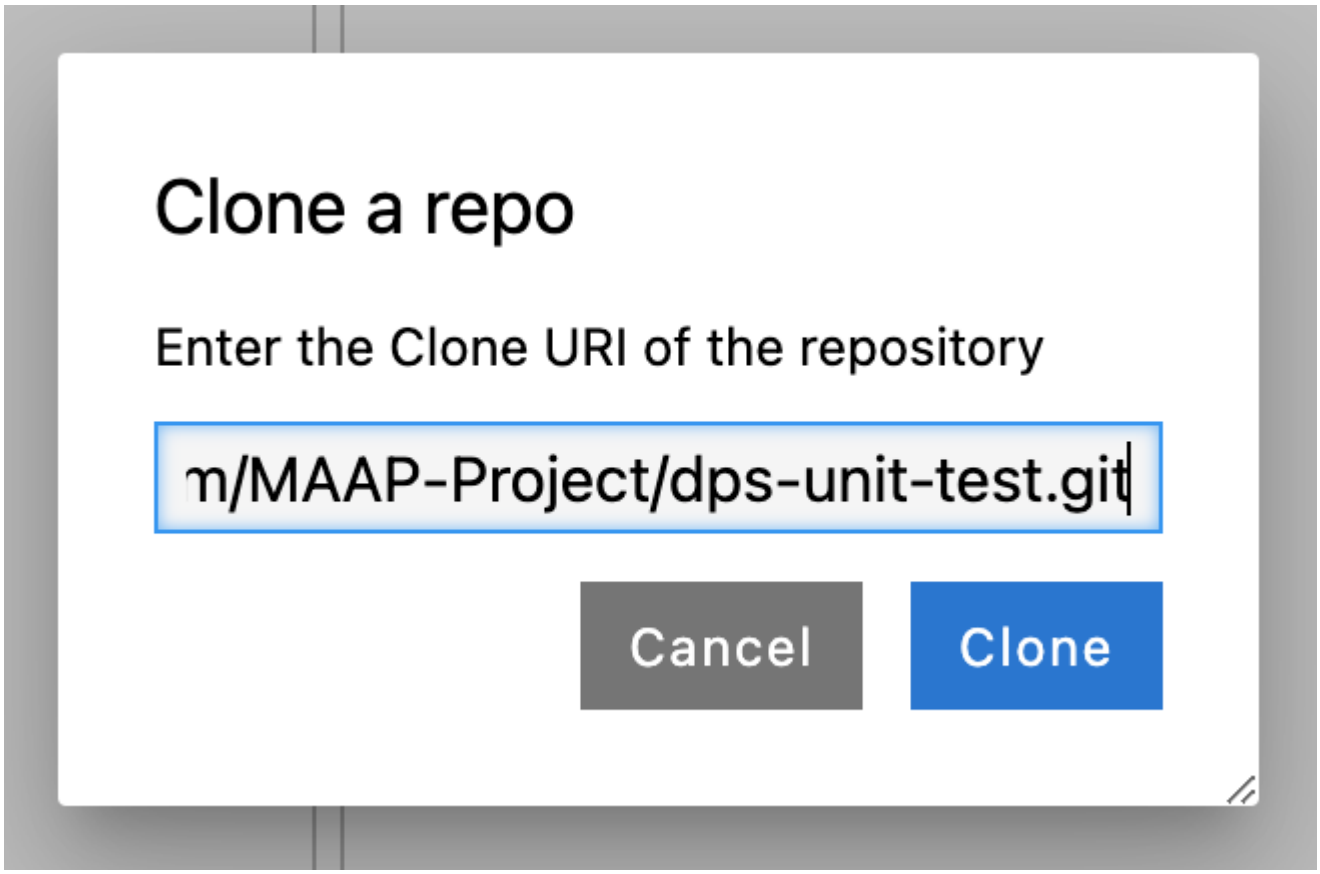

NASA MAAP
Workspaces (4)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
vanilla-bpckr
basic-71c56
basic-ioowq
r-khn07
File
Edit
Filter files by
/ algorithms /
Name

2. Copy the Github clone link from https://github.com/MAAP-Project/dps-unit-test

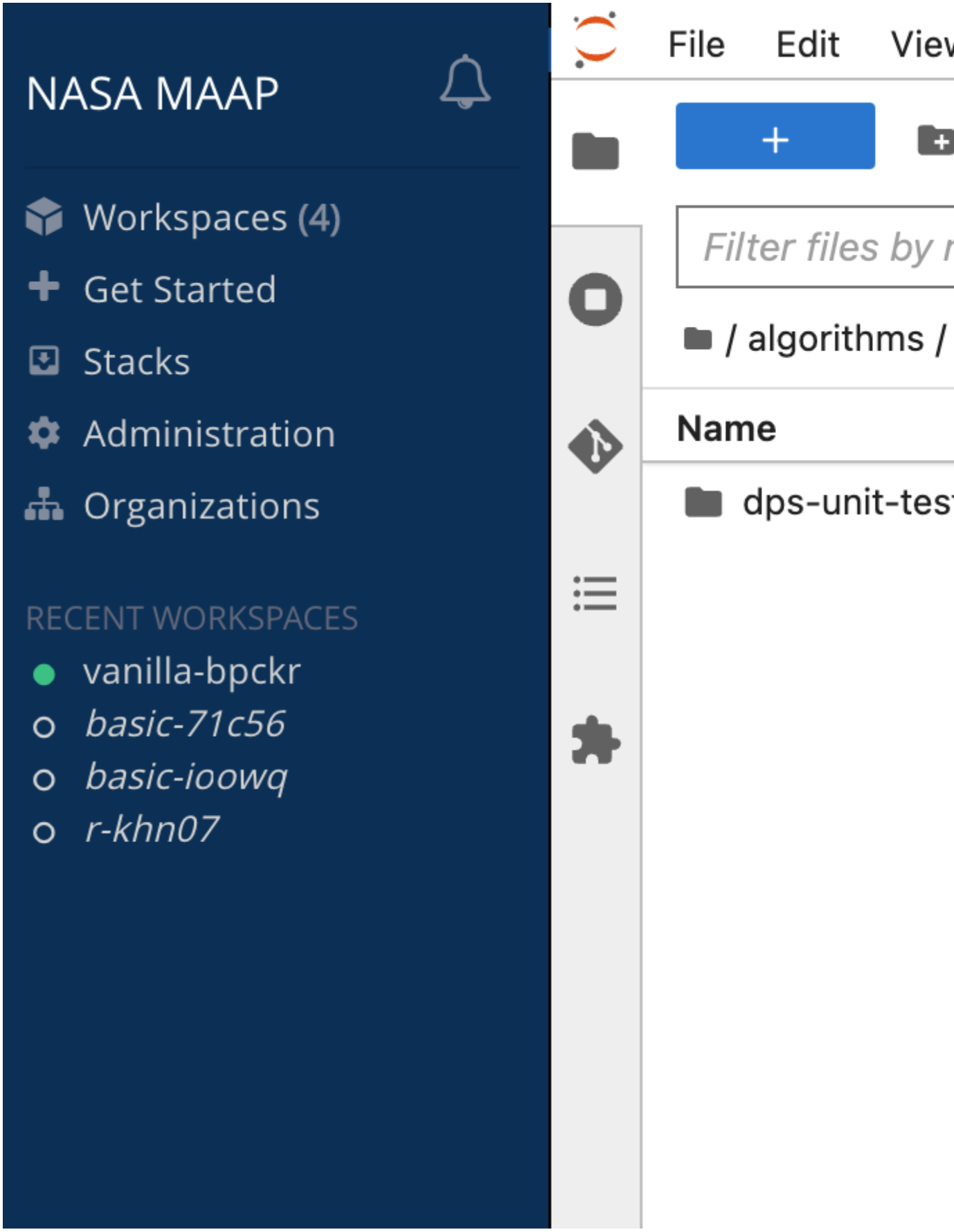

Search or jump to...
/
MAAP-Project / dps-unit-test
Pu
Code
Issues
Pull requests
main
1 branch
0 tags
marjo-luc Create LICENSE
.gitignore
Add Algo
LICENSE
Create LI
README.md
Initial co
algorithm_config.yaml
Update k
input_file.txt
Create in

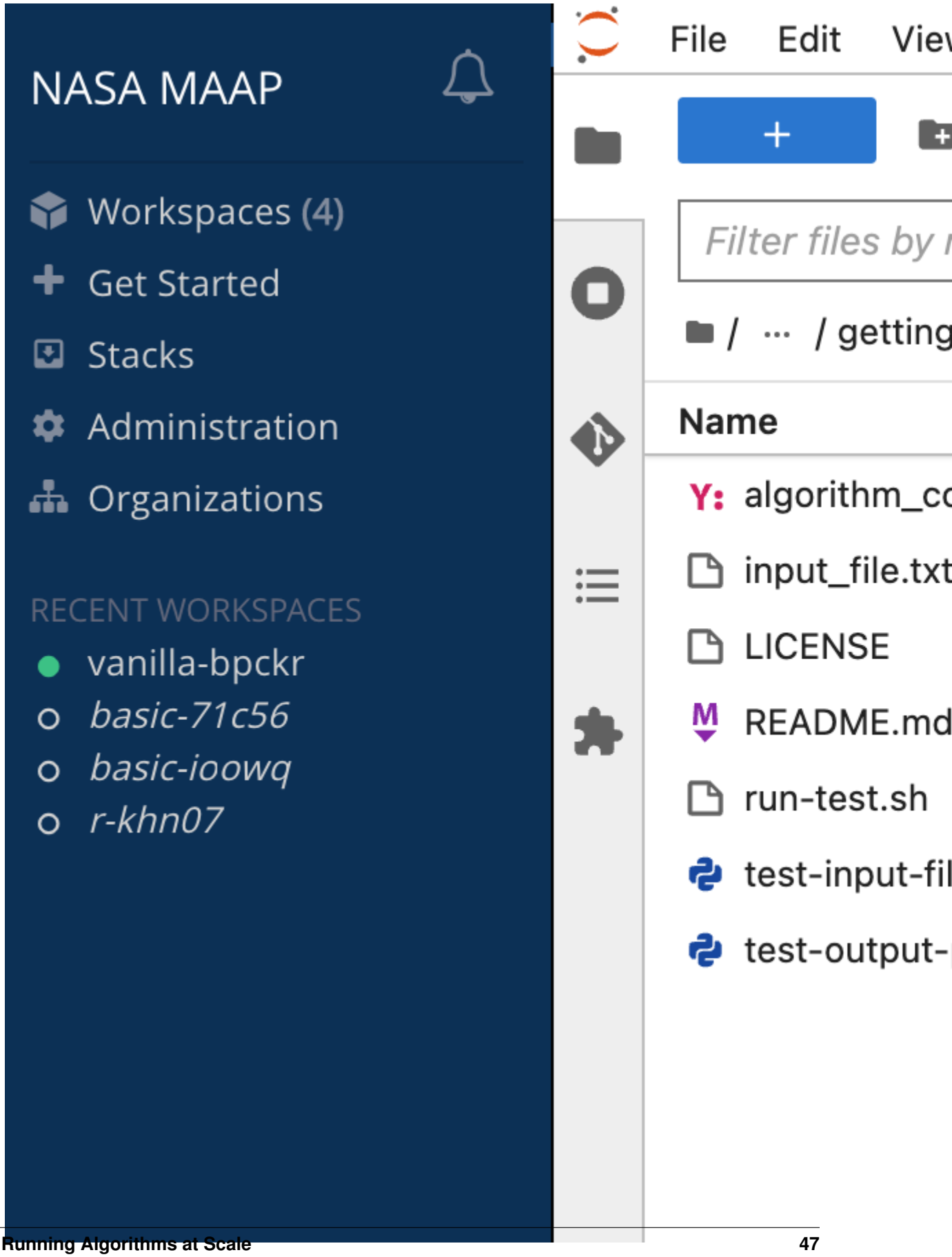
3. Open the built-in Jupyter Github UI to the left of the file browser. Choose "Clone a Repository" and paste in the .git link you copied from the Github repository.

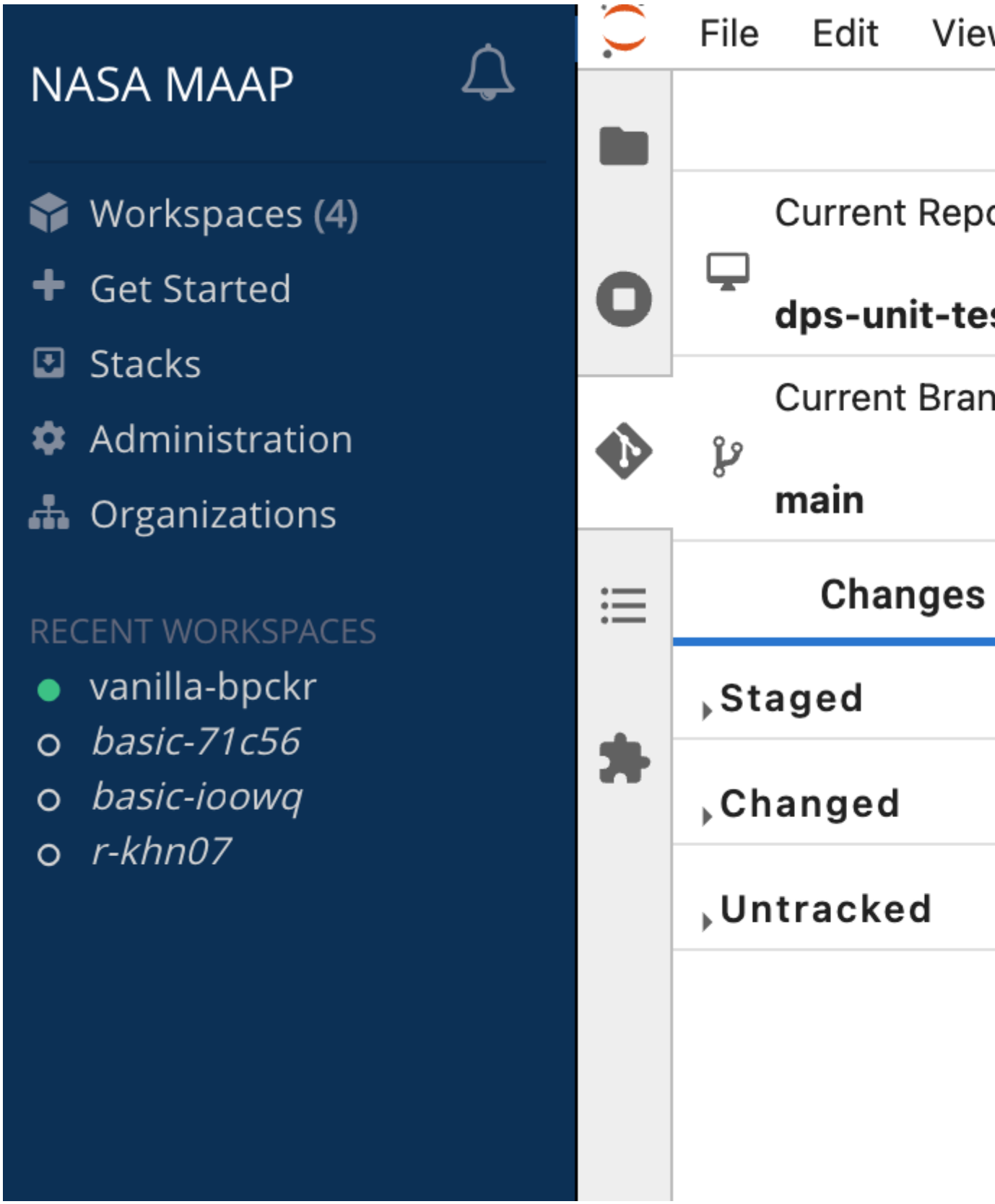




4. You should see a new folder created with the repo you cloned. If you browse to that folder and open up the Jupyter Github UI again, it will show you some info about that repo.


NASA MAAP
Workspaces (4)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
vanilla-bpckr
basic-71c56
basic-ioowq
r-khn07
File
Edit
Filter files by
/ algorithms /
Name
dps-unit-tes

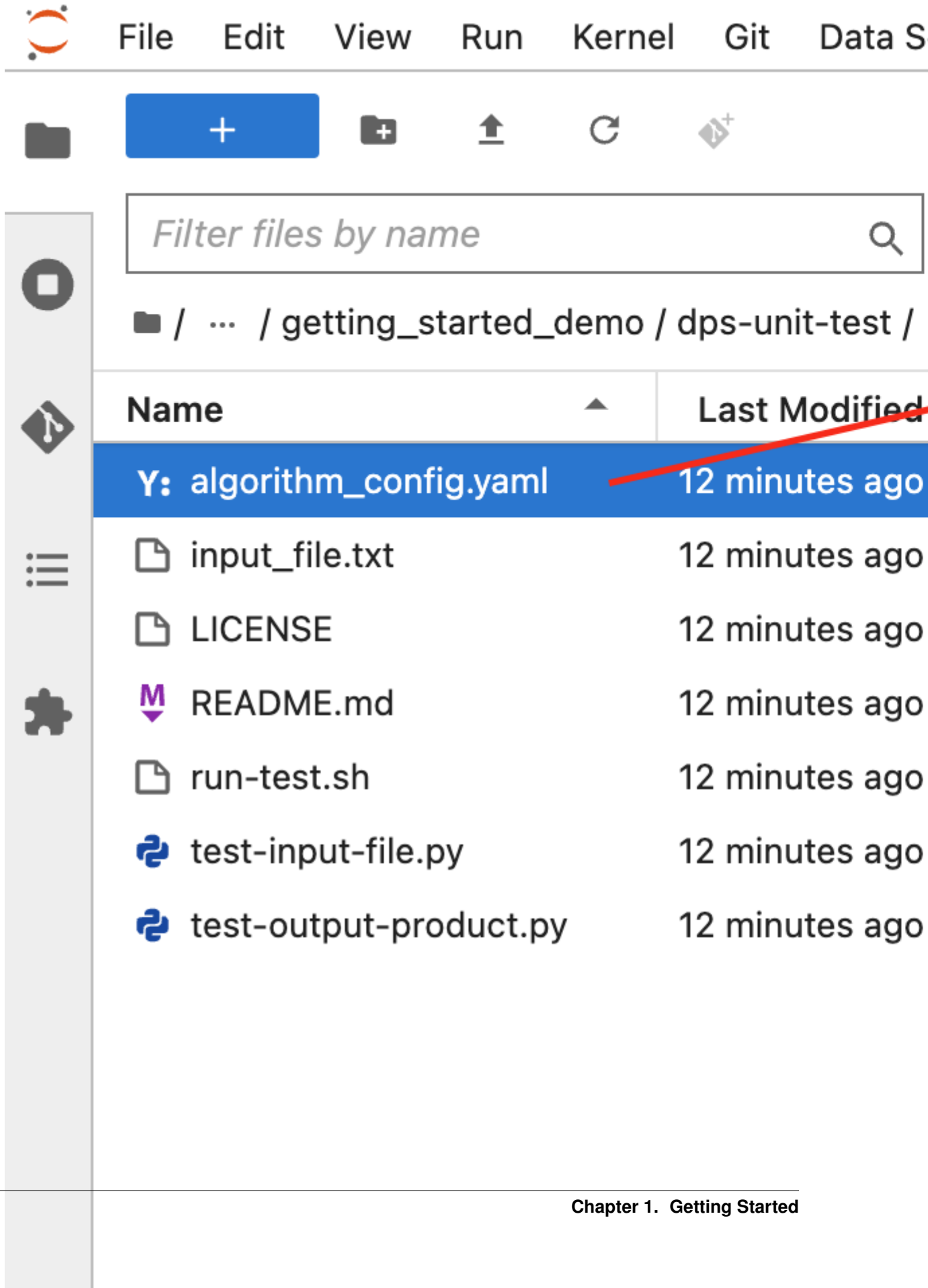

NASA MAAP
Workspaces (4)
Get Started
Stacks
Administration
Organizations
RECENT WORKSPACES
vanilla-bpckr
basic-71c56
basic-ioowq
r-khn07
File
Edit
Name
input_file.txt
LICENSE
README.md
run-test.sh



5. If you want to make changes to the code and have your own copy of it to register, Clone the code into your

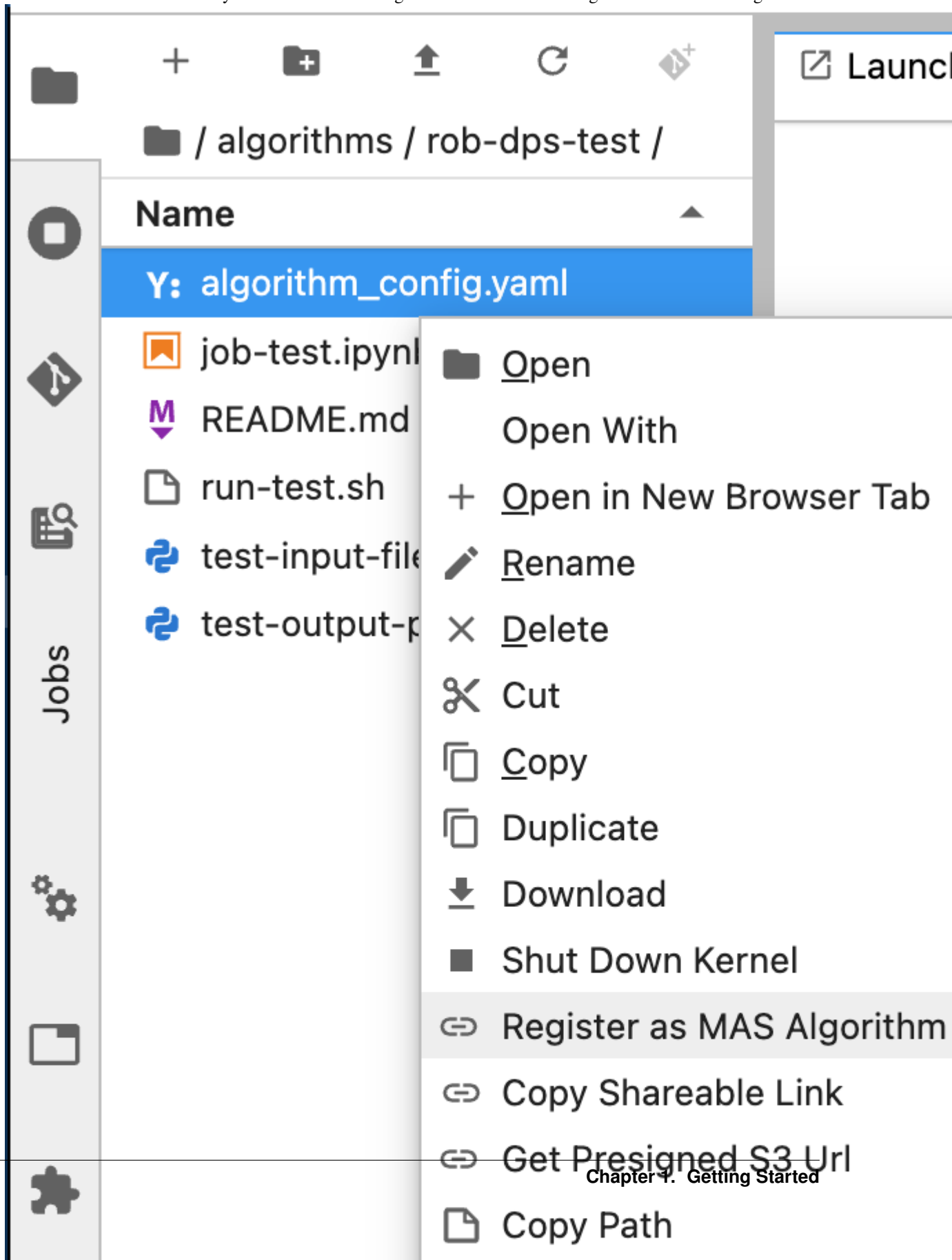
MAAP GitLab. The git link to your code is indicated in the `algorithm_config.yml`. If you would prefer to skip this for now, leave the repository_url in `algorithm_config.yml` pointed at the "root" user (`repository_url: https://repo.dit.maap-project.org/root/dps-unit-test.git`)

6. Rename the algorithm to personalize it. You do this by opening up the `algorithm_config.yml` file and changing the `algo_name` field.


File
Edit
View
Run
Kernel
Git
Data S
Filter files by name
/ ··· / getting_started_demo / dps-unit-test /
Name
Last Modified
algorithm_config.yaml
12 minutes ago
input_file.txt
12 minutes ago
LICENSE
12 minutes ago
README.md
12 minutes ago
run-test.sh
12 minutes ago
test-input-file.py
12 minutes ago
test-output-product.py
12 minutes ago

Register the algorithm

1. Make sure code is ready and saved -> right click file -> “Register as MAS Algorithm”



2. This automatically creates `algorithm_config.yaml` file with the presets if it is not already present (which, in this example case, it is present). There is only one for any directory. At this point you would normally edit the configuration file, then repeat step 1 and click "OK" to register. For this example we did this in step 5 in the previous section.
3. Outputs (if any) should be written to a folder named `outputs`. There are none in the example we are using here.

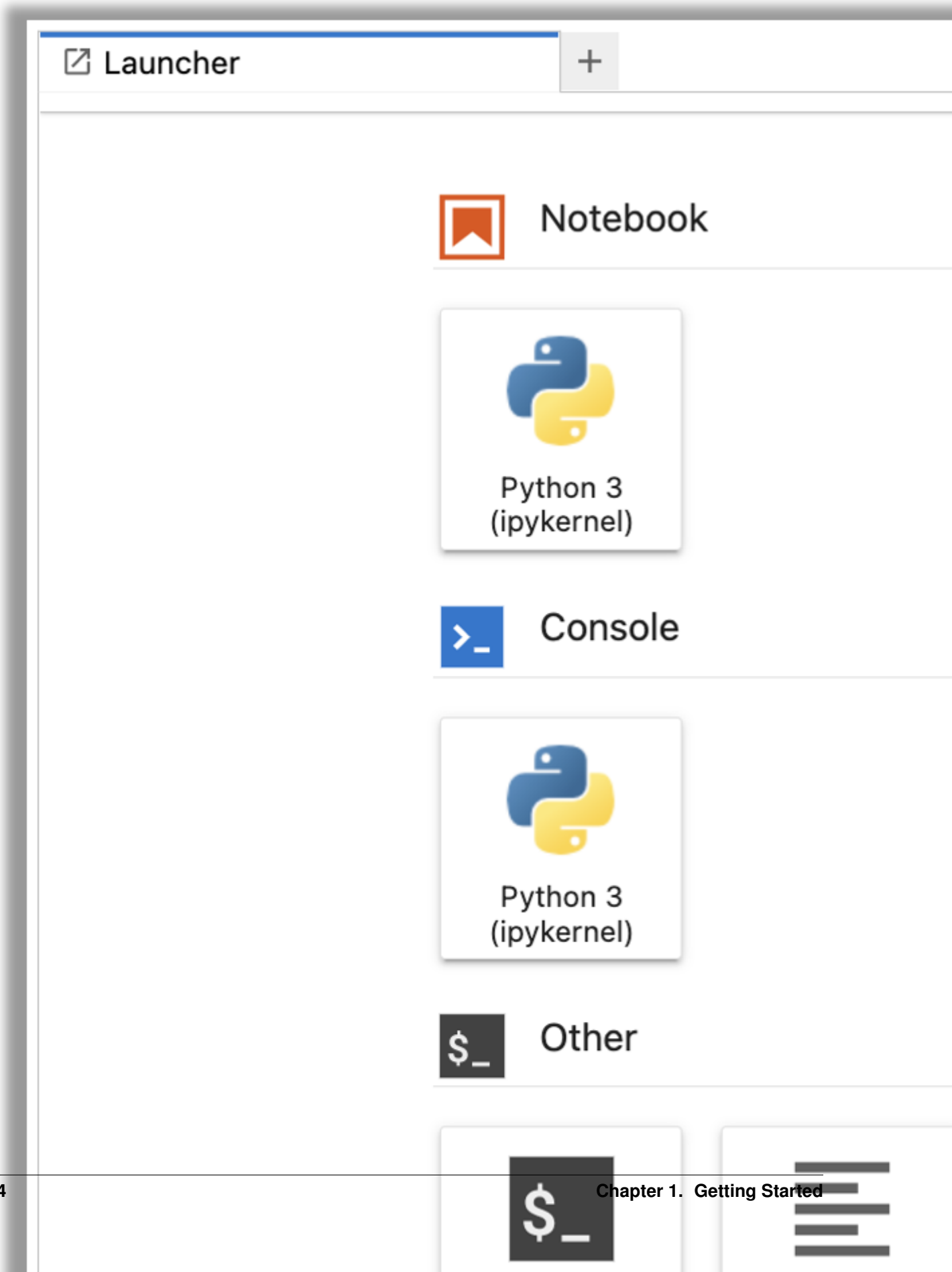
Note: It can take some time to register an algorithm. You can determine if it has completed when you see it appear in the Jobs UI (see below) or in the menus under DPS/MAS Operations > List Algorithms.

1.5.2 Run the Algorithm as a Job and Monitor it

The Jobs UI

MAAP is configured to run up to 4,000 concurrent jobs. There are two additional ways to run a job: via the Jobs UI in the Launcher, or via a call to the maap-py Python library.

The Jobs UI will let you run and monitor jobs easily. You can find full documentation in the system reference guide for *the Jobs UI*. You can also find specific documentation on *how to submit jobs* and *how to monitor jobs* in the System Reference Guide FAQs.


Launcher
Notebook
Python 3
(ipykernel)
Console
Python 3
(ipykernel)
Other

Some alternative methods of running the job are found below.

Pop-up

- Click DPS/MAS Operations menu -> Execute DPS Job
- Select your algorithm from the dropdown
- A new popup will ask for inputs; if it doesn't take inputs, the popup will say so.
- Click OK again to view the ID for the job just submitted.

OR

maap-py

Import the `maap-py` library: if in Jupyter, click the small blue MAAP button in the top left corner to automatically insert code. If using a script, add these lines manually at the top of your notebook:

```
from maap.maap import MAAP
maap = MAAP()
```

Pass your algorithm's name, version, required inputs, and username to the function `maap.submitJob (identifier is job- algo_name:algo_version)` Check result: `maap.getJobResult()`

CHAPTER 2

Science Examples

2.1 HLSL30 Search and Composite

Authors: Nathan Thomas (GSFC/UMD), Sumant Jha (MSFC/USRA), Aimee Barciauskas (DevSeed), Alex Mandel (DevSeed)

Date: December 19, 2022

Description: In this tutorial, we will search the LPDAAC for Harmonized Landsat Sentinel-2 (HLS) 30m optical imagery that intersects an AOI. We will filter the catalog based on a cloud cover % and build a maximum-NDVI (Normalized Difference Vegetation Index) composite image, including a suite of popular indices, which will help give us an in-depth look at vegetation health.

2.1.1 Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

2.1.2 About the Data

Harmonized Landsat Sentinel-2 30m

Harmonized Landsat Sentinel-2 (HLS) was developed in response to a greater need for moderate-to-high resolution imagery to track various short-term landcover changes. Data are gathered by the Landsat-8 and Landsat-9 satellites, which carry the Operational Land Imager (OLI), as well as the Sentinel-2A and Sentinel-2B satellites, which carry the Multi-Spectral Instrument (MSI). With combined measurements from both the Landsat and Sentinel satellites, HLS imagery has global coverage with a spatial resolution of 30m and a temporal resolution of 2-3 days. (Source: HLS Overview Page)

Note about HLS datasets: The Sentinel and Landsat assets have been "harmonized" in the sense that these products have been generated to use the same spatial resolution and grid system. Thus, the HLS S30 and L30 products can be used interchangeably in algorithms and are "stackable". However, the individual band assets are specific to each provider.

2.1.3 Additional Resources

- HLSL30 v002 Dataset Landing Page
- Landsat 8 Bands and Combinations Blog
- HLSS30 v002 Dataset Landing Page
- Sentinel 2 Bands and Combinations Blog
- Harmonized Landsat Sentinel-2 (HLS) User Guide

2.1.4 Importing and Installing Packages

We will begin by installing any packages we need and importing the packages that we will use.

If needed the following packages should be installed:

[1]:
```
# cleanup data: removes data files that should be replaced
!rm -rf ./local-s3.json
!rm -rf ./sample.json

if False:
    !conda install -c conda-forge pystac-client -y
    !conda install -c conda-forge rio-tiler -y
```

Prerequisites

- geopandas
- folium
- pystac-client
- rio_tiler

[ ]:
```
# Uncomment the following lines to install these packages if you haven't already.
# !pip install geopandas
# !pip install folium
# !pip install pystac-client
# !pip install rio_tiler
```

We will now import a suite of packages that we will need:

[3]:
```
from maap.maap import MAAP
maap = MAAP(maap_host='api.maap-project.org')
import geopandas as gpd
import folium
import h5py
import pandas
import matplotlib
import matplotlib.pyplot as plt
from shapely.geometry import Polygon
```

(continues on next page)

(continued from previous page)

```
from pystac_client import Client
import datetime
import os
import rasterio as rio
import boto3
import json
import botocore
from rasterio.session import AWSSession
from rio_tiler.io import COGReader
import numpy as np
import matplotlib
import matplotlib.pyplot as plt
```

```
/opt/conda/lib/python3.7/site-packages/geopandas/_compat.py:115: UserWarning: The␣
↪Shapely GEOS version (3.11.1-CAPI-1.17.1) is incompatible with the GEOS version␣
↪PyGEOS was compiled with (3.8.1-CAPI-1.13.3). Conversions between both will be slow.
  shapely_geos_version, geos_capi_version_string
```

2.1.5 Creating an AOI

To begin we will create a polygon in a forested area in Virginia, USA. We will do this by providing a number of lat/lon coordinates and creating an AOI.

[4]:
```
lon_coords = [-80, -80, -79., -79, -80]
lat_coords = [39, 38.2, 38.2, 39, 39]

polygon_geom = Polygon(zip(lon_coords, lat_coords))
crs = 'epsg:4326'
AOI = gpd.GeoDataFrame(index=[0], crs=crs, geometry=[polygon_geom])
AOI_bbox = AOI.bounds.iloc[0].to_list()
```

We can visualize this polygon using a folium interactive map.

[5]:
```
m = folium.Map([38.5,-79.3], zoom_start=9, tiles='OpenStreetMap')
folium.GeoJson(AOI).add_to(m)
folium.LatLngPopup().add_to(m)
m
```

[5]:
```
<folium.folium.Map at 0x7f223c58c090>
```

2.1.6 Accessing the HLS Data

To be able to access the HLS imagery we need to activate a 'rasterio' AWS session. This will give us the required access keys that we need to search and read data from the LPDAAC S3 bucket.

[6]:
```
def get_aws_session_DAAC():
    """Create a Rasterio AWS Session with Credentials"""
    creds = maap.aws.earthdata_s3_credentials('https://data.lpdaac.earthdatacloud.
↪nasa.gov/s3credentials')
    boto3_session = boto3.Session(
        aws_access_key_id=creds['accessKeyId'],
        aws_secret_access_key=creds['secretAccessKey'],
        aws_session_token=creds['sessionToken'],
        region_name='us-west-2'
```

(continues on next page)

(continued from previous page)

```
    )
    return AWSSession(boto3_session)
```

[7]:
```
print('Getting AWS credentials...')
aws_session = get_aws_session_DAAC()
print('Finished')
```

```
Getting AWS credentials...
Finished
```

Now that we have our session credentials set up, we can search the HLS catalog using the following function, filtering by spatial extent (our AOI) and a time window:

[8]:
```
def query_stac(year, bbox, max_cloud, api, start_month_day, end_month_day):
    print('opening client')
    catalog = Client.open(api)

    date_min = str(year) + '-' + start_month_day
    print('start_month_day:\t\t', start_month_day)
    print('end_month_day:\t\t', end_month_day)
    date_max = str(year) + '-' + end_month_day
    start_date = datetime.datetime.strptime(date_min, "%Y-%m-%d")
    end_date = datetime.datetime.strptime(date_max, "%Y-%m-%d")
    start = start_date.strftime("%Y-%m-%dT00:00:00Z")
    end = end_date.strftime("%Y-%m-%dT23:59:59Z")

    print('start date, end date:\t\t', start, end)

    print('\nConducting HLS search now...')

    search = catalog.search(
        collections=["HLSL30.v2.0"],
        datetime=[start,end],
        bbox=bbox,
        max_items=500, # for testing, and keep it from hanging
        # query={"eo:cloud_cover":{"lt":20}} #doesn't work
    )
    print(f"Search query parameters:\n{search}\n")
    results = search.get_all_items_as_dict()

    return results
```

Here we run our STAC search and write the results out to a JSON file so we can access it later.

[9]:
```
search = query_stac(2020, AOI_bbox, 20, 'https://cmr.earthdata.nasa.gov/stac/LPCLOUD',
↪ '06-01', '09-30')

with open("./sample.json", "w") as outfile:
    json.dump(search, outfile)
```

```
opening client
start_month_day:                06-01
end_month_day:                  09-30
start date, end date:             2020-06-01T00:00:00Z 2020-09-30T23:59:59Z

Conducting HLS search now...
Search query parameters:
```

(continues on next page)

(continued from previous page)

```
<pystac_client.item_search.ItemSearch object at 0x7f223c514850>
```

So far, we have not filtered by cloud cover, which is a common filtering parameter for optical imagery. We can use the metadata files included in our STAC search to find the amount of cloud cover in each file and decide if it meets our threshold or not. We will set a cloud cover threshold of 50%. While we are doing this, we will also change the URLs to access the optical imagery from "https://" to AWS S3 URLs ("S3://").

[10]:
```
def write_local_data_and_catalog_s3(catalog, bands, save_path, cloud_cover, s3_path=
↪"s3://"):
    '''Given path to a response json from a sat-api query, make a copy changing urls␣
↪to local paths'''
    creds = maap.aws.earthdata_s3_credentials('https://data.lpdaac.earthdatacloud.
↪nasa.gov/s3credentials')
    aws_session = boto3.session.Session(
        aws_access_key_id=creds['accessKeyId'],
        aws_secret_access_key=creds['secretAccessKey'],
        aws_session_token=creds['sessionToken'],
        region_name='us-west-2')
    s3 = aws_session.client('s3')

    with open(catalog) as f:
        clean_features = []
        asset_catalog = json.load(f)

        # Remove duplicate scenes
        features = asset_catalog['features']

        for feature in features:
            umm_json = feature['links'][6]['href']
            umm_data = !curl -i {umm_json}
            for line in umm_data:
                if "CLOUD_COVERAGE" in line:
                    cc_perc = (int(line.split("CLOUD_COVERAGE")[-1].split('"')[4]))
                    if cc_perc > cloud_cover:
                        pass
                    else:
                        try:
                            for band in bands:
                                output_file = feature['assets'][band]['href'].replace(
↪'https://data.lpdaac.earthdatacloud.nasa.gov/', s3_path)
                                bucket_name = output_file.split("/")[2]
                                s3_key = "/".join(output_file.split("/")[3:])
                                head = s3.head_object(Bucket = bucket_name, Key = s3_
↪key, RequestPayer='requester')
                                if head['ResponseMetadata']['HTTPStatusCode'] == 200:
                                    feature['assets'][band]['href'] = output_file
                            clean_features.append(feature)
                        except botocore.exceptions.ClientError as e:
                            if e.response['Error']['Code'] == "404":
                                print(f"The object does not exist. {output_file}")
                            else:
                                raise
        # save and updated catalog with local paths
        asset_catalog['features'] = clean_features
        local_catalog = catalog.replace('sample', 'local-s3')
```

(continues on next page)

(continued from previous page)

```
        with open(local_catalog,'w') as jsonfile:
            json.dump(asset_catalog, jsonfile)

        return local_catalog
```

When run, this will create a new JSON file that will only include files that meet our cloud cover threshold and have S3 URLs.

[11]:
```
local_cat = write_local_data_and_catalog_s3('./sample.json', ['B02','B03','B04','B05',
↪'B06','B07','Fmask'], './', 60, s3_path="s3://")
```

Now that we have images that meet our requirements, we will composite them into a multiband image for our AOI. We will composite the image on a band-by-band basis, so we first have to get a list of all the file paths for each band.

[12]:
```
def GetBandLists(inJSON, bandnum, tiles=['']):
    bands = dict({2:'B02', 3:'B03', 4:'B04', 5:'B05', 6:'B06', 7:'B07',8:'Fmask'})
    BandList = []
    with open(inJSON) as f:
        response = json.load(f)
    for i in range(len(response['features'])):
        try:
            getBand = response['features'][i]['assets'][bands[bandnum]]['href']
            # check 's3' is at position [:2]
            if getBand.startswith('s3', 0, 2):
                BandList.append(getBand)
        except Exception as e:
            print(e)

    BandList.sort()
    return BandList
```

We will build a band list of file paths for each image band. We will also access the 'fmask' band to mask out clouds.

[13]:
```
blue_bands = GetBandLists('./local-s3.json', 2)
green_bands = GetBandLists('./local-s3.json', 3)
red_bands = GetBandLists('./local-s3.json', 4)
nir_bands = GetBandLists('./local-s3.json', 5)
swir_bands = GetBandLists('./local-s3.json', 6)
swir2_bands = GetBandLists('./local-s3.json', 7)
fmask_bands = GetBandLists('./local-s3.json', 8)

print('number of each images in each band = ', len(blue_bands))
```

```
number of each images in each band =  21
```

2.1.7 Reading in HLS Data and Creating Composite

We will not read all of the HLS data, as we only need what's included in our AOI. To do this "windowed read" we need to know the dimensions, in pixels, of the window. To do this we need to convert our AOI to a projected coordinate system (UTM) and calculate the number of columns and rows we will need, using a pixel resolution of 30m.

[14]:
```
def get_shape(bbox, res=30):
    left, bottom, right, top = bbox
    width = int((right-left)/res)
    height = int((top-bottom)/res)
```

(continues on next page)

(continued from previous page)

```
    return height,width

# convert to m
AOI_utm = AOI.to_crs('epsg:32617')
height, width = get_shape(AOI_utm.bounds.iloc[0].to_list())
```

When building a maximum-NDVI composite, the first data we need is the red and NIR bands to make an NDVI band for each image. We will use 'riotiler' to perform a windowed read of our data. We will also read the 'fmask' layer as we can use this to mask out unwanted pixels.

[15]:
```
def ReadData(file, in_bbox, height, width, epsg="epsg:4326", dst_crs="epsg:4326"):
    '''Read a window of data from the raster matching the tile bbox'''
    with COGReader(file) as cog:
        img = cog.part(in_bbox, bounds_crs=epsg, max_size=None, dst_crs=dst_crs,␣
↪height=height, width=width)

    return (np.squeeze(img.as_masked().astype(np.float32)) * 0.0001)
```

Our AWS session has an expiry time. Now would be a good time to renew our access key to ensure we do not encounter any timeout issues.

[16]:
```
print('Getting AWS credentials...')
aws_session = get_aws_session_DAAC()
print('Finished')
```

```
Getting AWS credentials...
Finished
```

Using our renewed session, for each band we will read in the relevant band in each of our images, creating an array of image bands.

[17]:
```
with rio.Env(aws_session):
    print('reading red bands...')
    red_stack = np.array([ReadData(red_bands[i], AOI_bbox, height, width, epsg="epsg:
↪4326", dst_crs="epsg:4326") for i in range(len(red_bands))])
    print('reading nir bands...')
    nir_stack = np.array([ReadData(nir_bands[i], AOI_bbox, height, width, epsg="epsg:
↪4326", dst_crs="epsg:4326") for i in range(len(nir_bands))])
    print('reading fmask bands...')
    fmask_stack = np.array([ReadData(fmask_bands[i], AOI_bbox, height, width, epsg=
↪"epsg:4326", dst_crs="epsg:4326") for i in range(len(fmask_bands))])
    print('finished')

print("number of red_bands = ", np.shape(red_stack)[0])
```

```
reading red bands...
reading nir bands...
reading fmask bands...
finished
number of red_bands =  21
```

Now that we have our red band array and NIR band array we can make an NDVI image. While we do this, we will also apply our 'fmask' to remove any pixels that contain clouds.

[18]:
```
ndvi_stack =  np.ma.array(np.where(((fmask_stack==1)), -9999, (nir_stack-red_stack)/
↪(nir_stack+red_stack)))
ndvi_stack = np.where((~np.isfinite(ndvi_stack)) | (ndvi_stack>1), -9999, ndvi_stack)
```

```
/opt/conda/lib/python3.7/site-packages/ipykernel_launcher.py:1: RuntimeWarning:␣
↪divide by zero encountered in true_divide
  """Entry point for launching an IPython kernel.
```

At this point, we can plot our images and see which ones contain cloud coverage. These images have gaps where there is no data or cloud coverage.

[19]:
```
fig, axes = plt.subplots(3,7, figsize=(33,30))

for i, ax in enumerate(axes.flat):
    ndvi_stack[i] = np.where((ndvi_stack[i]>1) | (ndvi_stack[i]<-1), 0, ndvi_stack[i])
    ax.axes.xaxis.set_visible(False)
    ax.axes.yaxis.set_visible(False)
    ax.imshow(ndvi_stack[i], cmap='viridis', clim=(0.1, 0.5))
```

Now that we have a stack of NDVI bands, we can create an index band that maps the pixel position from each band where the NDVI value is greatest. We can use this to locate the pixels we want to use in our composite.

[20]:
```
# Create Bool mask where there is no value in any of the NDVI layers
print("Make NDVI valid mask")
print("shape:\t\t", np.ma.array(ndvi_stack).shape)
```

(continues on next page)

(continued from previous page)

```
MaxNDVI = np.ma.max(np.ma.array(ndvi_stack),axis=0)
BoolMask = np.ma.getmask(MaxNDVI)
del MaxNDVI

## Get the argmax index positions from the stack of NDVI images
print("Get stack nan mask")
ndvi_stack = np.ma.array(ndvi_stack)
print("Calculate Stack max NDVI image")
NDVImax = np.nanargmax(ndvi_stack,axis=0)
## create a tmp array (binary mask) of the same input shape
NDVItmp = np.ma.zeros(ndvi_stack.shape, dtype=bool)

## for each dimension assign the index position (flattens the array to a LUT)
print("Create LUT of max NDVI positions")
for i in range(np.shape(ndvi_stack)[0]):
    NDVItmp[i,:,:]=NDVImax==i
```

```
Make NDVI valid mask
shape:            (21, 3006, 2951)
Get stack nan mask
Calculate Stack max NDVI image
Create LUT of max NDVI positions
```

Now that we have our NDVI lookup table and a list of all the images for each band, we can make composites based on the maximum NDVI value. For this, we will use the following two functions:

[21]:
```
def CollapseBands(inArr, NDVItmp, BoolMask):
    inArr = np.ma.masked_equal(inArr, 0)
    inArr[np.logical_not(NDVItmp)]=0
    compImg = np.ma.masked_array(inArr.sum(0), BoolMask)
    #print(compImg)
    return compImg

def CreateComposite(file_list, NDVItmp, BoolMask, in_bbox, height, width, epsg, dst_
↪crs):
    MaskedFile = [ReadData(file_list[i], in_bbox, height, width, epsg, dst_crs) for i␣
↪in range(len(file_list))]
    Composite=CollapseBands(MaskedFile, NDVItmp, BoolMask)
    return Composite
```

For each band, we will read all the images (for the required band) and create a composite based on the maximum NDVI value.

[22]:
```
aws_session = get_aws_session_DAAC()
with rio.Env(aws_session):
    print('Creating Blue Composite')
    blue_comp = CreateComposite(blue_bands, NDVItmp, BoolMask, AOI_bbox, height,␣
↪width, "epsg:4326", "epsg:4326")
    print('Creating Green Composite')
    green_comp = CreateComposite(green_bands, NDVItmp, BoolMask, AOI_bbox, height,␣
↪width, "epsg:4326", "epsg:4326")
    print('Creating Red Composite')
    red_comp = CreateComposite(red_bands, NDVItmp, BoolMask, AOI_bbox, height, width,
↪"epsg:4326", "epsg:4326")
    print('Creating NIR Composite')
    nir_comp = CreateComposite(nir_bands, NDVItmp, BoolMask, AOI_bbox, height, width,
↪"epsg:4326", "epsg:4326")
```

(continues on next page)

(continued from previous page)

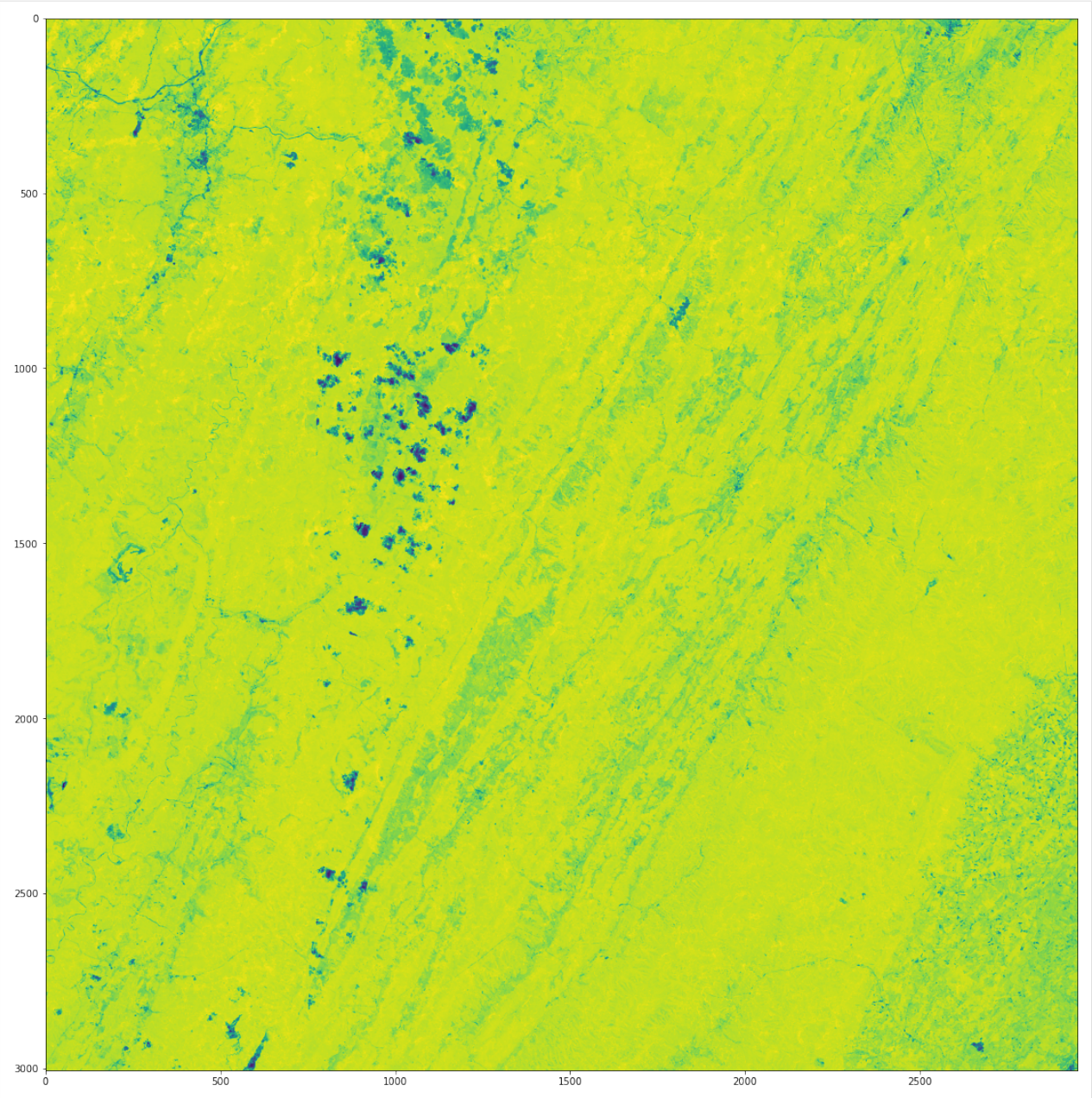
```
    print('Creating SWIR Composite')
    swir_comp = CreateComposite(swir_bands, NDVItmp, BoolMask, AOI_bbox, height,␣
↪width, "epsg:4326", "epsg:4326")
    print('Creating SWIR2 Composite')
    swir2_comp = CreateComposite(swir2_bands, NDVItmp, BoolMask, AOI_bbox, height,␣
↪width, "epsg:4326", "epsg:4326")
    print('Creating NDVI Composite')
    ndvi_comp = CollapseBands(ndvi_stack, NDVItmp, BoolMask)
    print('Creating fmask Composite')
    fmask_comp = CollapseBands(fmask_stack, NDVItmp, BoolMask)
```

```
Creating Blue Composite
Creating Green Composite
Creating Red Composite
Creating NIR Composite
Creating SWIR Composite
Creating SWIR2 Composite
Creating NDVI Composite
Creating fmask Composite
```

We can look at our NDVI composite image and see we now have a complete image for our AOI.

[23]:
```
fig, axes = plt.subplots(1,1, figsize=(20,20))
ax.axes.xaxis.set_visible(False)
ax.axes.yaxis.set_visible(False)
plt.imshow(np.where(fmask_comp==1, -9999, ndvi_comp))
```

[23]:
```
<matplotlib.image.AxesImage at 0x7f223c43f6d0>
```



Now that we have a 7-band composite image, we can use these bands to calculate a suite of common vegetation indices using the following functions. These indices will give us a better look at vegetation health by giving us information on vegetation water content, greenness, and more.

[24]:
```
# SAVI
def calcSAVI(red, nir):
    savi = ((nir - red)/(nir + red + 0.5))*(1.5)
    print('\tSAVI Created')
    return savi

# NDMI
def calcNDMI(nir, swir):
    ndmi = (nir - swir)/(nir + swir)
```

(continues on next page)

(continued from previous page)

```
    print('\tNDMI Created')
    return ndmi

# EVI
def calcEVI(blue, red, nir):
    evi = 2.5 * ((nir - red) / (nir + 6 * red - 7.5 * blue + 1))
    print('\tEVI Created')
    return evi

# NBR
def calcNBR(nir, swir2):
    nbr = (nir - swir2)/(nir + swir2)
    print('\tNBR Created')
    return nbr

# MSAVI
def calcMSAVI(red, nir):
    msavi = (2 * nir + 1 - np.sqrt((2 * nir + 1)**2 - 8 * (nir - red))) / 2
    print('\tMSAVI Created')
    return msavi
```

We can call these functions and make our additional indices.

[25]:
```
# calculate covars
print("Generating covariates")
SAVI = calcSAVI(red_comp, nir_comp)
#print("NDMI")
NDMI = calcNDMI(nir_comp, swir_comp)
#print("EVI")
EVI = calcEVI(blue_comp, red_comp, nir_comp)
#print("NBR")
NBR = calcNBR(nir_comp, swir2_comp)
MSAVI = calcMSAVI(red_comp, nir_comp)
```

```
Generating covariates
        SAVI Created
        NDMI Created
        EVI Created
        NBR Created
        MSAVI Created
```

We have a suite of 12 bands now, and we can merge them together into a single 12-band image stack.

[26]:
```
print("\nCreating raster stack...\n")
stack = np.transpose([blue_comp, green_comp, red_comp, nir_comp, swir_comp, swir2_
↪comp, ndvi_comp, SAVI, MSAVI, NDMI, EVI, NBR], [0, 1, 2])
stack = np.where(fmask_comp==1, -9999, stack)
print(np.shape(stack))
```

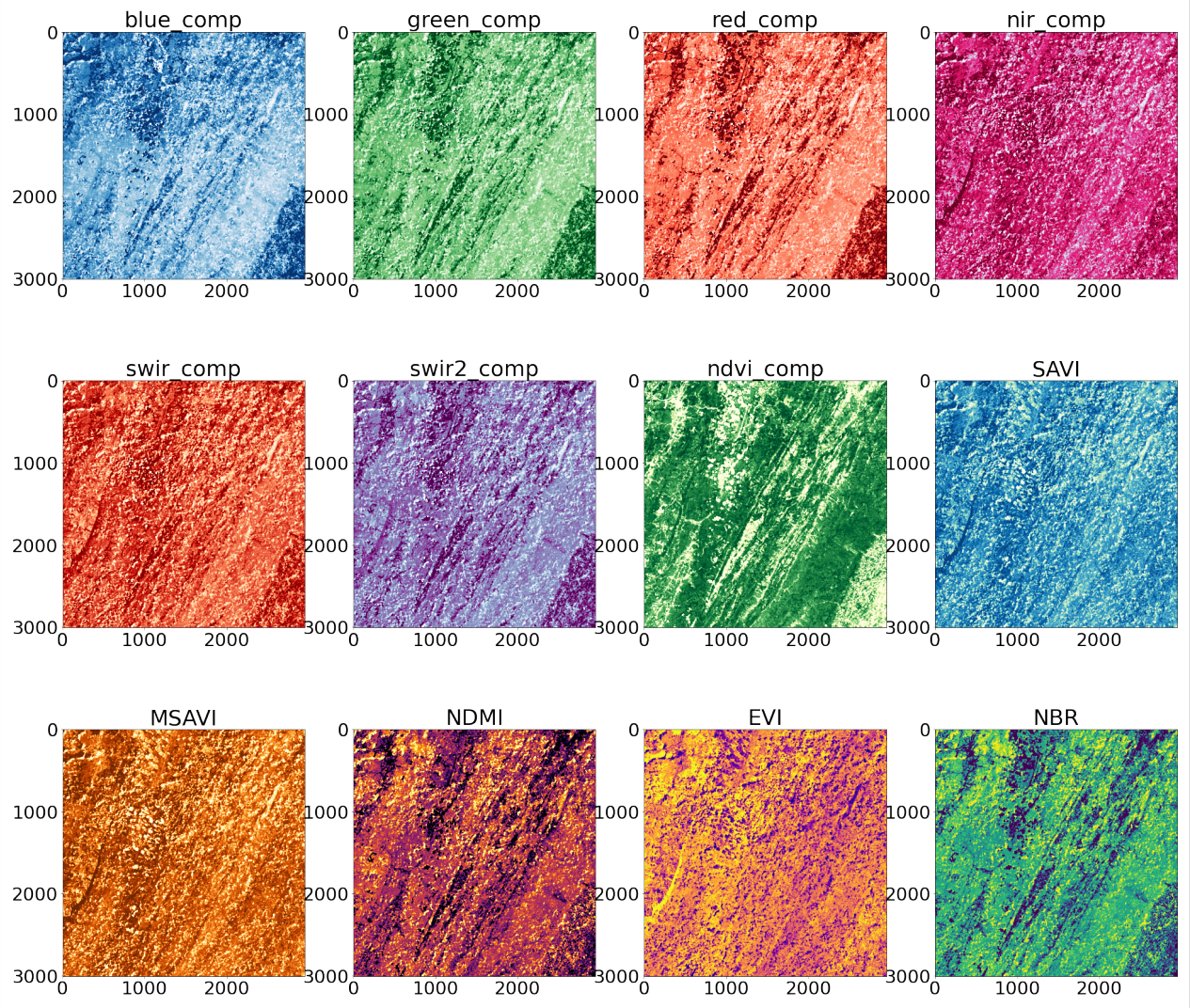
```

Creating raster stack...

(12, 3006, 2951)
```

2.1.8 Display Results

We can look at each of these bands by using 'matplotlib' to plot each one individually.

[27]:
```
n: int = len(stack)
#topo_cmaps = ["bone","Spectral", "magma", "RdBu", "coolwarm"]
topo_cmaps = ['Blues','Greens','Reds','PuRd','OrRd','BuPu','YlGn','GnBu','YlOrBr',
↪'inferno','plasma', 'viridis']
print(stack.shape)
bandnames = ['blue_comp', 'green_comp', 'red_comp', 'nir_comp', 'swir_comp', 'swir2_
↪comp', 'ndvi_comp', 'SAVI', 'MSAVI', 'NDMI', 'EVI', 'NBR']

font = {'size'   : 30}
matplotlib.rc('font', **font)

#axs = [[0,0],[0,1],[0,2],[0,3],[0,4],[0,5],[1,0],[1,1],[1,2],[1,3],[1,4],[1,5],[2,0],
↪[2,1],[2,2],[2,3],[2,4],[2,5]]
fig, axes = plt.subplots(3,4, figsize=(33,30))
print(axes.flat)
for i, ax in enumerate(axes.flat):
    ax.imshow(stack[i], cmap=topo_cmaps[i], clim=(np.percentile(stack[i], 10), np.
↪percentile(stack[i], 90)))
    ax.set_title(bandnames[i])
```

```
(12, 3006, 2951)
<numpy.flatiter object at 0x561c7d6fe260>
```



We can also visualize our composite NDVI band on our interactive 'folium' map. You can see that even though we found multiple images, by using a windowed read we were able to just read and process the data we needed.

[28]:
```
m = folium.Map(location=[38.6, -78.5], zoom_start=9, tiles='CartoDB positron')
AOI_bx = AOI.bounds
#folium.GeoJson(AOI, style_function=lambda x: {'fillColor': 'orange','opacity':0}).
↪add_to(m)
geo_r = folium.raster_layers.ImageOverlay(np.ma.getdata(ndvi_comp), opacity=1,␣
↪bounds=[[AOI_bx['miny'][0],AOI_bx['minx'][0]],[AOI_bx['maxy'][0],AOI_bx['maxx
↪'][0]]])
geo_r.add_to(m)

m
```

[28]:
```
<folium.folium.Map at 0x7f236144a950>
```

2.2 GEDI_L2A Search and Visualize

This tutorial aims to provide information and code to help users get started working with the Global Ecosystem Dynamics Investigation (GEDI) Level 2A (GEDI02_A) product using the MAAP. Information about the GEDI02_A product may be found at the Data Set Landing Page. We will start by importing the packages which will allow us to search for, access, explore, and visualize GEDI02_A product data.

Note: This Jupyter notebook utilizes the `folium` and `pystac_client` packages. If you do not have these packages installed, uncomment the lines and run the following code block.

[1]:
```
# !pip install folium
# !pip install pystac_client
```

For this tutorial, we will import `boto3`, `folium`, `h5py`, `pandas`, `exists` from `os.path` and `Client` from `pystac_client` as shown in the following codeblock.

[2]:
```
# Install packages
import boto3
import folium
import h5py
import os
import pandas as pd
from maap.maap import MAAP
from os.path import exists
from pystac_client import Client
```

2.2.1 Searching for and accessing GEDI02_A data

As of the time of the writing of this tutorial (2/10/23), two recommended ways for searching and accessing GEDI02_A data for use on the MAAP ADE are through the `maap-stac` as well as through NASA's Common Metadata Repository (CMR). The methods for using these two ways are different and documented in the following two sub-sections.

Via `maap-stac`

To search for data from the GEDI02_A product, we will use the `Client` package to open the `maap-stac` URL <https://stac.maap-project.org/>`__ and assign this to a variable (`client` in this case).

[3]:
```
# Open the maap-stac URL with the Client package
URL = 'https://stac.maap-project.org/'
client = Client.open(URL)
```

Now we can use the client specified above to search for data within the GEDI02_A product. Let's search for the first item that is found in the GEDI02_A collection and assign this to a variable (`search` in this case).

[4]:
```
collection = 'GEDI02_A' # assign collection name
# Search for 1st item found in the collection
search = client.search(
    max_items = 1,
    collections = collection,
)
```

Let's inspect this item using the `get_all_items()` function.

```
[5]: # Inspect first item
item = search.get_all_items()[0]
item
```

```
[5]: <Item id=GEDI02_A_2021272190541_O15849_04_T03030_02_003_02_V002>
```

After running the code block above, you should receive an output such as the output shown above. Click the arrow next to the item to see information such as the ID, bounding box coordinates, datetime, and more. In order to access the data, we will use the `item` variable from the above section in order to extract the necessary information to set and display `bucket`, `key`, and `filename` variables. These will be useful to use as arguments when downloading the file.

```
[6]: # Use the item variable to extract information about the bucket, key, and file name
href = item.assets['data'].href
path_parts = href.split('/')
bucket = path_parts[2]
key = href.split(bucket)[1][1:]
filename = path_parts[-1]
# Display arguments
bucket, key, filename
```

```
[6]: ('nasa-maap-data-store',
 'file-staging/nasa-map/GEDI02_A___002/2021.09.29/GEDI02_A_2021272190541_O15849_04_
↪T03030_02_003_02_V002.h5',
 'GEDI02_A_2021272190541_O15849_04_T03030_02_003_02_V002.h5')
```

Now let's set an `s3` variable using the `boto3.client` function. We can use the function `download_file` along with the arguments we set in the previous block to download the GEDI02_A data we need.

```
[7]: # Set s3 variable
s3 = boto3.client('s3')
# If file already exists, do not download file
if exists(filename):
    print("File already downloaded")
# Otherwise, download the file
else:
    s3.download_file(bucket, key, filename)
    print("Finished downloading")
```

```
Finished downloading
```

After the previous block has finished running, we should see the message `Finished downloading` and the file should appear in the same directory that the Jupyter notebook is in.

Via NASA's CMR

To search for data from the GEDI02_A product using NASA's CMR, we invoke the `MAAP` constructor, setting the `maap_host` argument to `'api.ops.maap-project.org'`.

```
[8]: # Invoke the MAAP using the MAAP host argument
maap = MAAP(maap_host='api.ops.maap-project.org')
```

Now we can use the `searchGranule` function to find granule data within the collection, using the collection short name "GEDI_02A". Note that we can use `searchGranule`'s `cmr_host` argument to specify a CMR instance external to MAAP and the `readable_granule_name` argument to find granules matching either granule UR or producer granule id (please see the API documentation for more information). In order to download data from NASA's CMR, we will set a variable to the first result from the `results` we obtained.

[9]:
```
# Search for granule data using CMR host name and collection short name, and readable_
↪granule_name arguments
results = maap.searchGranule(
    cmr_host='cmr.earthdata.nasa.gov',
    short_name='GEDI02_A',
    readable_granule_name = "GEDI02_A_2021272190541_O15849_04_T03030_02_003_02_V002.h5
↪")
# Download first result
filename = results[0].getData()
```

If desired, the `print` function can be utilized to see the file name and directory.

[10]:
```
# Print file directory
print(filename)
```

```
./GEDI02_A_2021272190541_O15849_04_T03030_02_003_02_V002.h5
```

2.2.2 Explore

Now that we have downloaded the data, let's look into what it contains.

[11]:
```
# Create variable containing info from the file we downloaded
gediL2A = h5py.File(filename, 'r')
```

GEDI02_A data has data for 8 different beams. Let's create a list of beam names to help explore the data.

[12]:
```
# Create list of beam names
beamNames = [g for g in gediL2A.keys() if g.startswith('BEAM')]
beamNames
```

[12]:
```
['BEAM0000',
 'BEAM0001',
 'BEAM0010',
 'BEAM0011',
 'BEAM0101',
 'BEAM0110',
 'BEAM1000',
 'BEAM1011']
```

Now let's explore the information available for one of the beams (in this case 'BEAM0000').

[13]:
```
# Get list of objects in the data pertaining to 'BEAM0000'
beam = beamNames[0]
gediL2A_objs = []
gediL2A.visit(gediL2A_objs.append)
gediSDS = [o for o in gediL2A_objs if isinstance(gediL2A[o], h5py.Dataset)]
[i for i in gediSDS if beam in i][0:20]
```

[13]:
```
['BEAM0000/ancillary/l2a_alg_count',
 'BEAM0000/beam',
 'BEAM0000/channel',
 'BEAM0000/degrade_flag',
 'BEAM0000/delta_time',
 'BEAM0000/digital_elevation_model',
 'BEAM0000/digital_elevation_model_srtm',
 'BEAM0000/elev_highestreturn',
 'BEAM0000/elev_lowestmode',
```

(continues on next page)

(continued from previous page)

```
 'BEAM0000/elevation_bias_flag',
 'BEAM0000/elevation_bin0_error',
 'BEAM0000/energy_total',
 'BEAM0000/geolocation/elev_highestreturn_a1',
 'BEAM0000/geolocation/elev_highestreturn_a2',
 'BEAM0000/geolocation/elev_highestreturn_a3',
 'BEAM0000/geolocation/elev_highestreturn_a4',
 'BEAM0000/geolocation/elev_highestreturn_a5',
 'BEAM0000/geolocation/elev_highestreturn_a6',
 'BEAM0000/geolocation/elev_lowestmode_a1',
 'BEAM0000/geolocation/elev_lowestmode_a2']
```

2.2.3 Visualize

Now that we've seen the various labels within the /BEAM0000 group, let's use this information to visualize the GEDI orbit path for our scenes. To start, we shall get samples for various shots, the beam number, longitude, latitude, and quality flags. We can use these samples to create and display a `pandas` dataframe.

[14]:
```
# Set variables for shot, beam number, longitude, latitude, and quality flag samples
lonSample, latSample, shotSample, qualitySample, beamSample = [], [], [], [], []
lats = gediL2A[f'{beamNames[0]}/lat_lowestmode'][()]
lons = gediL2A[f'{beamNames[0]}/lon_lowestmode'][()]
shots = gediL2A[f'{beamNames[0]}/shot_number'][()]
quality = gediL2A[f'{beamNames[0]}/quality_flag'][()]
for i in range(len(shots)):
    if i % 100 == 0:
        shotSample.append(str(shots[i]))
        lonSample.append(lons[i])
        latSample.append(lats[i])
        qualitySample.append(quality[i])
        beamSample.append(beamNames[0])
# Create a pandas dataframe containing the sample information
latslons = pd.DataFrame({'Beam': beamSample, 'Shot Number': shotSample, 'Longitude':␣
↪lonSample,
                         'Latitude': latSample, 'Quality Flag': qualitySample})
# Display the dataframe
latslons
```

[14]:

	Beam	Shot Number	Longitude	Latitude	Quality Flag
0	BEAM0000	158490000400502383	52.337698	0.599011	0
1	BEAM0000	158490000400502483	52.367554	0.556129	0
2	BEAM0000	158490000400502583	52.396932	0.514800	0
3	BEAM0000	158490000400502683	52.426548	0.472746	0
4	BEAM0000	158490000400502783	52.456248	0.430695	0
...	...	...	...	...	...
1685	BEAM0000	158490000400670883	135.837810	-51.605016	0
1686	BEAM0000	158490000400670983	135.921291	-51.605653	0
1687	BEAM0000	158490000400671083	136.003206	-51.606299	0
1688	BEAM0000	158490000400671183	136.086629	-51.606741	0
1689	BEAM0000	158490000400671283	136.168523	-51.607159	0

[1690 rows x 5 columns]

We can now create a map of the orbit path using the dataframe that we have created and utilizing the `folium.Map` and `folium.Circle` functions. Include `map` in the code block to inspect the map which is created within the Jupyter notebook.

[15]:
```
# Create a map
map = folium.Map(
    location=[
        latslons.Latitude.mean(),
        latslons.Longitude.mean()
    ], zoom_start=3, control_scale=True
)
# Add variables to the map
for index, location_info in latslons.iterrows():
    folium.Circle(
        [location_info["Latitude"], location_info["Longitude"]],
        popup=f"Shot Number: {location_info['Shot Number']}"
    ).add_to(map)
# Display map
map
```

[15]:
```
<folium.folium.Map at 0x7f04c05a8690>
```

References:

- https://lpdaac.usgs.gov/resources/e-learning/getting-started-gedi-l2a-version-2-data-python/
- https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html

2.3 GEDI_L2B Search and Visualize

Authors: Samuel Ayers (UAH), Sumant Jha (MSFC/USRA), Anish Bhusal (UAH), Alex Mandel (DevSeed), Aimee Barciauskas (DevSeed)

Date: December 19, 2022

Description: In this tutorial, we will use the integrated Earthdata search function in MAAP Algorithm Development Environment (ADE) to search for GEDI L2B data for an area of interest. We will then download some of this data and read its attributes in preparation for some analysis. We will perform a spatial subset on the data to reduce data volumes, and then make some basic plots of our data.

2.3.1 Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

2.3.2 About the Data

GEDI L2B Canopy Cover and Vertical Profile Metrics Data Global Footprint Level V002

This dataset provides Global Ecosystem Dynamics Investigation (GEDI) Level 2 (L2) data, which has the purpose of extracting biophysical metrics and consists of Canopy Cover and Vertical Profile Metrics. These metrics are derived from the L1B waveform, and include canopy cover, Plant Area Index (PAI), Plant Area Volume Density (PAVD), and Foliage Height Diversity (FHD). GEDI is attached to the International Space Station (ISS) and collects data globally between 51.6° N and 51.6° S latitudes at the highest resolution and densest sampling of any light detection and ranging (lidar) instrument in orbit to date; specifically, GEDI L2B data has a spatial resolution of 25m. (Source: GEDI L2B CMR Search)

2.3.3 Additional Resources

- GEDI_L2B Version 2 Dataset Landing Page
- GEDI Level 2 Version 2 User Guide
- The GEDI Website

2.3.4 Importing and Installing Packages

We will begin by installing any packages we need and importing the packages that we will use.

Prerequisites

- geopandas
- folium

[ ]:
```
# Uncomment the following lines to install these packages if you haven't already.
# !pip install geopandas
# !pip install folium
```

[38]:
```
from maap.maap import MAAP
maap = MAAP(maap_host='api.maap-project.org')
import geopandas as gpd
import folium
import h5py
import pandas
import matplotlib
import matplotlib.pyplot as plt
import shapely
import os
```

2.3.5 Search Data Using an AOI

We will search and download GEDI L2B data using the bounding box of a vector AOI. Firstly, an AOI over the Shenandoah National Park will be created and then we will plot its location on a map.

[39]:
```
# Using bounding coordinates to create a polygon around Shenandoah National Park
lon_coords = [-78.32129105072025, -78.04618813890727, -78.72985973163064, -79.
↪0158578082679, -78.32129105072025]
lat_coords = [38.88703610703791, 38.74909216350823, 37.88789051477522, 38.
↪03177640342157, 38.88703610703791]

polygon_geom = shapely.geometry.polygon.Polygon(zip(lon_coords, lat_coords))
crs = 'epsg:4326'
AOI = gpd.GeoDataFrame(index=[0], crs=crs, geometry=[polygon_geom])
```

We can get the bounding box of the AOI so we can use it as a spatial limit on our data search. GeoPandas has a function for returning the spatial coordinates of a bounding box:

[40]:
```
# Get the bounding box of the shp
bbox = AOI.bounds
# print the bounding box coords
print('minx = ', bbox['minx'][0])
print('miny = ', bbox['miny'][0])
```

(continues on next page)

(continued from previous page)

```
print('maxx = ', bbox['maxx'][0])
print('maxy = ', bbox['maxy'][0])
```

```
minx =  -79.0158578082679
miny =  37.88789051477522
maxx =  -78.04618813890727
maxy =  38.88703610703791
```

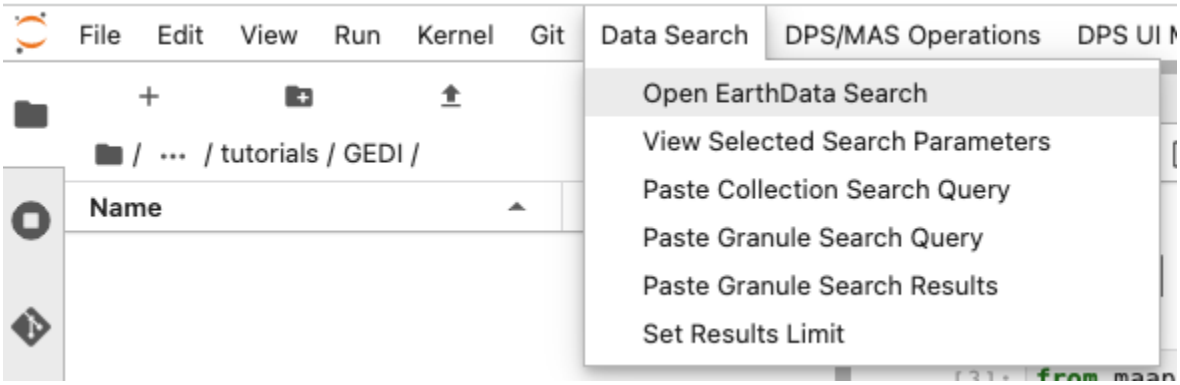
Let's look at our AOI on an interactive map using folium.

[41]:
```
m = folium.Map(location=[38.5, -78], zoom_start=9, tiles='CartoDB positron')
geo_j = folium.GeoJson(data=AOI, style_function=lambda x: {'fillColor': 'orange'})
geo_j.add_to(m)
m
```

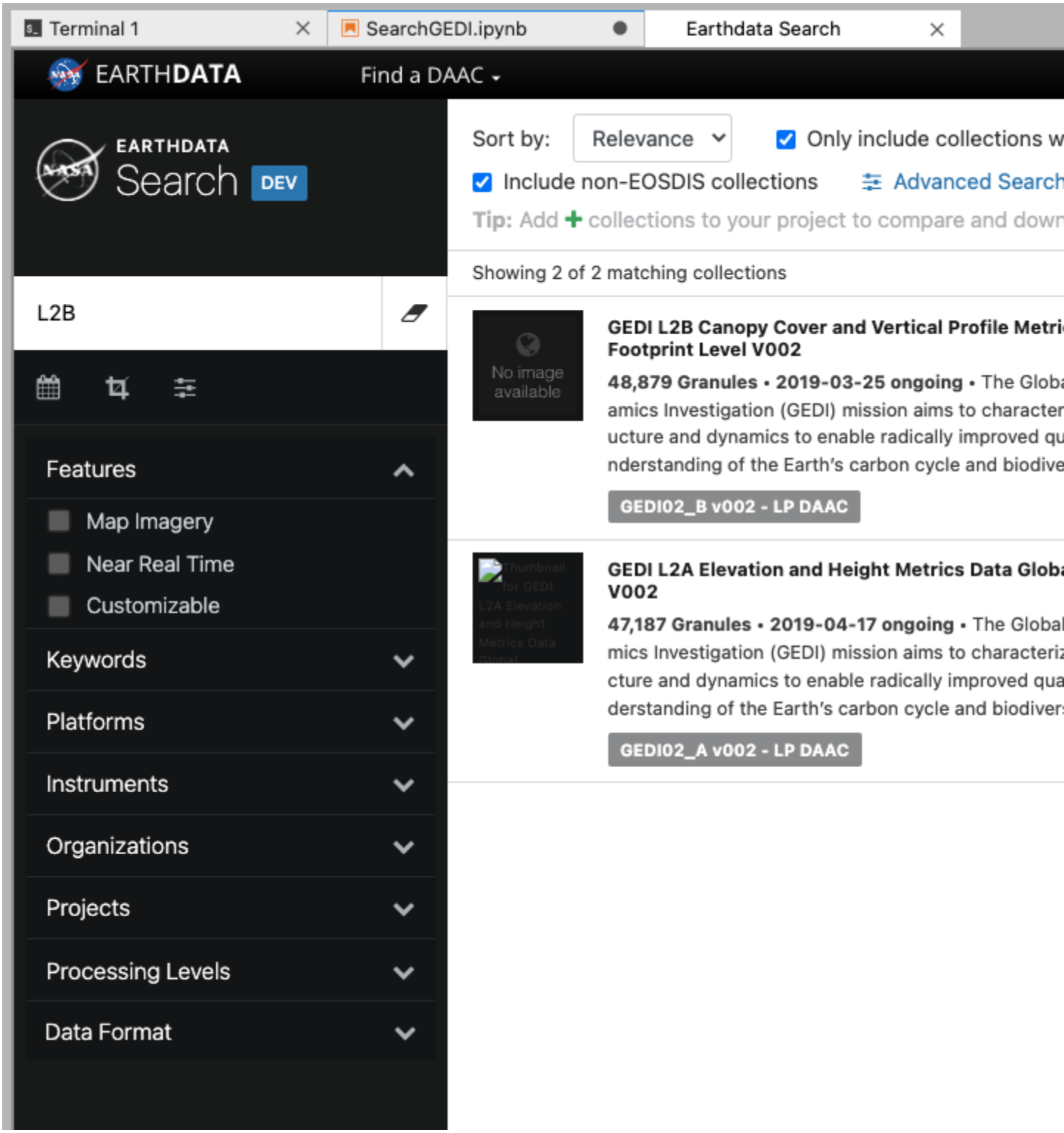
[41]:
```
<folium.folium.Map at 0x7fa48a66b710>
```

2.3.6 Search using the EarthData Search Integration

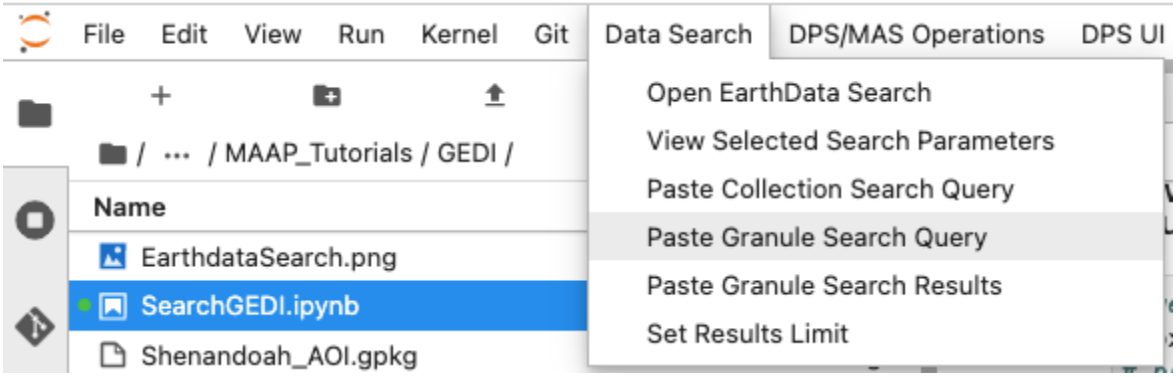
Search Data: To search GEDI data we can use the EarthData search integration in the MAAP ADE. Open the Earthdata search toolbar by clicking on the following:



This will open up the EarthData search interface in a new tab. We can use the search bar to search GEDI L2B data. By entering "L2B" in the search bar, we can see the relevant GEDI data is filtered. Click on the dataset to get more details.



While we have been searching for data, the MAAP ADE has been keeping a track of our search parameters. This means that we can easily insert our search for GEDI data straight into our notebook.



2.3.7 Inspect and Filter through the Data

This gives us our search parameters in our notebook using the GEDI dataset ID. The default limit for the number of returned results is 1000. Running this will produce 1000 results, but we can look at the first one by indexing the list of returned results. This is what the data entry looks like. You can see a large amount of metadata for the file along with the URL for where this specific file is stored.

[42]:
```
data = maap.searchGranule(cmr_host='cmr.earthdata.nasa.gov', concept_id="C1908350066-
↪LPDAAC_ECS", limit=1000)[0]
data
```

[42]:
```
{'concept-id': 'G2011069580-LPDAAC_ECS',
 'collection-concept-id': 'C1908350066-LPDAAC_ECS',
 'revision-id': '10',
 'format': 'application/echo10+xml',
 'Granule': {'GranuleUR': 'SC:GEDI02_B.002:2433465280',
  'InsertTime': '2021-02-22T19:16:30.675Z',
  'LastUpdate': '2021-09-16T13:36:07.965Z',
  'Collection': {'DataSetId': 'GEDI L2B Canopy Cover and Vertical Profile Metrics␣
↪Data Global Footprint Level V002'},
  'DataGranule': {'SizeMBDataGranule': '16.5413',
   'ProducerGranuleId': 'GEDI02_B_2019108002012_O01959_01_T03909_02_003_01_V002.h5',
   'DayNightFlag': 'UNSPECIFIED',
   'ProductionDateTime': '2021-02-21T14:45:08Z'},
  'PGEVersionClass': {'PGEVersion': '003'},
  'Temporal': {'RangeDateTime': {'BeginningDateTime': '2019-04-18T00:20:12.000000Z',
    'EndingDateTime': '2019-04-18T01:52:53.000000Z'}},
  'Spatial': {'HorizontalSpatialDomain': {'Geometry': {'GPolygon': {'Boundary': {
↪'Point': [{'PointLongitude': '80.2890335089',
         'PointLatitude': '-4.6168623465'},
        {'PointLongitude': '82.4542052313', 'PointLatitude': '-1.5568021223'},
        {'PointLongitude': '83.6402279084', 'PointLatitude': '0.1277767863'},
        {'PointLongitude': '83.6814118126', 'PointLatitude': '0.0987901632'},
        {'PointLongitude': '82.4953794073', 'PointLatitude': '-1.5858244175'},
        {'PointLongitude': '80.3302671214',
         'PointLatitude': '-4.6459691487'}]}}}}},
  'OrbitCalculatedSpatialDomains': {'OrbitCalculatedSpatialDomain': {'StartOrbitNumber
↪': '1959',
    'StopOrbitNumber': '1959'}},
  'MeasuredParameters': {'MeasuredParameter': {'ParameterName': 'Level2B'}},
  'Platforms': {'Platform': {'ShortName': 'ISS',
    'Instruments': {'Instrument': {'ShortName': 'GEDI',
      'Sensors': {'Sensor': {'ShortName': 'LIDAR'}}}}}},
```

(continues on next page)

(continued from previous page)

```
  'Campaigns': {'Campaign': {'ShortName': 'GEDI'}},
  'AdditionalAttributes': {'AdditionalAttribute': [{'Name': 'identifier_product_doi',
     'Values': {'Value': '10.5067/GEDI/GEDI02_B.002'}},
    {'Name': 'identifier_product_doi_authority',
     'Values': {'Value': 'https://doi.org'}},
    {'Name': 'Reference_Ground_Track', 'Values': {'Value': '3909'}},
    {'Name': 'SEGMENT_NUMBER', 'Values': {'Value': '01'}}]},
  'InputGranules': {'InputGranule': ['GEDI01_B_2019108002012_O01959_01_T03909_02_005_
↪01_V002.h5',
    'GEDI02_A_2019108002012_O01959_01_T03909_02_003_01_V002_10algs.h5']},
  'OnlineAccessURLs': {'OnlineAccessURL': {'URL': 'https://e4ftl01.cr.usgs.gov//GEDI_
↪L1_L2/GEDI/GEDI02_B.002/2019.04.18/GEDI02_B_2019108002012_O01959_01_T03909_02_003_
↪01_V002.h5',
    'URLDescription': 'GEDI02_B_2019108002012_O01959_01_T03909_02_003_01_V002.h5.␣
↪MimeType: application/x-hdfeos',
    'MimeType': 'application/x-hdfeos'}},
  'OnlineResources': {'OnlineResource': [{'URL': 'https://doi.org/10.5067/GEDI/GEDI02_
↪B.001 ',
     'Description': 'The Landing Page for this file may be accessed directly from␣
↪this link',
     'Type': 'DOI',
     'MimeType': 'text/html'},
    {'URL': 'https://e4ftl01.cr.usgs.gov//WORKING/BRWS/Browse.001/2021.02.23/GEDI02_B_
↪2019108002012_O01959_01_T03909_02_003_01_V002.png',
     'Description': 'This Browse file may be downloaded directly from this link',
     'Type': 'BROWSE',
     'MimeType': 'image/jpeg'},
    {'URL': 'https://e4ftl01.cr.usgs.gov//GEDI_L1_L2/GEDI/GEDI02_B.002/2019.04.18/
↪GEDI02_B_2019108002012_O01959_01_T03909_02_003_01_V002.h5.xml',
     'Description': 'This Metadata file may be downloaded directly from this link',
     'Type': 'EXTENDED METADATA',
     'MimeType': 'text/xml'}]},
  'Orderable': 'true',
  'DataFormat': 'HDF5',
  'Visible': 'true'}}
```

So far, this search function requests all of the GEDI data but we can add a spatial subset filter using our AOI from above to limit the results. Adding a spatial filter returns 259 GEDI L2B files that intersect with our AOI.

[43]:
```
data_aoi = maap.searchGranule(cmr_host='cmr.earthdata.nasa.gov', concept_id=
↪"C1908350066-LPDAAC_ECS", bounding_box="-79.0158578082679,37.88789051477522,-78.
↪04618813890727,38.887036107037915", limit=1000)
print(len(data_aoi))
```

```
259
```

This is more data than we need, so let's look at the contents of a single GEDI file. Firstly we need to bring it from the server side (S3) to our local side. This can be done using the MAAP function getData. We'll create a new data directory, and then we will pull the 7th file in our search results.

[44]:
```
# set data directory
dataDir = './data'

# check if directory exists -> if directory doesn't exist, directory is created
if not os.path.exists(dataDir):
    os.mkdir(dataDir)
```

[45]:
```
# pulling the 7th file into the new directory
gedi_data = data_aoi[6].getData(dataDir)
print(gedi_data)
```

```
./data/GEDI02_B_2019145051352_O02537_03_T04809_02_003_01_V002.h5
```

GEDI data has 8 beams. So, we will check that all beams are in our file and print a list of the available beams.

[46]:
```
gedi_h5_file = h5py.File(gedi_data, 'r')
gedi_keys = list(gedi_h5_file.keys())
gedi_beams = ['BEAM0000', 'BEAM0001', 'BEAM0010', 'BEAM0011', 'BEAM0101', 'BEAM0110',
↪'BEAM1000', 'BEAM1011']
gedi_beams_lst = []
for gedi_beam_name in gedi_keys:
    if gedi_beam_name in gedi_beams:
        gedi_beams_lst.append(gedi_beam_name)
print(gedi_beams_lst)
```

```
['BEAM0000', 'BEAM0001', 'BEAM0010', 'BEAM0011', 'BEAM0101', 'BEAM0110', 'BEAM1000',
↪'BEAM1011']
```

For each beam, we need to get all of the data and metrics associated with it. For this, we have a function that will gather all of the metrics we want and put them into a geopandas dataframe:

[47]:
```
def get_gedi_df(gedi_h5_file, gedi_beam_name):
    gedi_beam = gedi_h5_file[gedi_beam_name]

    # Get location info.
    gedi_beam_geoloc = gedi_beam['geolocation']
    # Get land cover data.
    gedi_beam_landcover = gedi_beam['land_cover_data']

    gedi_beam_df = pandas.DataFrame(
            {'elevation_bin0'          : gedi_beam_geoloc['elevation_bin0'],
             'elevation_lastbin'       : gedi_beam_geoloc['elevation_lastbin'],
             'height_bin0'             : gedi_beam_geoloc['height_bin0'],
             'height_lastbin'          : gedi_beam_geoloc['height_lastbin'],
             'shot_number'             : gedi_beam_geoloc['shot_number'],
             'solar_azimuth'           : gedi_beam_geoloc['solar_azimuth'],
             'solar_elevation'         : gedi_beam_geoloc['solar_elevation'],
             'latitude_bin0'           : gedi_beam_geoloc['latitude_bin0'],
             'latitude_lastbin'        : gedi_beam_geoloc['latitude_lastbin'],
             'longitude_bin0'          : gedi_beam_geoloc['longitude_bin0'],
             'longitude_lastbin'       : gedi_beam_geoloc['longitude_lastbin'],
             'degrade_flag'            : gedi_beam_geoloc['degrade_flag'],
             'digital_elevation_model' : gedi_beam_geoloc['digital_elevation_model'],
             'landsat_treecover'       : gedi_beam_landcover['landsat_treecover'],
             'modis_nonvegetated'      : gedi_beam_landcover['modis_nonvegetated'],
             'modis_nonvegetated_sd'   : gedi_beam_landcover['modis_nonvegetated_sd'],
             'modis_treecover'         : gedi_beam_landcover['modis_treecover'],
             'modis_treecover_sd'      : gedi_beam_landcover['modis_treecover_sd'],
             'beam'                    : gedi_beam['beam'],
             'cover'                   : gedi_beam['cover'],
             'master_frac'             : gedi_beam['master_frac'],
             'master_int'              : gedi_beam['master_int'],
             'num_detectedmodes'       : gedi_beam['num_detectedmodes'],
             'omega'                   : gedi_beam['omega'],
             'pai'                     : gedi_beam['pai'],
```

(continues on next page)

(continued from previous page)

```
            'pgap_theta'             : gedi_beam['pgap_theta'],
            'pgap_theta_error'       : gedi_beam['pgap_theta_error'],
            'rg'                     : gedi_beam['rg'],
            'rh100'                  : gedi_beam['rh100'],
            'rhog'                   : gedi_beam['rhog'],
            'rhog_error'             : gedi_beam['rhog_error'],
            'rhov'                   : gedi_beam['rhov'],
            'rhov_error'             : gedi_beam['rhov_error'],
            'rossg'                  : gedi_beam['rossg'],
            'rv'                     : gedi_beam['rv'],
            'sensitivity'            : gedi_beam['sensitivity'],
            'stale_return_flag'      : gedi_beam['stale_return_flag'],
            'surface_flag'           : gedi_beam['surface_flag'],
            'l2a_quality_flag'       : gedi_beam['l2a_quality_flag'],
            'l2b_quality_flag'       : gedi_beam['l2b_quality_flag']})

    gedi_beam_gdf = gpd.GeoDataFrame(gedi_beam_df, crs='EPSG:4326',
                                     geometry=gpd.points_from_xy(gedi_beam_df.
↪longitude_lastbin,
                                                                 gedi_
↪beam_df.latitude_lastbin))
    return gedi_beam_gdf
```

To access the data with this function, we can call the function for each beam id that we have:

[48]:
```
BEAM0000 = get_gedi_df(gedi_h5_file, 'BEAM0000')
BEAM0001 = get_gedi_df(gedi_h5_file, 'BEAM0001')
BEAM0010 = get_gedi_df(gedi_h5_file, 'BEAM0010')
BEAM0011 = get_gedi_df(gedi_h5_file, 'BEAM0011')
BEAM0101 = get_gedi_df(gedi_h5_file, 'BEAM0101')
BEAM0110 = get_gedi_df(gedi_h5_file, 'BEAM0110')
BEAM1000 = get_gedi_df(gedi_h5_file, 'BEAM1000')
BEAM1011 = get_gedi_df(gedi_h5_file, 'BEAM1011')
```

Now we can look at the data in one of the dataframes (beams). We can see that there are 108583 GEDI shots (108583 entries in each column).

[49]:
```
print(BEAM0000.head())
print('number of rows = ', len(BEAM0000))
```

```
   elevation_bin0  elevation_lastbin  height_bin0  height_lastbin  \
0      533.320205         532.870899 -9999.000000    -9999.000000
1      533.610735         533.161429 -9999.000000    -9999.000000
2      835.586133         790.954687    42.284134       -2.373195
3      822.361914         780.875651    38.280968       -3.229356
4      788.459662         760.602475    25.957445       -1.915897

         shot_number  solar_azimuth  solar_elevation  latitude_bin0  \
0  25370000300165164     -34.511044       -10.500773      51.821288
1  25370000300165165     -34.510292       -10.501078      51.821282
2  25370000300165166     -34.509689       -10.501341      51.821258
3  25370000300165167     -34.508926       -10.501648      51.821253
4  25370000300165168     -34.508163       -10.501955      51.821249

   latitude_lastbin  longitude_bin0  ...  rhov  rhov_error  rossg  \
0         51.821288     -127.096372  ...   0.6     -9999.0    0.5
1         51.821282     -127.095550  ...   0.6     -9999.0    0.5
```

(continues on next page)

(continued from previous page)

```
2         51.821260     -127.094876  ...   0.6     -9999.0    0.5
3         51.821255     -127.094048  ...   0.6     -9999.0    0.5
4         51.821250     -127.093210  ...   0.6     -9999.0    0.5

            rv  sensitivity  stale_return_flag  surface_flag  \
0 -9999.000000   -14.502214                  1             1
1 -9999.000000    -8.689309                  1             1
2  3615.732178     0.965246                  0             1
3  4290.229980     0.969620                  0             1
4  2557.771240     0.943713                  0             1

   l2a_quality_flag  l2b_quality_flag                     geometry
0                 0                 0  POINT (-127.09637 51.82129)
1                 0                 0  POINT (-127.09555 51.82128)
2                 1                 1  POINT (-127.09485 51.82126)
3                 1                 1  POINT (-127.09403 51.82126)
4                 1                 1  POINT (-127.09320 51.82125)

[5 rows x 41 columns]
number of rows =  108583
```

2.3.8 Create the Subset and Display the Data

Before displaying this data we can do a spatial subset to remove the data outside of our AOI by doing a spatial subset. We use the Geopandas clip function to clip out the GEDI data based on the extent of our AOI. This reduces the size of the dataframe from 108583 rows to 508.

[50]:
```
BEAM0000_sub = gpd.clip(BEAM0000, AOI)
len(BEAM0000_sub)
```

[50]:
```
508
```

We can display this subset of data over our AOI extent.

[51]:
```
m = folium.Map(location=[38.5, -78], zoom_start=9, tiles='CartoDB positron')

geo_j = folium.GeoJson(data=AOI, style_function=lambda x: {'fillColor': 'orange'})
geo_j.add_to(m)

geo_g = folium.GeoJson(data=BEAM0000_sub,marker = folium.CircleMarker(radius = 1, #␣
↪Radius in metres
                                        weight = 0, #outline weight
                                        fill_color = '#FF0000',
                                        fill_opacity = 1))
geo_g.add_to(m)
m
```

[51]:
```
<folium.folium.Map at 0x7fa48b58f890>
```

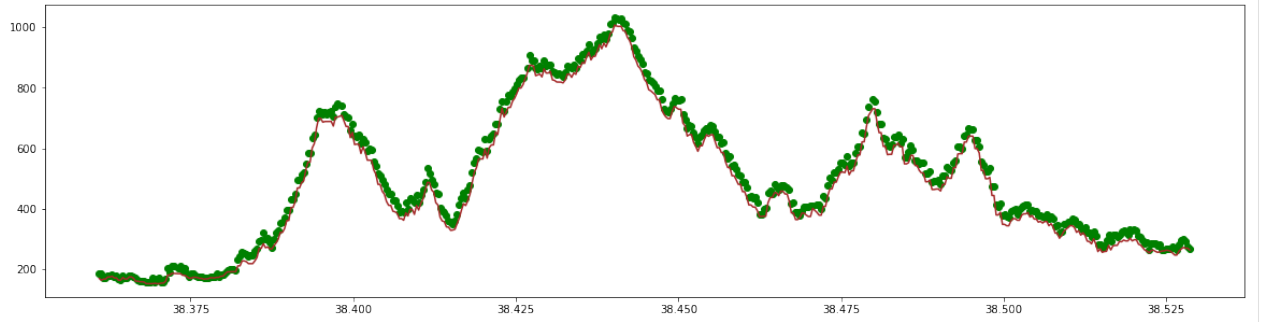
Now we have this subset of data, we can look at some of the GEDI metrics over our area of interest. The 'rh100' metric should give us the top of the canopy heights. What does this look like over Shenandoah National Park? We will look at the ground height and the 'rh100' above ground. The DEM metric is in meters (m) and the 'rh100' metric is in centimeters (cm) above the ground height. Therefore we must add the 'rh100' value to the ground height to get the total height of the tree for display purposes. The 'rh100' metric is also converted to meters (m) to normalize the units

[52]:
```
latitude = BEAM0000_sub['latitude_lastbin']
rh100 = BEAM0000_sub['rh100']
ground = BEAM0000_sub['digital_elevation_model']

TreeHeight = ground + rh100/100
```

Finally, we can make a plot of the ground surface and the canopy heights above the ground surface. We can see the forest canopies in green above the topographically complex ground in brown.

[53]:
```
plt.figure(figsize=(20, 5))
plt.scatter(latitude, TreeHeight, c='green')
plt.plot(latitude, ground, c='brown')
```

[53]:
```
[<matplotlib.lines.Line2D at 0x7fa48b7eb7d0>]
```



Now you have this basic structure you can investigate some of the other metrics and GEDI beams to understand more about the data.

2.4 GEDI_L3 Search and Download

Authors: Anish Bhusal (UAH), Sumant Jha (MSFC/USRA), Jamison French (DevSeed), Aimee Barciauskas (DevSeed), Alex Mandel (DevSeed)

Date: February 8, 2023

Description: In this tutorial, we will search for GEDI L3 data in NASA's Earthdata Search, and then download a GeoTIFF file from the ORNL DAAC S3. We will then visualize the GeoTIFF file, which consists of canopy heights.

2.4.1 Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

2.4.2 About the Data

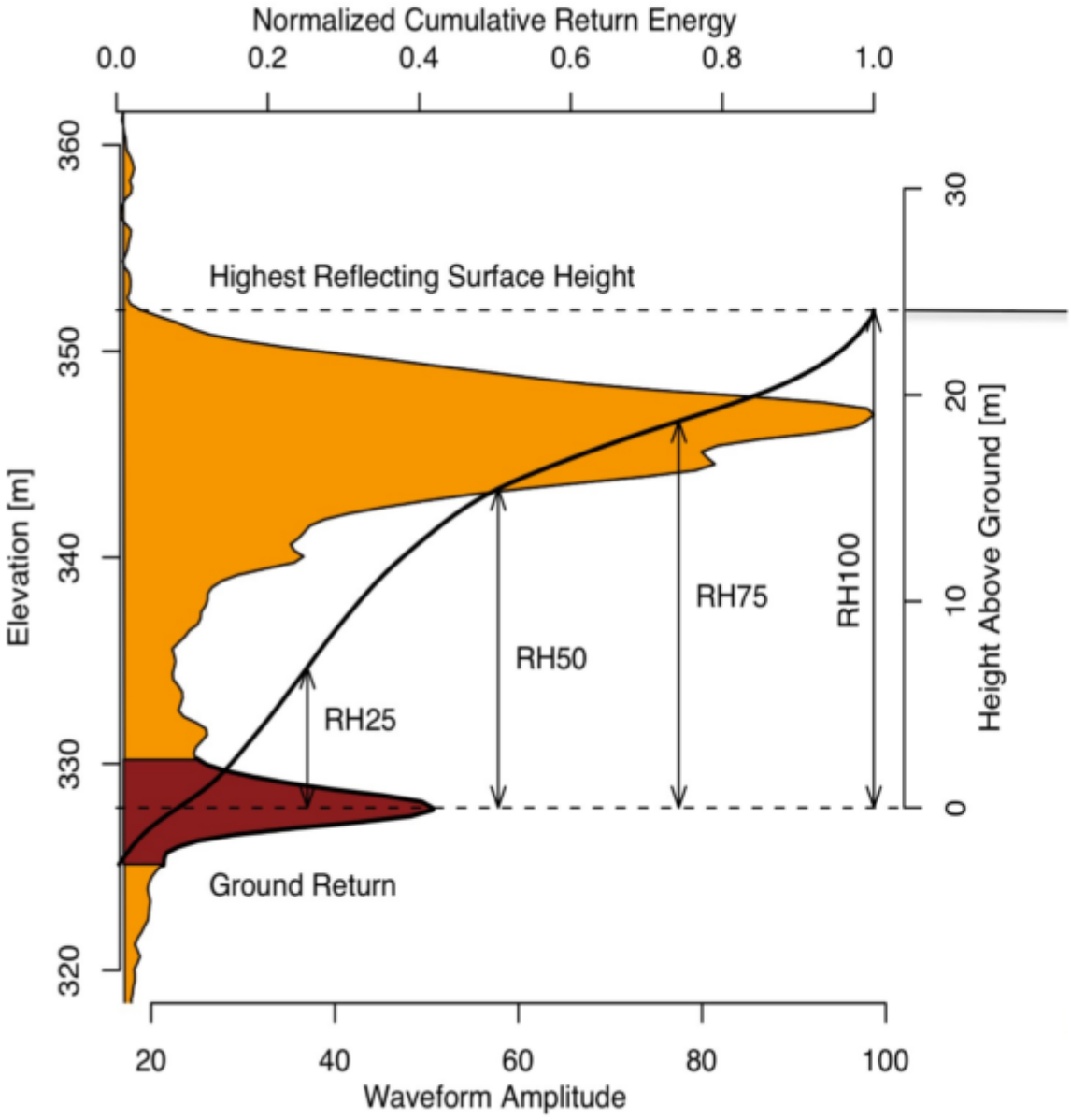
GEDI L3 Gridded Land Surface Metrics, Version 2

This dataset provides Global Ecosystem Dynamics Investigation (GEDI) Level 3 (L3) gridded mean canopy height, standard deviation of canopy height, mean ground elevation, standard deviation of ground elevation, and counts of laser footprints per 1-km x 1-km grid cells globally within -52 and 52 degrees latitude. GEDI is attached to the

International Space Station (ISS) and collects data globally between 51.6° N and 51.6° S latitudes at the highest resolution and densest sampling of any light detection and ranging (lidar) instrument in orbit to date.

GEDI L3 data products are gridded by spatially interpolating Level 2 footprint estimates of canopy cover, canopy height, Leaf Area Index (LAI), vertical foliage profile and their uncertainties. Level 2 data contains terrain elevation, canopy height, RH metrics and Leaf Area Index (LAI). The raw waveforms are collected by GEDI system and processed to provide canopy height and profile metrics. These metrics provide easy to use and interpret information about the vertical distribution of canopy material.

Source: GEDI L3 Gridded Land Surface Metrics, Version 2 Data Set Landing Page



Source: https://gedi.umd.edu/data/products/

Figure: Sample of GEDI lidar waveform (left). The light brown area under the curve represents the return energy from the canopy, while the dark brown area signifies the return from the underlying typography. The black line is cumulative return energey, starting from the bottom of the ground return (normalized to 0) to the top of canopy (normalized to 1).

The diagram on the right shows the distribution of trees that produced the waveform.

2.4.3 Additional Resources

- GEDI L3 Gridded Land Surface Metrics, Version 2 User Guide
- GEDI Overview Page, LPDAAC

2.4.4 Importing and Installing Packages

This tutorial assumes you've all packages installed. If you haven't already, uncomment the following lines to install these packages.

[ ]:
```
# !pip install geopandas
# !pip install contextily
# !pip install backoff
# !pip install folium
# !pip install geojsoncontour
```

[16]:
```
from maap.maap import MAAP
import pandas as pd
import folium
from rasterio.plot import show
import rasterio
import boto3
import os
```

After importing necessary packages, the next step is to initialize MAAP constructor using `api.maap-project.org` as `maap_host` argument.

[17]:
```
maap = MAAP(maap_host="api.maap-project.org")
```

Now, the next step is to seach granules from the CMR. To generate following query, you can use EarthData search feature from MAAP ADE. Refer to this tutorial for more info.

[18]:
```
results=maap.searchGranule(cmr_host='cmr.earthdata.nasa.gov',concept_id="C2153683336-
↪ORNL_CLOUD", limit=1000)
```

The above query gives 1000 results by default. The number of necessary results can be changed using `limit` argument. We can view the `GranuleUR` from results using:

[19]:
```
[result['Granule']['GranuleUR'] for result in results]
```

[19]:
```
['GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2020287_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2021104_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2020287_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2021104_002_02.
↪tif',
```

(continues on next page)

(continued from previous page)

```
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2021216_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2021216_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2022019_002_03.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2022019_002_03.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2022019_002_03.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2022019_002_03.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2022019_002_03.tif']
```

Before downloading a particular tif, let's catch the collection and file name for first item in results:

[20]:
```
granule_ur=results[0]['Granule']['GranuleUR'].split(".")
collection_name=granule_ur[0]
file_name=granule_ur[1]
```

[21]:
```
print(f"collection name: {collection_name} | file_name: {file_name}")
```

```
collection name: GEDI_L3_LandSurface_Metrics_V2 | file_name: GEDI03_elev_lowestmode_
↪stddev_2019108_2020287_002_02
```

2.4.5 Download file from ORNL DAAC S3

To download the file from the source, temporary s3 credentials are required for `maap` package. You can explicitly request `s3_cred_endpoint` for the credentials. The code below wraps that request to get credentials and download the file to your workspace.

[22]:
```
def get_s3_creds(url):
    return maap.aws.earthdata_s3_credentials(url)

def get_s3_client(s3_cred_endpoint):
    creds=get_s3_creds(s3_cred_endpoint)
    boto3_session = boto3.Session(
            aws_access_key_id=creds['accessKeyId'],
            aws_secret_access_key=creds['secretAccessKey'],
            aws_session_token=creds['sessionToken']
    )
    return boto3_session.client("s3")

def download_s3_file(s3, bucket, collection_name, file_name):
    os.makedirs("/projects/gedi_l3", exist_ok=True) # create directories, as necessary
    download_path=f"/projects/gedi_l3/{file_name}.tif"
    s3.download_file(bucket, f"gedi/{collection_name}/data/{file_name}.tif", download_
↪path)
    return download_path
```

[23]:
```
s3_cred_endpoint= 'https://data.ornldaac.earthdata.nasa.gov/s3credentials'
s3=get_s3_client(s3_cred_endpoint)
```

[24]:
```
bucket="ornl-cumulus-prod-protected"
download_path=download_s3_file(s3, bucket, collection_name, file_name)
download_path
```

[24]:
```
'/projects/gedi_l3/GEDI03_elev_lowestmode_stddev_2019108_2020287_002_02.tif'
```

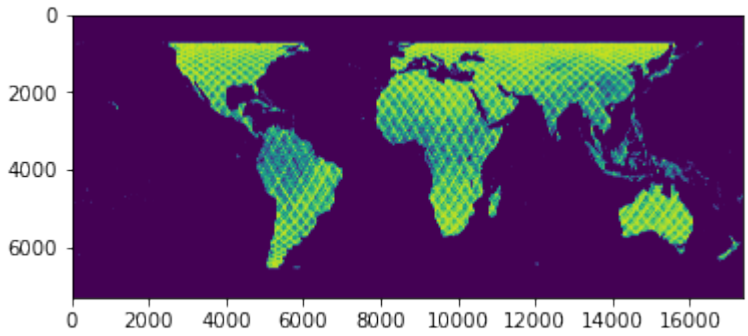
Now, we have the file in our local workspace. It's time to visualize it using `rasterio` package

2.4.6 [Optional] Visualization using Rasterio

The downloaded file is too big to read and visualize directly so we might need to scale it down and view it as a small thumbnail.

[25]:
```
def show_thumbnail(path):
    src=rasterio.open(path)
    oview = src.overviews(1)[0]
    thumbnail = src.read(1, out_shape=(1, int(src.height // oview), int(src.width //␣
↪oview)))
    show(thumbnail)
```

[26]:
```
show_thumbnail(download_path)
```



2.4.7 [Optional] Overlay Raster Layer on top of Folium Map

To properly visualize the canopy heights, we need to display the TIF image on the map. The TIF image file may be too memory and compute-intensive for the kernel causing the process to exit.

[27]:
```
# tif=rasterio.open(download_path)
# arr=tif.read()
# bounds=tif.bounds
```

[28]:
```
#import numpy as np

# x1,y1,x2,y2=bounds
# bbox=[(bounds.bottom, bounds.left), (bounds.top, bounds.right)]
# m=folium.Map(location=[14.59, 120.98], zoom_start=10)
# img = folium.raster_layers.ImageOverlay(image=np.moveaxis(arr, 0, -1), bounds=bbox,␣
↪opacity=0.9, interactive=True, cross_origin=False, zindex=1)
# m
```

2.5 GEDI_L4A Subset and Visualize

2.5.1 [Optional] Install Python Packages

This notebook contains some cells marked as optional, meaning that you can use this notebook without necessarily running such cells.

However, if you do wish to run the optional cells, you must install the following Python packages, which might not already be installed in your environment:

- `geopandas`: for reading your AOI (GeoJson file), as well as for reading the job output (GeoPackage file containing the subset)
- `contextily`: for visually verifying your AOI
- `backoff`: for repeatedly polling the job status (after submission) until the job has been completed (either successfully or not)
- `folium`: for visualizing your data on a Leaflet map
- `geojsoncontour`: for converting your matplotlib contour plots to geojson

[ ]:
```
# Uncomment the following lines to install these packages if you haven't already.
# !pip install geopandas
# !pip install contextily
# !pip install backoff
# !pip install folium
# !pip install geojsoncontour
```

A job can be submitted without these packages, but installing them in order to run the optional cells may make it more convenient for you to visually verify both your AOI and the subset output produced by your job.

2.5.2 Obtain Username

[18]:
```
from maap.maap import MAAP

maap = MAAP(maap_host="api.maap-project.org")
username = maap.profile.account_info()["username"]
username
```

```
WARNING:maap.maap:Unable to load config file from source maap.cfg
WARNING:maap.maap:Unable to load config file from source ./maap.cfg
WARNING:maap.maap:Unable to load config file from source /projects/maap.cfg
```

[18]:
```
'smk0033'
```

2.5.3 Define the Area of Interest

You may use either a publicly available GeoJSON file for your AOI, such as those available at geoBoundaries, or you may create a custom GeoJSON file for your AOI. The following 2 subsections cover both cases.

Using a geoBoundary GeoJSON File

If your AOI is a publicly available geoBoundary, you can obtain the URL for the GeoJSON file using the function below. You simply need to supply an ISO3 value and a level. To find the appropriate ISO3 and level values, see the table on the geoBoundaries site.

[19]:
```
import requests


def get_geo_boundary_url(iso3: str, level: int) -> str:
    response = requests.get(
        f"https://www.geoboundaries.org/api/current/gbOpen/{iso3}/ADM{level}"
    )
    response.raise_for_status()
    return response.json()["gjDownloadURL"]


# If using a geoBoundary, uncomment the following assignment, supply
# appropriate values for `<iso3>` and `<level>`, then run this cell.

# Example (Gabon level 0): get_geo_boundary("GAB", 0)

# aoi = get_geo_boundary_url("<iso3>", <level>)
```

Using a Custom GeoJSON File

Alternatively, you can make your own GeoJSON file for your AOI and place it within your `my-public-bucket` folder within the ADE.

Based upon where you place your GeoJSON file under `my-public-bucket`, you can construct the URL for a job's `aoi` input value.

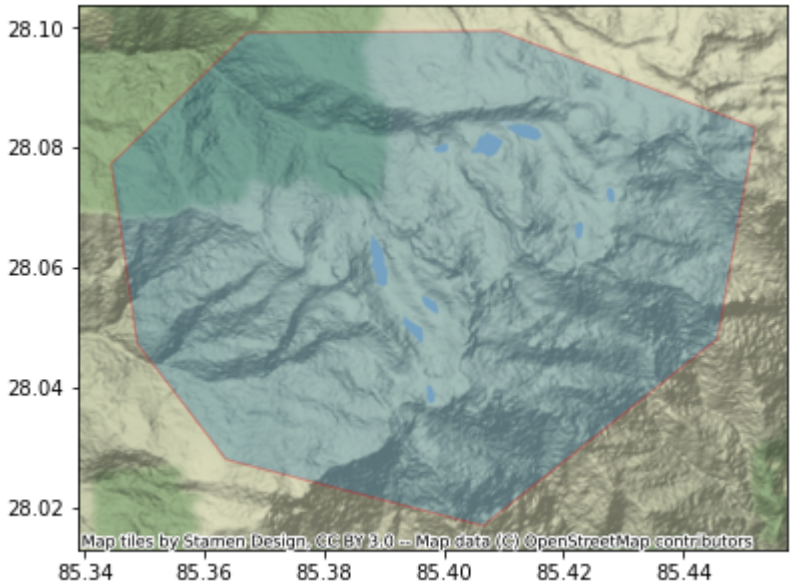
For example, if the relative path of your AOI GeoJSON file under `my-public-bucket` is `path/to/my-aoi.geojson` (avoid using whitespace in the path and filename), the URL you would supply as the value of a job's `aoi` input would be the following (where `{username}` is replaced with your username as output from the previous section):

```
f"https://maap-ops-workspace.s3.amazonaws.com/shared/{username}/path/to/my-aoi.geojson
↪"`
```

If this is the case, use the cell below.

[20]:
```
#aoi = f"https://maap-ops-workspace.s3.amazonaws.com/shared/{username}/langtang_np.
↪geojson"

#for your convenience you can use this geoJSON file but if you have your own geojson,␣
↪use the commented link as example format
aoi = f"https://maap-ops-workspace.s3.amazonaws.com/shared/anisbhsl/langtang_np.
↪geojson"
```

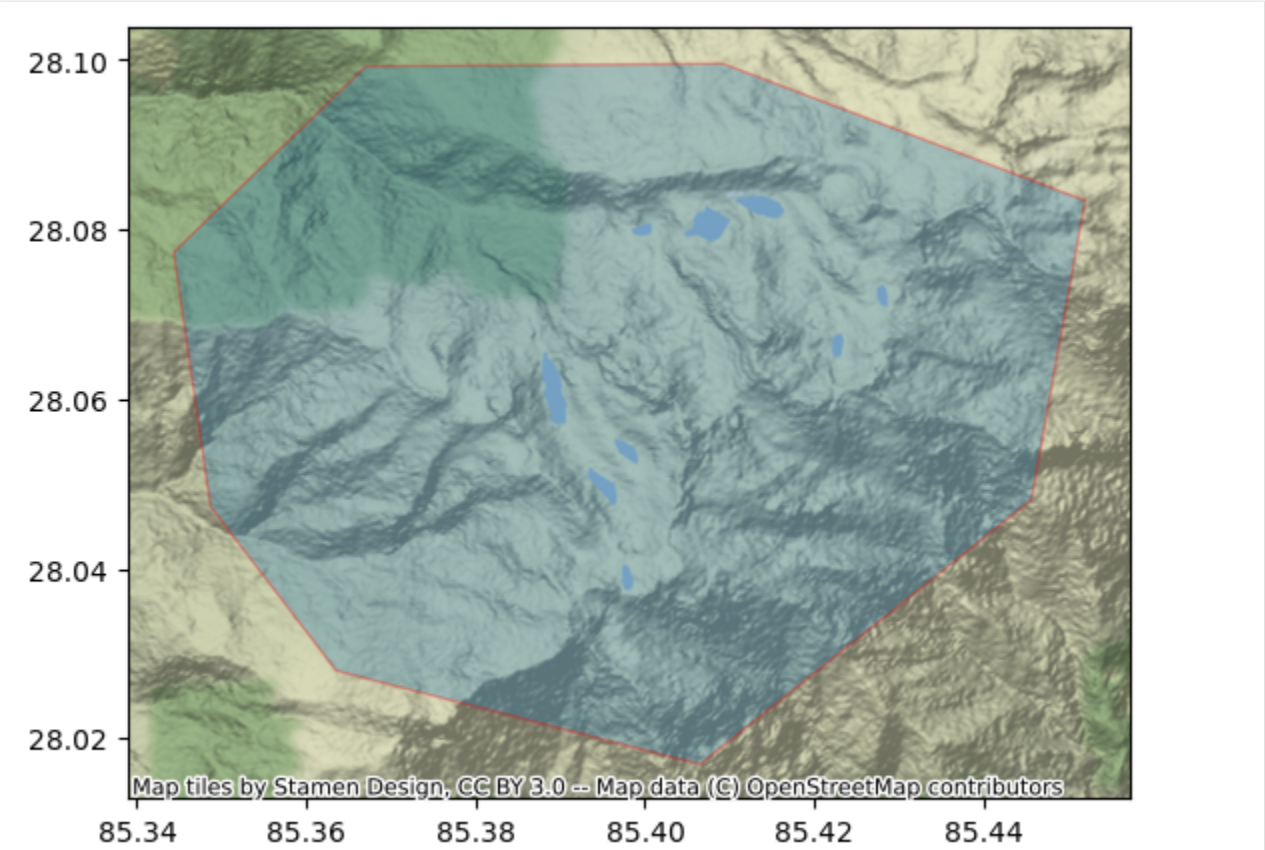
This example uses the AOI of Gosaikunda Lake region inside Langtang National Park. You can also create your own GeoJSON file for your AOI using sites like geojson.io



2.5.4 [Optional] Visually Verify your AOI

If you want to visually verify your AOI before proceeding, you may run the following cell, if you have the `geopandas` and `contextily` Python packages installed.

[21]:
```
try:
    import geopandas as gpd
    import contextily as ctx
except:
    print(
        "If you wish to visually verify your AOI, "
        "you must install the `geopandas` and `contextily` packages."
    )
else:
    aoi_gdf = gpd.read_file(aoi)
    aoi_epsg4326 = aoi_gdf.to_crs(epsg=4326)
    ax = aoi_epsg4326.plot(figsize=(10, 5), alpha=0.3, edgecolor="red")
    ctx.add_basemap(ax, crs=4326)
```



2.5.5 Submit a Job

When supplying input values for a GEDI subsetting job, to use the default value for a field (where indicated), use a dash (`"-"`) as the input value.

- `aoi` (required): URL to a GeoJSON file representing your area of interest, as explained above.
- `doi`: Digital Object Identifier (DOI) of the GEDI collection to subset, or a logical name representing such a DOI. Valid logical names: L1B, L2A, L2B, L4A
- `columns`: Comma-separated list of column names to include in the output file.
- `query`: Query expression for subsetting the rows in the output file.
- `limit`: Maximum number of GEDI granule data files to download (among those that intersect the specified AOI). (Default: 10000)

It is recommended to use `maap-dps-worker-32gb` queues when submitting a job with a large aoi.

[22]:
```
inputs = dict(
    aoi=aoi,
    doi="L4A",
    lat="lat_lowestmode",
    lon="lon_lowestmode",
    beams="coverage",
    columns="agbd, agbd_se, sensitivity, geolocation/sensitivity_a2, elev_lowestmode",
    query="l2_quality_flag == 1 and l4_quality_flag == 1 and sensitivity > 0.95 and␣
↪`geolocation/sensitivity_a2` > 0.95",
```

(continues on next page)

(continued from previous page)

```
    limit=10,
    temporal="-",
    output="gedi_subset.gpkg"
)

result = maap.submitJob(
    identifier="gedi-subset",
    algo_id="gedi-subset",
    version="0.6.0",
    queue="maap-dps-worker-32gb",
    username=username,
    **inputs,
)

job_id = result.id
job_id or result
```

[22]:
```
'72fca5ba-935a-49a2-802f-1dcfd3a5628c'
```

2.5.6 Get the Job's Output File

Now that the job has been submitted, we can use the `job_id` to check the job status until the job has been completed.

[23]:
```
from urllib.parse import urlparse


def job_status_for(job_id: str) -> str:
    return maap.getJobStatus(job_id)


def job_result_for(job_id: str) -> str:
    return maap.getJobResult(job_id)[0]


def to_job_output_dir(job_result_url: str) -> str:
    return f"/projects/my-private-bucket/{job_result_url.split(f'/{username}/')[1]}"
```

If you have installed the `backoff` Python package, running the following cell will automatically repeatedly check your job's status until the job has been completed. Otherwise, you will have to manually repeatedly rerun the following cell until the output is either `'Succeeded'` or `'Failed'`.

[24]:
```
try:
    import backoff
except:
    job_status = job_status_for(job_id)
else:
    # Check job status every 2 minutes
    @backoff.on_predicate(
        backoff.constant,
        lambda status: status not in ["Deleted", "Succeeded", "Failed"],
        interval=120,
    )
    def wait_for_job(job_id: str) -> str:
        return job_status_for(job_id)
```

(continues on next page)

(continued from previous page)

```
    job_status = wait_for_job(job_id)

job_status
```

```
INFO:backoff:Backing off wait_for_job(...) for 0.9s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 18.1s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 49.5s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 6.8s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 42.4s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 26.7s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 86.6s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 117.0s (Accepted)
INFO:backoff:Backing off wait_for_job(...) for 17.9s (Running)
INFO:backoff:Backing off wait_for_job(...) for 95.7s (Running)
```

[24]:
```
'Succeeded'
```

[25]:
```
assert job_status == "Succeeded", (
    job_result_for(job_id)
    if job_status == "Failed"
    else f"Job {job_id} has not yet completed ({job_status}). Rerun the prior cell."
)

output_url = job_result_for(job_id)
output_dir = to_job_output_dir(output_url)
output_file = f"{output_dir}/gedi_subset.gpkg"
print(f"Your subset results are in the file {output_file}")
```

```
Your subset results are in the file /projects/my-private-bucket/dps_output/gedi-
↪subset/0.6.0/2023/06/27/20/05/21/764642/gedi_subset.gpkg
```

2.5.7 [Optional] Visually Verify the Results

If you installed the `geopandas` Python package, you can visually verify the output file by running the following cell.

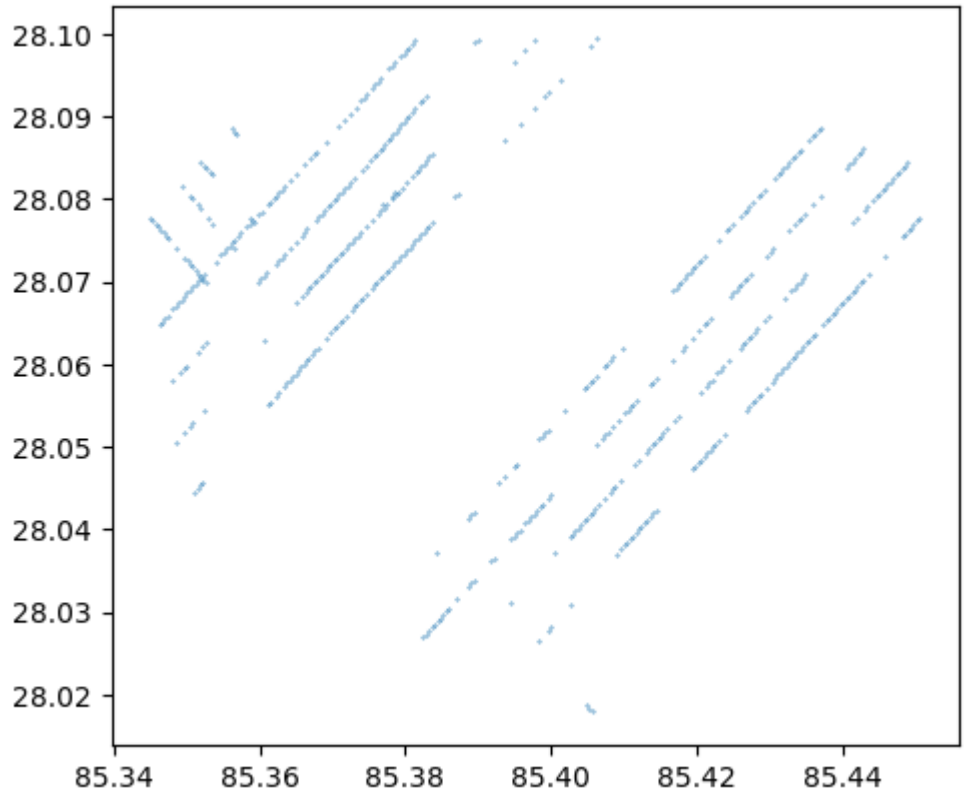
[26]:
```
try:
    import geopandas as gpd
    import matplotlib.pyplot as plt
except:
    print(
        "If you wish to visually verify your output file, "
        "you must install the `geopandas` package."
    )
else:
    gedi_gdf = gpd.read_file(output_file)
    print(gedi_gdf.head())
    sensitivity_colors = plt.cm.get_cmap("viridis_r")
    gedi_gdf.plot(markersize = 0.1)
```

```
                                            filename  \
0  GEDI04_A_2020064181434_O06951_02_T04323_02_002...
1  GEDI04_A_2020064181434_O06951_02_T04323_02_002...
2  GEDI04_A_2020064181434_O06951_02_T04323_02_002...
3  GEDI04_A_2020064181434_O06951_02_T04323_02_002...
4  GEDI04_A_2020064181434_O06951_02_T04323_02_002...
```

(continues on next page)

(continued from previous page)

```
   geolocation/sensitivity_a2  sensitivity  elev_lowestmode    agbd_se  \
0                    0.959091     0.959091      3256.992432  11.047880
1                    0.968629     0.968629      3282.666748  11.057747
2                    0.962079     0.962079      3314.994141  11.052886
3                    0.962610     0.962610      3351.686035  11.045983
4                    0.968436     0.968436      3436.938721  11.047858

         agbd                    geometry
0  169.741974  POINT (85.34628 28.06473)
1  235.977890  POINT (85.34667 28.06512)
2  189.141037  POINT (85.34705 28.06550)
3  180.132187  POINT (85.34742 28.06589)
4  191.731232  POINT (85.34817 28.06666)
```



2.5.8 Generate contour lines

Create a `lat`, `lon` mesh grid with elevation as a depth parameter. As shown in the plot above, the lines don't seem smooth. So we can apply `linear` or 'cubic' interpolation to smoothen those missing points.

[27]:
```
geometry = gedi_gdf["geometry"]
elevation=gedi_gdf["elev_lowestmode"]
```

[28]:
```
lon = geometry.x
lat = geometry.y
```

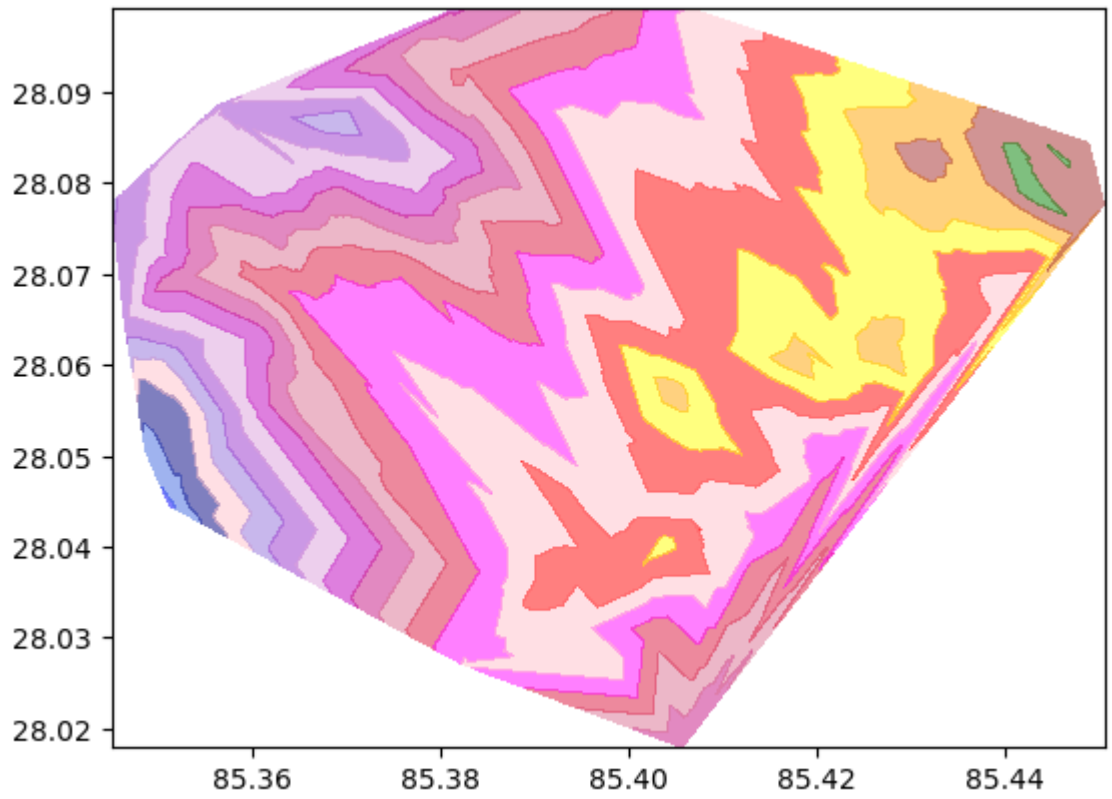
[29]:
```
import numpy as np

x=np.linspace(min(lon), max(lon), 1000)
y=np.linspace(min(lat), max(lat), 1000)
```

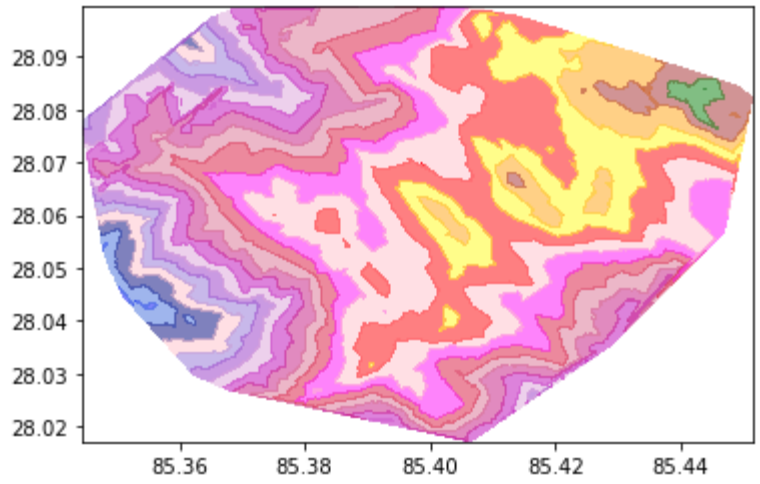
[30]:
```
from scipy.interpolate import griddata

x_mesh, y_mesh = np.meshgrid(x,y)
```

You may experiment with `nearest`, `linear`, and `cubic` interpolation methods to see which gives more smooth results.

[31]:
```
#grid the elevation
z_mesh = griddata((lon, lat), elevation, (x_mesh, y_mesh), method='linear')
```

[32]:
```
colors=['blue','royalblue', 'navy','pink',  'mediumpurple',  'darkorchid',  'plum',
↪'m', 'mediumvioletred', 'palevioletred', 'crimson',
        'magenta','pink','red','yellow','orange', 'brown','green', 'darkgreen']
levels=len(colors)
contourf = plt.contourf(x_mesh, y_mesh, z_mesh, levels, alpha=0.5,  colors=colors,␣
↪linestyles='None', vmin=elevation.min(), vmax=elevation.max())
```





Now we need to plot this contour into an interactive map for better visualization.

2.5.9 Plot the contour lines in folium

You may need to install `geojsoncontour`, `mapclassify`, and `folium`, if you don't already have them installed. We need to convert this `contourf` into geoJSON format.

[33]:
```
import folium
from folium import plugins
import branca
import geojsoncontour
```

[34]:
```
geojson = geojsoncontour.contourf_to_geojson(
    contourf=contourf,
    min_angle_deg=3.0,
    ndigits=5,
    stroke_width=1,
    unit='ft',
    fill_opacity=0.1,
)
```

[35]:
```
#create map view
m = folium.Map([lat.mean(), lon.mean()], zoom_start=12, tiles="OpenStreetMap")

folium.GeoJson(
    geojson,
    style_function=lambda x:{
        'color': x['properties']['stroke'],
        'weight':    x['properties']['stroke-width'],
        'fillColor': x['properties']['fill'],
        'opacity':   0.5,
    }
).add_to(m)

cm = branca.colormap.LinearColormap(colors, vmin=elevation.min(), vmax=elevation.
↪max()).to_step(levels)
cm.caption='Elevation (in m)'
m.add_child(cm)
```

(continues on next page)

(continued from previous page)

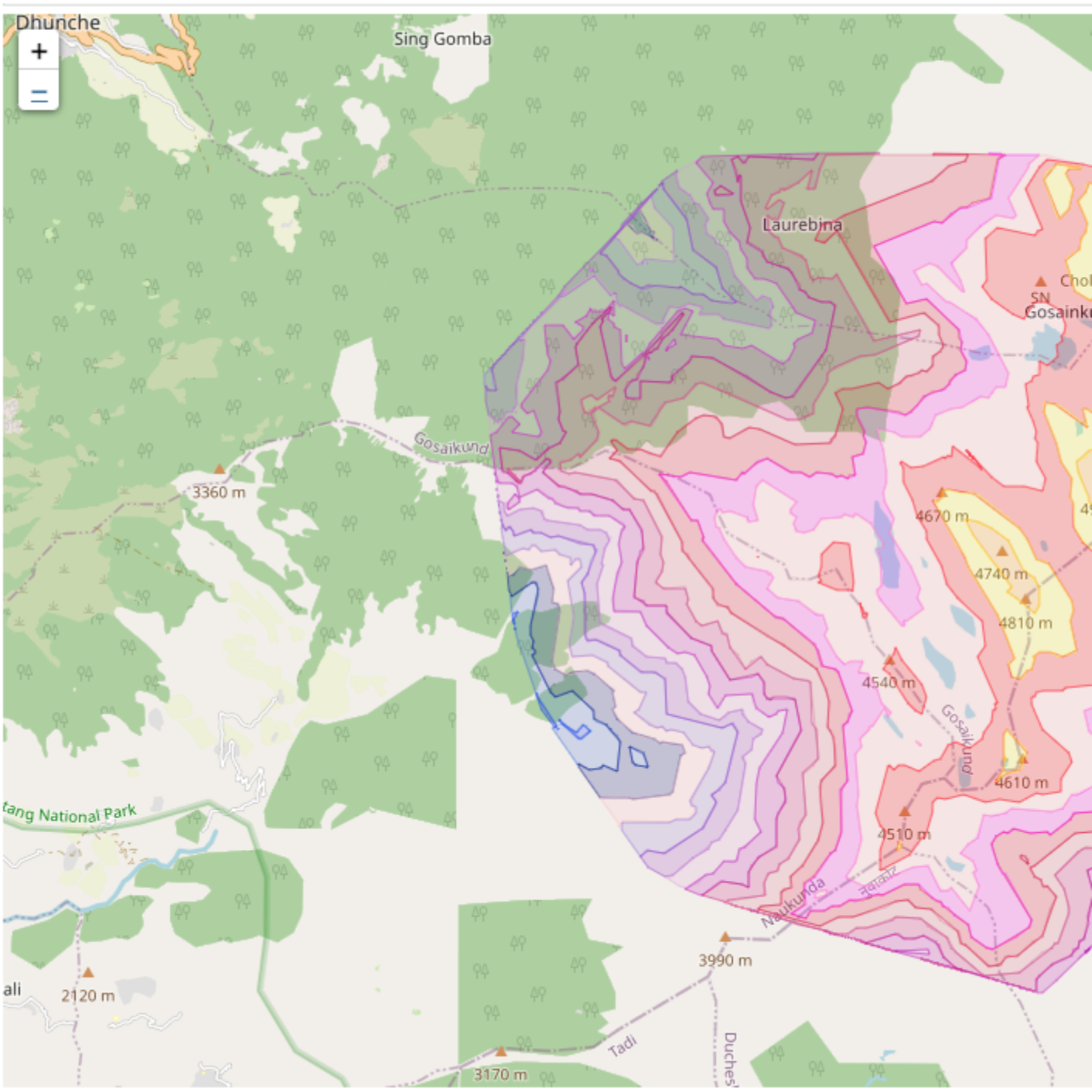
```
#legend
plugins.Fullscreen(position='topright', force_separate_button=True).add_to(m)
```

[35]:
```
<folium.plugins.fullscreen.Fullscreen at 0x7f356182c4c0>
```

[36]:
```
m
```

[36]:
```
<folium.folium.Map at 0x7f356182c670>
```

Now you have an interactive visualization of a contour plot.



2.6 GEDI_L4B Search and Visualize

Authors: Nikita Susan (UAH), Aimee Barciauskas (DevSeed), Sumant Jha (MSFC/USRA), Alex Mandel (DevSeed)

Date: April 7, 2023

Description: In this example, we demonstrate how to access the GEDI L4B collection and granule data on the MAAP ADE, and then visualize the data using matplotlib.

2.6.1 Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

2.6.2 About the Data

GEDI L4B Gridded Aboveground Biomass Density, Version 2

This dataset provides Global Ecosystem Dynamics Investigation (GEDI) Level 4 (L4) data, which has the purpose of providing mean aboveground biomass density (AGBD) and consists of the GEDI_L4A and GEDI_L4B collections. GEDI_L4B uses a sample present within each 1km cell to statistically infer mean AGBD. GEDI is attached to the International Space Station (ISS) and collects data globally between 51.6° N and 51.6° S latitudes at the highest resolution and densest sampling of any light detection and ranging (lidar) instrument in orbit to date; specifically, GEDI L4B data has a spatial resolution of 1km. (Source: GEDI_L4B Version 2 User Guide)

2.6.3 Additional Resources

- GEDI_L4B Version 2 Dataset Landing Page
- The GEDI Website
- Earthdata Search

2.6.4 Importing Packages

Within your Jupyter Notebook, start by importing the **maap package**. Then invoke the **MAAP** constructor, setting the maap_host argument to **'api.maap-project.org'**.

[22]:
```
from maap.maap import MAAP
from matplotlib import pyplot
import os
import pprint
import rasterio
import boto3

maap = MAAP(maap_host="api.maap-project.org")
```

2.6.5 Search for the Collection and Associated Granules

Now, we will search for the collection using the collection short name:

[23]:
```
collection = maap.searchCollection(cmr_host='cmr.earthdata.nasa.gov', short_name=
↪"GEDI_L4B_Gridded_Biomass_2017", limit=100)
print(collection)
```

```
[{'concept-id': 'C2244602422-ORNL_CLOUD', 'revision-id': '7', 'format': 'application/
↪echo10+xml', 'Collection': {'ShortName': 'GEDI_L4B_Gridded_Biomass_2017', 'VersionId
↪': '2', 'InsertTime': '2022-03-29T00:00:00Z', 'LastUpdate': '2023-06-12T20:25:17Z',
↪'LongName': 'GEDI L4B Gridded Aboveground Biomass Density, Version 2', 'DataSetId':
↪'GEDI L4B Gridded Aboveground Biomass Density, Version 2', 'Description': "This␣
↪Global Ecosystem Dynamics Investigation (GEDI) L4B product provides 1 km␣ (continues on next page)
↪km,  hereafter) estimates of mean aboveground biomass density (AGBD) based on␣
↪observations from mission week 19 starting on 2019-04-18 to mission week 138 ending␣
↪on 2021-08-04. The GEDI L4A Footprint Biomass product converts each high quality␣
↪waveform to an AGBD prediction, and the L4B product uses the sample present within␣
↪the borders of each 1 km cell to statistically infer mean AGBD. The gridding␣
↪procedure is described in the GEDI L4B Algorithm Theoretical Basis Document (ATBD).␣
```

(continued from previous page)

Next, we can search for granules using the searchGranule function and the concept ID from our collection search above:

[24]:
```
COLLECTIONID = collection[0]['concept-id']
results = maap.searchGranule(cmr_host='cmr.earthdata.nasa.gov',concept_
↪id=COLLECTIONID)  # COLLECTIONID 'C2244602422-ORNL_CLOUD'
print(f'Got {len(results)} results')
results[0]['Granule']
```

```
Got 10 results
```

[24]:
```
{'GranuleUR': 'GEDI_L4B_Gridded_Biomass.GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif',
 'InsertTime': '2022-03-29T00:00:00Z',
 'LastUpdate': '2023-04-10T21:58:24Z',
 'Collection': {'ShortName': 'GEDI_L4B_Gridded_Biomass_2017',
  'VersionId': '2'},
 'DataGranule': {'DataGranuleSizeInBytes': '20103343',
  'SizeMBDataGranule': '20.103343',
  'Checksum': {'Value':
↪'025a141348906d5e612262218c496a2d468446ca30875439be6651d851bfbe23',
   'Algorithm': 'SHA-256'},
  'DayNightFlag': 'BOTH',
  'ProductionDateTime': '2022-03-29T00:00:00Z'},
 'Temporal': {'RangeDateTime': {'BeginningDateTime': '2019-04-18T00:00:00Z',
   'EndingDateTime': '2021-08-04T23:59:59Z'}},
 'Spatial': {'HorizontalSpatialDomain': {'Geometry': {'BoundingRectangle': {
↪'WestBoundingCoordinate': '-180',
     'NorthBoundingCoordinate': '52',
     'EastBoundingCoordinate': '180',
     'SouthBoundingCoordinate': '-52'}}}},
 'MeasuredParameters': {'MeasuredParameter': [{'ParameterName': 'LIDAR WAVEFORM'},
   {'ParameterName': 'BIOMASS'},
   {'ParameterName': 'TERRESTRIAL ECOSYSTEMS'}]},
 'Platforms': {'Platform': {'ShortName': 'ISS',
   'Instruments': {'Instrument': {'ShortName': 'GEDI'}}}},
 'Campaigns': {'Campaign': {'ShortName': 'GEDI'}},
 'Price': '0',
 'OnlineAccessURLs': {'OnlineAccessURL': [{'URL': 'https://data.ornldaac.earthdata.
↪nasa.gov/protected/gedi/GEDI_L4B_Gridded_Biomass/data/GEDI04_B_MW019MW138_02_002_05_
↪R01000M_PS.tif',
    'URLDescription': 'Download GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif',
    'MimeType': 'image/tiff'},
   {'URL': 'https://data.ornldaac.earthdata.nasa.gov/public/gedi/GEDI_L4B_Gridded_
↪Biomass/data/GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif.sha256',
    'URLDescription': 'Download GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif.sha256'},
   {'URL': 's3://ornl-cumulus-prod-protected/gedi/GEDI_L4B_Gridded_Biomass/data/
↪GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif',
    'URLDescription': 'This link provides direct download access via S3 to the granule
↪',
    'MimeType': 'image/tiff'}]},
 'OnlineResources': {'OnlineResource': [{'URL': 'https://daac.ornl.gov/GEDI/guides/
↪GEDI_L4B_Gridded_Biomass.html',
    'Description': 'ORNL DAAC Data Set Documentation',
    'Type': "USER'S GUIDE"},
   {'URL': 'https://doi.org/10.3334/ORNLDAAC/2017',
    'Description': 'Data set Landing Page DOI URL',
```

(continues on next page)

(continued from previous page)

```
    'Type': 'DATA SET LANDING PAGE'},
   {'URL': 'https://daac.ornl.gov/daacdata/gedi/GEDI_L4B_Gridded_Biomass/comp/GEDI_
↪L4B_ATBD_v1.0.pdf',
    'Description': 'Data Set Documentation',
    'Type': 'GENERAL DOCUMENTATION'},
   {'URL': 'https://daac.ornl.gov/daacdata/gedi/GEDI_L4B_Gridded_Biomass/comp/GEDI_
↪L4B_Gridded_Biomass.pdf',
    'Description': 'Data Set Documentation',
    'Type': 'GENERAL DOCUMENTATION'},
   {'URL': 'https://webmap.ornl.gov/sdat/pimg/2017_9.png',
    'Description': 'GEDI L4B Gridded Prediction Stratum, Version 2, Mission Weeks 19-
↪138',
    'Type': 'BROWSE',
    'MimeType': 'image/png'},
   {'URL': 'https://data.ornldaac.earthdata.nasa.gov/s3credentials',
    'Description': 'api endpoint to retrieve temporary credentials valid for same-
↪region direct s3 access',
    'Type': 'VIEW RELATED INFORMATION'}]},
 'Orderable': 'false',
 'DataFormat': 'COG'}
```

2.6.6 Accessing and Downloading the Granule from ORNL DAAC S3

Before downloading, we'll get the collection and file name:

[25]:
```
granule_ur=results[0]['Granule']['GranuleUR'].split(".")
collection_name=granule_ur[0]
file_name=granule_ur[1]

print(collection_name)
print(file_name)
```

```
GEDI_L4B_Gridded_Biomass
GEDI04_B_MW019MW138_02_002_05_R01000M_PS
```

Now we'll proceed to get tempory s3 credentials, and then download the tif file to our workspace:

[26]:
```
def get_s3_creds(url):
    return maap.aws.earthdata_s3_credentials(url)

def get_s3_client(s3_cred_endpoint):
    creds=get_s3_creds(s3_cred_endpoint)
    boto3_session = boto3.Session(
            aws_access_key_id=creds['accessKeyId'],
            aws_secret_access_key=creds['secretAccessKey'],
            aws_session_token=creds['sessionToken']
    )
    return boto3_session.client("s3")

def download_s3_file(s3, bucket, collection_name, file_name):
    os.makedirs("/projects/gedi_l4b", exist_ok=True) # create directories, as␣
↪necessary
    download_path=f"/projects/gedi_l4b/{file_name}.tif"
    s3.download_file(bucket, f"gedi/{collection_name}/data/{file_name}.tif", download_
↪path)
    return download_path
```

[27]:
```
s3_cred_endpoint= 'https://data.ornldaac.earthdata.nasa.gov/s3credentials'
s3=get_s3_client(s3_cred_endpoint)
```

[28]:
```
bucket="ornl-cumulus-prod-protected"
download_path=download_s3_file(s3, bucket, collection_name, file_name)
download_path
```

[28]:
```
'/projects/gedi_l4b/GEDI04_B_MW019MW138_02_002_05_R01000M_PS.tif'
```

Open the local file using rasterio, and print the shape of the data to verify if the file was read properly:

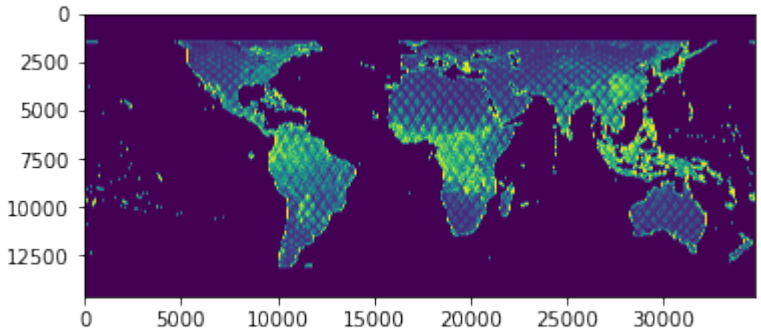
[29]:
```
src = rasterio.open(download_path)
data = src.read(1)

print(data.shape)
```

```
(14616, 34704)
```

2.6.7 Plot the Data

Finally, we'll use matplotlib to visualize our .tif file:

[30]:
```
pyplot.imshow(data)
pyplot.show()
```



2.7 ICESat-02 ATL03 Subset and Visualize

About ATL03 Data: (ATL03) contains height above the WGS 84 ellipsoid (ITRF2014 reference frame), latitude, longitude, and time for all photons downlinked by the Advanced Topographic Laser Altimeter System (ATLAS) instrument on board the Ice, Cloud and land Elevation Satellite-2 (ICESat-2) observatory. The ATL03 product was designed to be a single source for all photon data and ancillary information needed by higher-level ATLAS/ICESat-2 products. As such, it also includes spacecraft and instrument parameters and ancillary data not explicitly required for ATL03. (source: https://nsidc.org/data/atl03/versions/5)

Required packages:

You will need to install the following required packages if not already present in your working environment: maap-py, pandas, geopandas, folium, shapely, h5glance, h5py

2.7.1 Import relevant python modules

[1]:
```
# Import the MAAP package
from maap.maap import MAAP

# Invoke the MAAP constructor using the maap_host argument
maap = MAAP(maap_host='api.ops.maap-project.org')

# Import pandas dataframe
import pandas as pd

# Import libraries needed for visualizing data spatial extent
import geopandas as gpd
import folium
from shapely.geometry import Polygon,Point

# Import H5glance to interactively explore H5 file in notebook
from h5glance import H5Glance

# Import H5py to read h5 file
import h5py
```

```
/opt/conda/lib/python3.7/site-packages/geopandas/_compat.py:115: UserWarning: The␣
↪Shapely GEOS version (3.11.1-CAPI-1.17.1) is incompatible with the GEOS version␣
↪PyGEOS was compiled with (3.8.1-CAPI-1.13.3). Conversions between both will be slow.
  shapely_geos_version, geos_capi_version_string
```

2.7.2 Decide on a subset of ATL03 data to load using spatial extent and date range and visualize extent using folium

[2]:
```
# Create a variable for short name of ATL03 data
short_name = 'ATL03'

# Create spatial extent for which we will search
# ATL03 granules on NASA CMR. We have chosen a very narrow
# area over Yosemite national park. This is okay,
# as we need just a few granules for demo purposes.

# Create Latitude, Longitude list.
lat_coords = [37.700057,37.700057,37.758166,37.758166,37.700057]
lon_coords = [-119.680359,-119.680359,-119.538910,-119.538910,-119.680359]

# Create bounding box
spatial_extent = [lon_coords[0],lat_coords[0],lon_coords[2],lat_coords[2]]

# Reformat bounding box to work with NASA CMR API
spatial_extent = ','.join(str(coords) for coords in spatial_extent)

#Provide date range. It is just 1 day.
date_range = ['2021-02-02','2022-02-03']

# For folium purpose, provide the map center
map_center = [37.729139,-119.609670]

# Convert to AOI for visualizaton with folium
polygon_geom = Polygon(zip(lon_coords, lat_coords))
```

(continues on next page)

(continued from previous page)

```
# Provide relevant Coordinate Reference System
crs = 'epsg:4326'

# Convert to Geodataframe and back to list in that specific reference system
AOI = gpd.GeoDataFrame(index=[0], crs=crs, geometry=[polygon_geom])
AOI_bbox = AOI.bounds.iloc[0].to_list()
```

[3]:
```
# Visualize the spatial extent using folium.
m = folium.Map(map_center, zoom_start=12, tiles='OpenStreetMap')
folium.GeoJson(AOI).add_to(m)
folium.LatLngPopup().add_to(m)
m
```

[3]:
```
<folium.folium.Map at 0x7ff1580c11d0>
```

2.7.3 Search for available granules in the NASA CMR for the given spatial extent and date range and print the number of granules available.

[4]:
```
# Provide the name of cmr host to use with maap py
nasa_cmr_host = "cmr.earthdata.nasa.gov"
# Search granule using maap-py using subset criteria identified earlier.
data = maap.searchGranule(cmr_host=nasa_cmr_host,short_name = short_name, bounding_
↪box = spatial_extent,temporal= date_range,limit=1000)
# Check to see if the granule search was successfull in finding data within spatial␣
↪and temporal extent
print(len(data))
```

```
68
```

[5]:
```
#Use maap-py's getDatafunction to extract one of the HDF files from NSIDC servers.
# Print the name of file extracted. The file will be downloaded and stored in your␣
↪current directory

ice_data = data[0].getData()
print(ice_data)
```

```
./ATL03_20210202191800_06231006_005_01.h5
```

2.7.4 Read the H5 file to understand the data structure.

There are two ways to do this. We can just list keys and then go forward exploring each key one by one.

[6]:
```
# Open the H5 file and list the keys
ice_file = h5py.File(ice_data,'r')
list(ice_file.keys())
```

[6]:
```
['METADATA',
 'ancillary_data',
 'atlas_impulse_response',
 'ds_surf_type',
 'ds_xyz',
```

(continues on next page)

(continued from previous page)

```
 'gt1l',
 'gt1r',
 'gt2l',
 'gt2r',
 'gt3l',
 'gt3r',
 'orbit_info',
 'quality_assessment']
```

Or use the h5glance package to interactively list various keys, sub-keys and variables.

H5Glance lists all the keys and sub-keys and allows for the copying of path variables on the fly, all from within the Jupyter Notebook. Note: In the web version of this notebook which is shown here, not all sub-fields will be listed. But they can be accessed when using a Jupyter Notebook running on ADE or local machines.

[7]:
```
H5Glance(ice_file)
```

[7]:
```
./ATL03_20210202191800_06231006_005_01.h5/ (47 attributes)
├METADATA    (9 children) (3 attributes)
├ancillary_data     (34 children) (2 attributes)
├atlas_impulse_response    (2 children) (1 attributes)
├ds_surf_type      [int32: 5] (12 attributes)
├ds_xyz     [int32: 3] (12 attributes)
├gt1l        (5 children) (7 attributes)
├gt1r        (5 children) (7 attributes)
├gt2l        (5 children) (7 attributes)
├gt2r        (5 children) (7 attributes)
├gt3l        (5 children) (7 attributes)
├gt3r        (5 children) (7 attributes)
├orbit_info         (7 children) (2 attributes)
└quality_assessment         (9 children) (1 attributes)
```

2.7.5 Subset the data by required columns. In this case we need Latitude, Longitude, Photon Height and Along Track Distance.

We are using the copied path from the data tree generated by H5Glance above. Use h5py to read file and variables from the copied path.

[8]:
```
with h5py.File(ice_data,'r') as f:
    gt1l_lat = f['/gt1l/heights/lat_ph'][:]
    gt1l_lon = f['/gt1l/heights/lon_ph'][:]
    gt1l_height = f['/gt1l/heights/h_ph'][:]
    gt1l_dist_ph = f['/gt1l/heights/dist_ph_along'][:]
```

2.7.6 Show the subset data in a dataframe.

1. By using Pandas module.

[9]:
```
# Write latitude, longitude, photon height and along track distance for gt1l to a␣
↪dataframe
gt1l_df = pd.DataFrame({'Latitude': gt1l_lat, 'Longitude': gt1l_lon, 'Photon_Height':␣
↪gt1l_height,'Along_track_distance':gt1l_dist_ph})
```

(continues on next page)

(continued from previous page)

```
gt1l_df
```

[9]:

```
           Latitude   Longitude  Photon_Height  Along_track_distance
0         59.482065 -115.906874     611.887878             16.894447
1         59.482065 -115.906877     580.084106             16.980614
2         59.482064 -115.906882     527.962097             17.122099
3         59.482063 -115.906885     491.036133             17.222527
4         59.482059 -115.906875     616.919922             17.593958
...             ...         ...            ...                   ...
36561868  33.011390 -119.752303      69.065865              4.154013
36561869  33.011382 -119.752316     -16.501354              5.156162
36561870  33.011380 -119.752329    -108.562981              5.470425
36561871  33.011376 -119.752317     -17.947205              5.871534
36561872  33.011375 -119.752319     -37.489002              5.938596

[36561873 rows x 4 columns]
```

2. By using Geopandas module.

Create geopandas dataframe with a column for point locations in the geometry column, and other relevant variables.

[10]:

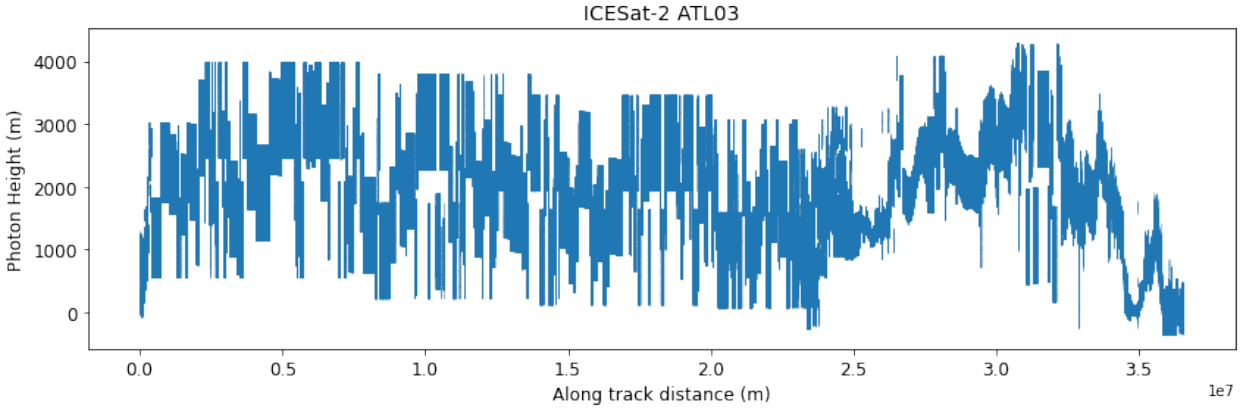
```
geometry = gpd.points_from_xy(gt1l_lon, gt1l_lat)
data = {'Latitude': gt1l_lat, 'Longitude': gt1l_lon, 'Photon_Height': gt1l_height,
↪'Along_track_distance':gt1l_dist_ph}
gdf = gpd.GeoDataFrame(data,geometry=geometry, crs='EPSG:4326')

# View the resulting geopandas dataframe
print(gdf.head())
```

```
    Latitude   Longitude  Photon_Height  Along_track_distance  \
0  59.482065 -115.906874     611.887878             16.894447
1  59.482065 -115.906877     580.084106             16.980614
2  59.482064 -115.906882     527.962097             17.122099
3  59.482063 -115.906885     491.036133             17.222527
4  59.482059 -115.906875     616.919922             17.593958

                     geometry
0  POINT (-115.90687 59.48207)
1  POINT (-115.90688 59.48206)
2  POINT (-115.90688 59.48206)
3  POINT (-115.90689 59.48206)
4  POINT (-115.90687 59.48206)
```

2.7.7 Visualize photon heights with respect to along track distance for this H5 file (using inputs from geodataframe)

[22]:

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(figsize=(14, 4))
gdf.Photon_Height.plot(ax=ax, ls='', marker='.', ms=0.01)
ax.set_xlabel('Along track distance (m)', fontsize=12);
ax.set_ylabel('Photon Height (m)', fontsize=12)
ax.set_title('ICESat-2 ATL03', fontsize=14)
ax.tick_params(axis='both', which='major', labelsize=12)
```



[ ]:

2.8 AfriSAR Search and Visualize

2.8.1 Description

This dataset provides gridded estimates of aboveground biomass (AGB) for four sites in Gabon at 0.25 ha (50 m) resolution derived with field measurements and airborne LiDAR data collected from 2010 to 2016. The sites represent a mix of forested, savannah, and some agricultural and disturbed landcover types: Lope site, within Lope National Park; Mabounie, mostly forested site; Mondah Forest, protected area; and the Rabi forest site, part of the Smithsonian Institution of Global Earth Observatories world-wide network of forest plots. Plot-level biophysical measurements of tree diameter and tree height (or estimated by allometry) were performed at 1 ha and 0.25 ha scales on multiple plots at each site and used to derive AGB for each tree and then summed for each plot. Aerial LiDAR scans were used to construct digital elevation models (DEM) and digital surface models (DSM), and then the DEM and DSM were used to construct a canopy height model (CHM) at 1 m resolution. After checking site-plot locations against the CHM, mean canopy height (MCH) was computed over each 0.25 ha. A single regression model relating MCH and AGB estimates, incorporating local height based on the trunk DBH (HD) relationships, was produced for all sites and combined with the CHM layer to construct biomass maps at 0.25 ha resolution.

Source: https://daac.ornl.gov/AFRISAR/guides/AfriSAR_AGB_Maps.html

[1]:
```
# Uncomment and run the line below if missing packages
# !pip install pystac-client rioxarray hvplot
```

Data type cannot be displayed: application/javascript, application/vnd.holoviews_load.v0+json

Data type cannot be displayed: application/javascript, application/vnd.holoviews_load.v0+json

[ ]:
```
import rioxarray
import rasterio as rio
from pystac_client import Client
import hvplot.xarray
```

(continues on next page)

(continued from previous page)

```
import warnings
warnings.filterwarnings("ignore")
```

2.8.2 Declare your collection of interest

You can discover available collections the follwoing ways:

1. Use the {STAC_API_URL}/collections API endpoint (JSON response)
2. Programmatically using pystac (see example in the list-collections.ipynb notebook)
3. In the STAC Browser : http://delta-staging-stac-browser.s3-website-us-east-1.amazonaws.com/

[2]:
```
STAC_API_URL = 'https://stac.maap-project.org/'
collection = 'AfriSAR_AGB_Maps_1681'
```

[3]:
```
# custom headers
headers = []

cat = Client.open(STAC_API_URL, headers=headers)
```

2.8.3 Listing all the available collections

[4]:
```
for collection in cat.get_collections():
    print(collection)
```

```
<CollectionClient id=AfriSAR_UAVSAR_Coreg_SLC>
<CollectionClient id=Landsat8_SurfaceReflectance>
<CollectionClient id=AfriSAR_AGB_Maps_1681>
<CollectionClient id=ABLVIS1B>
<CollectionClient id=GEDI02_B>
<CollectionClient id=AFLVIS2>
<CollectionClient id=BIOSAR1>
<CollectionClient id=AFRISAR_DLR>
<CollectionClient id=GEDI02_A>
```

2.8.4 Getting the collection of intrest 'AfriSAR_AGB_Maps_1681'

[5]:
```
collection = cat.get_collection('AfriSAR_AGB_Maps_1681')
collection
```

[5]:
```
<CollectionClient id=AfriSAR_AGB_Maps_1681>
```

2.8.5 Use satsearch to discover items in the 'AfriSAR_AGB_Maps_1681' collection

[6]:
```
search = cat.search(
    max_items = 15,
    limit = 5,
    collections = collection,
)
```

2.8.6 Adding items to an iterable

[7]:
```
items = [i.to_dict() for i in search.get_all_items()]
items[0]
```

[7]:
```
{'type': 'Feature',
 'stac_version': '1.0.0',
 'id': 'Rabi_AGB_50m',
 'properties': {'boxes': ['-1.94602 9.85914 -1.90031 9.90636'],
  'links': [{'rel': 'http://esipfed.org/ns/fedsearch/1.1/s3#',
    'href': 's3://nasa-maap-data-store/file-staging/nasa-map/AfriSAR_AGB_Maps_1681___
↪1/Rabi_AGB_50m.tif',
    'title': 'File to download',
    'hreflang': 'en-US'},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/metadata#',
    'href': 'https://daac.ornl.gov/AFRISAR/guides/AfriSAR_AGB_Maps.html',
    'title': 'ORNL DAAC Data Set Documentation (USERS GUIDE)',
    'hreflang': 'en-US'},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/metadata#',
    'href': 'https://dx.doi.org/10.3334/ORNLDAAC/1681',
    'title': 'Data set Landing Page DOI URL (DATA SET LANDING PAGE)',
    'hreflang': 'en-US'},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/metadata#',
    'href': 'https:/daac.ornl.gov/daacdata/afrisar/AfriSAR_AGB_Maps/comp/AfriSAR_AGB_
↪Maps.pdf',
    'title': 'Data Set Documentation (GENERAL DOCUMENTATION)',
    'hreflang': 'en-US'},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/metadata#',
    'href': 'https:/daac.ornl.gov/daacdata/afrisar/AfriSAR_AGB_Maps/comp/AfriSAR_AGB_
↪Maps_PlotDetails.csv',
    'title': 'Data Set Documentation (GENERAL DOCUMENTATION)',
    'hreflang': 'en-US'},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/data#',
    'href': 's3://nasa-maap-data-store/file-staging/nasa-map/AfriSAR_AGB_Maps_1681___1
↪',
    'hreflang': 'en-US',
    'inherited': True},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/documentation#',
    'href': 'https://daac.ornl.gov/AFRISAR/guides/AfriSAR_Mondah_Field_Data.html',
    'hreflang': 'en-US',
    'inherited': True},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/metadata#',
    'href': 'https://doi.org/10.3334/ORNLDAAC/1580',
    'hreflang': 'en-US',
    'inherited': True},
   {'rel': 'http://esipfed.org/ns/fedsearch/1.1/documentation#',
    'href': 'https://daac.ornl.gov/daacdata/afrisar/AfriSAR_Mondah_Field_Data/comp/
↪AfriSAR_Mondah_Field_Data.pdf',
    'hreflang': 'en-US',
    'inherited': True}],
```

(continues on next page)

(continued from previous page)

```
  'title': 'AfriSAR_AGB_Maps.Rabi_AGB_50m.tif',
  'updated': '2019-05-13T12:45:07+00:00',
  'datetime': '2016-02-01T00:00:00Z',
  'time_end': '2016-03-31T23:59:59.000Z',
  'proj:bbox': [595552.0, 9784881.0, 600802.0, 9789931.0],
  'proj:epsg': 32732.0,
  'concept_id': 'G1200115789-NASA_MAAP',
  'dataset_id': 'AfriSAR: Aboveground Biomass for Lope, Mabounie, Mondah, and Rabi␣
↪Sites, Gabon',
  'proj:shape': [101.0, 105.0],
  'time_start': '2016-02-01T00:00:00.000Z',
  'browse_flag': False,
  'data_center': 'NASA_MAAP',
  'granule_size': '0.014109',
  'proj:geometry': {'type': 'Polygon',
   'coordinates': [[[595552.0, 9784881.0],
     [600802.0, 9784881.0],
     [600802.0, 9789931.0],
     [595552.0, 9789931.0],
     [595552.0, 9784881.0]]]},
  'day_night_flag': 'BOTH',
  'proj:transform': [50.0,
   0.0,
   595552.0,
   0.0,
   -50.0,
   9789931.0,
   0.0,
   0.0,
   1.0],
  'original_format': 'ECHO10',
  'coordinate_system': 'CARTESIAN',
  'online_access_flag': False,
  'collection_concept_id': 'C1200115768-NASA_MAAP'},
 'geometry': {'type': 'Polygon',
  'coordinates': [[[9.859161573944016, -1.9460214484122889],
    [9.906363400458433, -1.9459965989464993],
    [9.906339295667408, -1.9003139362380275],
    [9.859138724101811, -1.900338201941798],
    [9.859161573944016, -1.9460214484122889]]]},
 'links': [{'rel': 'collection',
   'href': 'https://stac.maap-project.org/collections/AfriSAR_AGB_Maps_1681',
   'type': 'application/json'},
  {'rel': 'parent',
   'href': 'https://stac.maap-project.org/collections/AfriSAR_AGB_Maps_1681',
   'type': 'application/json'},
  {'rel': <RelType.ROOT: 'root'>,
   'href': 'https://stac.maap-project.org/',
   'type': <MediaType.JSON: 'application/json'>,
   'title': 'maap-stac'},
  {'rel': 'self',
   'href': 'https://stac.maap-project.org/collections/AfriSAR_AGB_Maps_1681/items/
↪Rabi_AGB_50m',
   'type': 'application/geo+json'}],
 'assets': {'data': {'href': 's3://nasa-maap-data-store/file-staging/nasa-map/AfriSAR_
↪AGB_Maps_1681___1/Rabi_AGB_50m.tif',
   'type': 'image/tiff; application=geotiff; profile=cloud-optimized',
```

(continues on next page)

(continued from previous page)

```
   'roles': ['data']},
  'metadata': {'href': 'https://daac.ornl.gov/AFRISAR/guides/AfriSAR_AGB_Maps.html',
   'roles': ['metadata']},
  'documentation': {'href': 'https://daac.ornl.gov/AFRISAR/guides/AfriSAR_Mondah_
↪Field_Data.html',
   'roles': ['documentation']}},
 'bbox': [9.859138724101811,
  -1.9460214484122889,
  9.906363400458433,
  -1.9003139362380275],
 'stac_extensions': ['https://stac-extensions.github.io/projection/v1.0.0/schema.json
↪',
  'https://stac-extensions.github.io/raster/v1.1.0/schema.json'],
 'collection': 'AfriSAR_AGB_Maps_1681'}
```

2.8.7 Extracting s3 link and reading it with rioxarray

[8]:
```
item = items[0]
s3_link = item['assets']['data']['href']
da = rioxarray.open_rasterio(s3_link)
da = da.squeeze('band', drop=True)
da
```

[8]:
```
<xarray.DataArray (y: 101, x: 105)>
[10605 values with dtype=float32]
Coordinates:
  * x            (x) float64 5.956e+05 5.956e+05 ... 6.007e+05 6.008e+05
  * y            (y) float64 9.79e+06 9.79e+06 9.79e+06 ... 9.785e+06 9.785e+06
    spatial_ref  int64 0
Attributes:
    AREA_OR_POINT:  Area
    _FillValue:     -9999.0
    scale_factor:   1.0
    add_offset:     0.0
```

2.8.8 Visualizing the read data with hvplot

[9]:
```
ds_masked = da.where(da != da._FillValue)

ds_masked.hvplot(
    'x', 'y',
    cmap='viridis',
    frame_height=400,
    frame_width=400
).redim.range(value=(0,da.max().values))
```

[9]:
```
:Image   [x,y]   (value)
```

[ ]:

[ ]:

[ ]:

CHAPTER 3

Technical Tutorials

3.1 Search

MAAP users are advised to use two catalogs:

1. Use NASA's Operational CMR to discover NASA-produced and curated data: https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html.
2. Use MAAP STAC for data not found in NASA CMR, and data produced by MAAP users: https://stac.maap-project.org/docs.

Warning: The https://cmr.maap-project.org catalog will be deprecated by **May 1, 2023**. Users should request collections they need from this catalog to be made discoverable in the MAAP STAC or NASA's Operational CMR if they're not already there.

More information on each catalog and migrating from MAAP's CMR here: MAAP's Dual Catalog.

3.1.1 MAAP's Dual Catalog

MAAP users are advised to use two catalogs:

1. Use NASA's Operational CMR to discover NASA-produced and curated data: https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html.
2. Use MAAP STAC for data not found in NASA CMR, and data produced by MAAP users: https://stac.maap-project.org/api.html.

Warning: The https://cmr.maap-project.org catalog will be deprecated by **May 1, 2023**. Users should request collections they need from this catalog to be made discoverable in the MAAP STAC or NASA's Operational CMR if they're not already there.

More information on each catalog and migrating from MAAP's CMR is detailed in the bottom of this page.

MAAP STAC

MAAP STAC (https://stac.maap-project.org) is dedicated to datasets not accessible via NASA's CMR, such as GEDI Cal/Val datasets, ESA datasets, and user-shared data products.

STAC discovery

Users can discover data in MAAP STAC using pystac-client or https://stac-browser.maap-project.org.

API documentation is available here: https://stac.maap-project.org/api.html (will return MAAP STAC results).

The general STAC API spec is here: https://api.stacspec.org/v1.0.0-rc.1/core/.

An example of using pystac-client is included above and in *Searching STAC Documentation*.

Data Access via STAC

Data assets (files) published to STAC have not moved from the S3 bucket `s3://nasa-maap-data-store`. ESA data is accessible via public HTTP access. NASA data in S3 is accessible publicly or via role-based bucket policy access.

Users are encouraged to use common AWS S3 libraries for NASA data access, such as Python's boto3.

Each item should have a "data" asset which includes a URL to the data.

For example, https://stac.maap-project.org/collections/BIOSAR1/items/biosar1_roi_lidar58 includes:

```
"assets": {
  "data": {
    "href": "https://bmap-catalogue-data.oss.eu-west-0.prod-cloud-ocb.orange-business.
↪com/Campaign_data/biosar1/biosar1_roi_lidar58.shx",
    "type": "application/octet-stream",
    "roles": [
      "data"
    ]
  },
}
```

NASA's Operational CMR

CMR Discovery

Users can discover data NASA's Operational CMR via its publicly accessible API: https://cmr.earthdata.nasa.gov and user interface: https://search.earthdata.nasa.gov.

CMR Search documentation can be found in *Searching Collections* and *Searching Granules* and https://cmr.earthdata.nasa.gov/search/site/docs/search/api.html.

CMR Access

For all NASA MAAP users, access to NASA'S Operational data is provided via a federated access token. Anything that is in NASA's Operational CMR should be accessed via maap-py so that the federated access token can be used. Users can also access data from LPDAAC (and possibly other DAACs in the future) without maap-py since the workspace should have access via a role-based bucket policy on the LPDAAC cloud bucket.

Anyone can access data through Earthdata Login as well.

Find more documentation about how to access data in CMR in the Access section of this documentation.

Migrating from MAAP's CMR

If you're migrating code from using https://cmr.maap-project.org, we're here to help. The documentation below should support migrating to https://cmr.earthdata.nasa.gov and https://stac.maap-project.org. If not, please contact the data team for assistance.

Migration Steps:

1. Identify where your code is using https://cmr.maap-project.org and which datasets are being discovered and accessed.
2. Once you've identified the datasets, use https://search.earthdata.nasa.gov or https://stac-browser.maap-project.org to find out if the dataset is available through NASA's Operational CMR or MAAP's STAC catalog. If you don't see your datasets in one of those places, reach out to the data team so they can prioritize that dataset for publication to MAAP STAC.
3. If the dataset is in NASA's Operational CMR and you're using MAAP's Python library `maap-py` to discover and access data, add the parameter `cmr_host="cmr.earthdata.nasa.gov"` to your `maap.searchCollection` and `maap.searchGranule` function calls. Update the `concept_id` to match the one from NASA's Operational CMR if you're using it to identify a specific collection or granule.
4. If the dataset is in MAAP STAC, use `pystac_client` (https://pystac-client.readthedocs.io/en/stable/) or an HTTP library to call the STAC HTTP API endpoints directly.

Examples:

Example of switching a granule search to NASA's Operational CMR:

The code below discovers granules from the `ABoVE LVIS L2 Geolocated Surface Elevation Product`:

```
COLLECTION_ID = 'C1200125288-NASA_MAAP'
results = maap.searchGranule(concept_id=COLLECTION_ID)
pprint(f'Got {len(results)} results')
```

This dataset exists in NASA's Operational CMR. Using https://search.earthdata.nasa.gov, I discovered the collection's `concept_id` by searching for "ABoVE LVIS L2 Geolocated Surface Elevation Product" and copying the `concept_id` from the URL of the result to modify the code below:

```
COLLECTION_ID = 'C1513105984-NSIDC_ECS'
results = maap.searchGranule(
  cmr_host='cmr.earthdata.nasa.gov',
```

(continues on next page)

(continued from previous page)

```
  concept_id=COLLECTION_ID
)
pprint(f'Got {len(results)} results')
```

Example of switching a granule search to MAAP STAC:

This code discovers granules from the `Landsat 8 Operational Land Imager (OLI) Surface Reflectance Analysis Ready Data (ARD) V1, Peru and Equatorial Western Africa, April 2013-January 2020`.

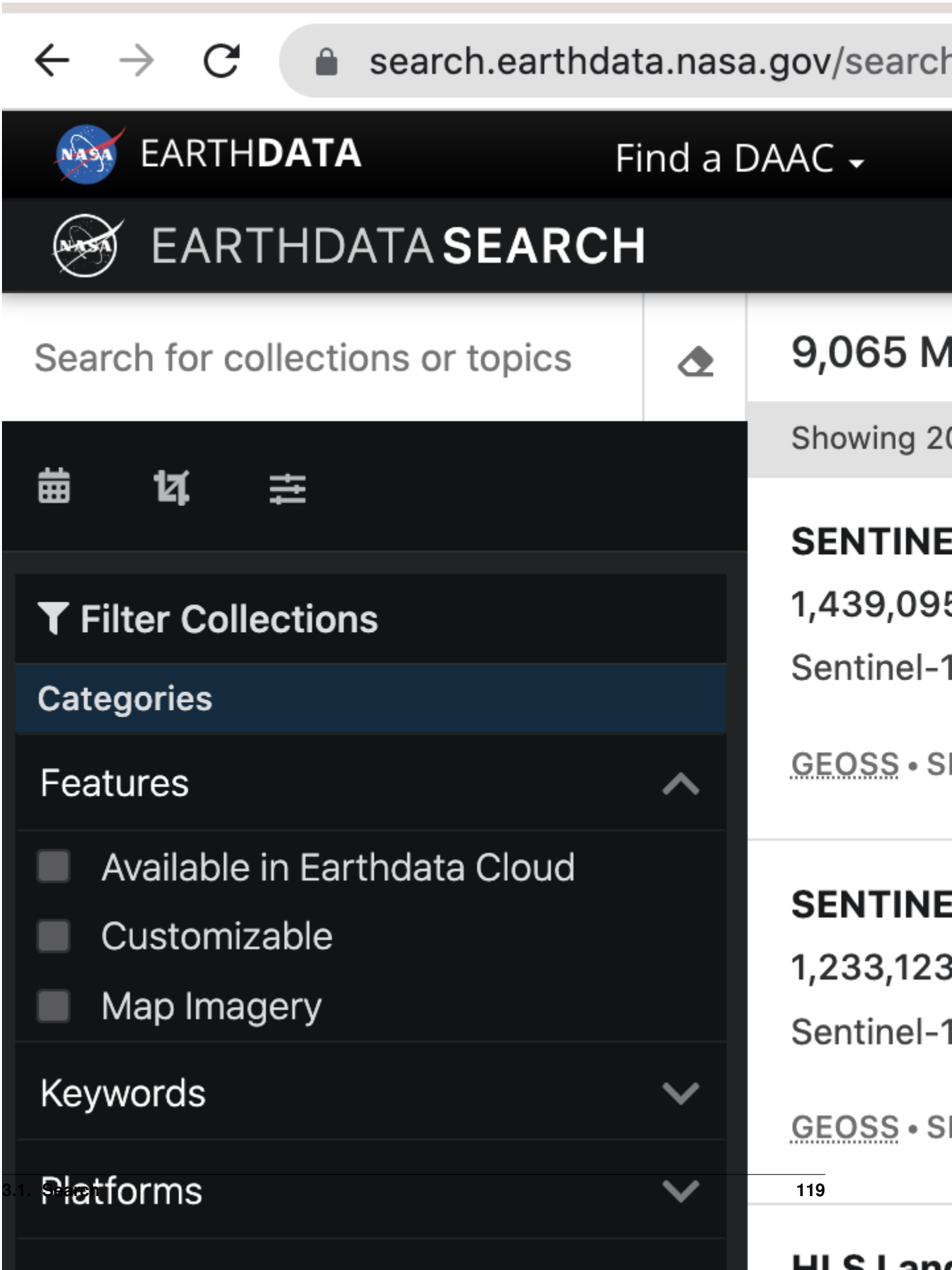
```
COLLECTION_ID = 'C1200110769-NASA_MAAP'

results = maap.searchGranule(concept_id=COLLECTION_ID)
pprint(f'Got {len(results)} results')
```

You can use https://stac-browser.maap-project.org to find the STAC collection ID for that dataset, which is `Landsat8_SurfaceReflectance`.

```
from pystac_client import Client
URL = 'https://stac.maap-project.org/'
cat = Client.open(URL)
for collection in cat.get_all_collections():
    print(collection)

collection = cat.get_collection('Landsat8_SurfaceReflectance')
items = collection.get_items()
```

3.1.2 Using the NASA Earthdata Search Client Graphical User Interface

The Earthdata Search Client (EDSC) allows users to search, preview, download, and access EOSDIS Earth observation data. NASA's EDSC is located at https://search.earthdata.nasa.gov. ESDC provides a graphical user interface (GUI) for NASA's Common Metadata Repository (CMR) to ease the process of discovering data.


search.earthdata.nasa.gov/search
EARTHDATA
Find a DAAC
EARTHDATA SEARCH
Search for collections or topics
9,065 M
Showing 20
Filter Collections
Categories
Features
Available in Earthdata Cloud
Customizable
Map Imagery
Keywords
Platforms
SENTINE
1,439,095
Sentinel-1
GEOSS • SI
SENTINE
1,233,123
Sentinel-1
GEOSS • SI

(*Image of the NASA EDSC GUI*)

Using the Earthdata Search Client

We can use the searchbar to search for a term (example: *GEDI*) to narrow the resulting collections. The search can be further refined by picking a temporal range from the calendar using the **Temporal** button and setting spatial boundaries using the **Spatial** button. Additionally, using the **Advanced search** lets you search by the HUC ID or region (more information about hydrological units may be found here). You can use the **Clear Filters** button to undo any filters that have been set. The sidebar to the left allows you to further refine your search by selecting one of the available facets. These include *Features*, *Keywords*, *Platforms*, *Instruments*, *Organizations*, *Projects*, and *Processing levels*.

We can also use the tools with the background map to refine our search. We can search spatially using **Search by spatial polygon**, **Search by spatial rectangle**, and **Search by spatial coordinate**. Layers may be edited using the **Edit layers** button and deleted using the **Delete layers** buttons. There are also options for **North Polar Stereographic**, **Geographic (Equirectangular)**, and **South Polar Stereographic** projections. There are options to **Zoom in** +, **Zoom home**, and **Zoom out** −. Finally, we can change the basemap by selecting the **Map layers** button.

The results of the search are displayed in the *Matching Collections* section. Collection names and summaries for each result are shown here. The **View collection details** button may be used to view related URLs and additional information about the selected collection. Also, collections may be added to a project using the **Add collection to the current project** button. Clicking anywhere else on a result allows you to see the granules within the collection available for download.

Using the Earthdata Search Client in the Algorithm Development Environment (ADE)

The EDSC can also be accessed within the ADE when using the MAAP JupyterLab IDE. When inside of a workspace, we can use the **Data Search** tab at the top of the Jupyter window to see the option to **Open EarthData Search**. We can also access the EDSC by selecting the **Commands** tab on the JupyterLab sidebar and scrolling down to the *Search* section where the **Open EarthData Search** option is located. Alternatively, we can use the search bar at the top of the JupyterLab sidebar to find this option, or press Ctrl+Shift+C to be taken directly to the search bar.

The JupyterLab Search extension is integrated with the embedded Earthdata Search Client allowing users to paste queries and query results into a notebook. This is accomplished by using the EDSC and navigating to the desired JupyterLab notebook. Next, use the **Data Search** tab at the top of the Jupyter window to see the **Paste Search Query** and **Paste Search Results** options can be selected (these can also be accessed by the **Commands** tab on the JupyterLab sidebar).

3.1.3 Searching for Collections in NASA's Operational CMR using maap-py

These examples walk through the MAAP API functionality of searching for collections based on specific parameters. Collections are groupings of files that share the same product specification. Searching for collections can be useful for finding individual files, known as granules, which are used for processing.

We begin by importing the `maap` package and creating a new MAAP class.

[1]:
```
# import the MAAP package to handle queries
from maap.maap import MAAP

# import printing package to help display outputs
from pprint import pprint
```

(continues on next page)

(continued from previous page)

```
# invoke the MAAP search client
maap = MAAP()
```

We can use the `maap.searchCollection` function to return a list of desired collections. Before using this function, let's use the `help` function to view the specific arguments and keywords for `maap.searchCollection`.

[2]:
```
# view help for the searchCollection function
help(maap.searchCollection)
```

```
Help on method searchCollection in module maap.maap:

searchCollection(limit=100, **kwargs) method of maap.maap.MAAP instance
    Search the CMR collections
    :param limit: limit of the number of results
    :param kwargs: search parameters
    :return: list of results (<Instance of Result>)
```

The help text is showing that `maap.searchCollection` accepts a limit and search parameters. The limit parameter limits the number of resulting collections returned by `maap.searchCollection`. Note that `limit=100` means that the *default limit* for results from the MAAP API is 100. `maap.searchCollection` accepts any additional search parameters that are included in the CMR. For a list of accepted parameters, please refer to the CMR Search Collections API reference.

In this example we will explore search options including:

1. Finding all Collections
2. Searching by temporal filter
3. Searching by spatial filter
4. Using the results from one search as inputs into another
5. Searching by additional attributes

Finding all Collections

Here we will demonstrate how to create a list containing all of the collections contained within the CMR. To do this, we will use the `maap.searchCollection` function without any additional search parameters.

[3]:
```
# search all collections
results = maap.searchCollection(cmr_host="cmr.earthdata.nasa.gov")

# print the number of collections
pprint(f'Got {len(results)} results')
```

```
'Got 100 results'
```

We get 100 results because of the default page limit. The result from the MAAP API is a list of collections where each element in the list is the metadata for that particular collection. To change the limit, type `limit=` and then a value within the parentheses after `maap.searchCollection()`.

Let's look at the metadata for the first collection in our list of results (`results[0]`) using `pprint`. For formatting purposes, we can use the `depth` parameter to control the number of levels of metadata detail to display. By default, there is no constraint on the depth. By setting a `depth` parameter (in this case `depth=2`), we can ensure that the next contained level is replaced by an ellipsis.

[9]:
```
# print the metadata for the first collection
# we use the depth parameter to set the layer of metadata detail in the results, with␣
↪(1) having the least detail
# (1) displays the concept ID, format, and revision ID
# adjust the depth to a larger value (6) if you would like to view all of the metadata
pprint(results[0], depth=2)
```

```
{'Collection': {'Campaigns': {...},
                'CollectionState': 'COMPLETE',
                'Contacts': {...},
                'DOI': {...},
                'DataSetId': '10 Days Synthesis of SPOT VEGETATION Images '
                             '(VGT-S10)',
                'Description': 'The VGT-S10 are near-global or continental, '
                               '10-daily composite images which are '
                               "synthesised from the 'best available' "
                               'observations registered in the course of every '
                               "'dekad' by the orbiting earth observation "
                               'system SPOT-VEGETATION. The products provide '
                               'data from all spectral bands (SWIR, NIR, RED, '
                               'BLUE), the NDVI and auxiliary data on image '
                               'acquisition parameters. The VEGETATION system '
                               'allows operational and near real-time '
                               'applications, at global, continental and '
                               'regional scales, in very broad environmentally '
                               'and socio-economically critical fields. The '
                               'VEGETATION instrument is operational since '
                               'April 1998, first with VGT1, from March 2003 '
                               'onwards, with VGT2. More information is '
                               'available on: '
                               'https://docs.terrascope.be/#/DataProducts/SPOT-VGT/
↪Level3/Level3',
                'InsertTime': '2023-04-07T08:27:18.514Z',
                'LastUpdate': '2023-04-07T08:27:18.514Z',
                'LongName': 'Not provided',
                'OnlineAccessURLs': {...},
                'OnlineResources': {...},
                'Platforms': {...},
                'ProcessingLevelId': 'NA',
                'ScienceKeywords': {...},
                'ShortName': 'urn:ogc:def:EOP:VITO:VGT_S10',
                'Spatial': {...},
                'Temporal': {...},
                'UseConstraints': {...},
                'VersionId': '1'},
 'concept-id': 'C2207472890-FEDEO',
 'format': 'application/echo10+xml',
 'revision-id': '5'}
```

The `Collection` key has all of the collection information including attributes, the archive center, spatial, and temporal information. The `concept-id` is a unique identifier for this collection. It can be used to further refine search results from the CMR, such as when searching for granule information.

Searching by Temporal Filter

Here we use a temporal filter to narrow down our results using the `temporal` keyword in our search. The temporal keyword takes datetime information in a specific format. The date format used is `YYYY-MM-DDThh:mm:ssZ` and

temporal search criteria may be either a single date or a date range. If one date is provided then it can be inferred as the start or end date. To define a start date and return all collections after the date, put a comma after the date (`YYYY-MM-DDThh:mm:ssZ,`). To define a end date and return all granules prior to the data, put a comma before the date (`,YYYY-MM-DDThh:mm:ssZ`). Lastly, to get a date range, provide the start date and end date separated by a comma (`YYYY-MM-DDThh:mm:ssZ,YYYY-MM-DDThh:mm:ssZ`).

In this example we will search for one month of data.

[5]:
```
datetimeRange = '2000-01-01T00:00:00Z,2000-01-31T23:59:59Z' # specify datetime range␣
↪to search for data in January 2000

results = maap.searchCollection(
    cmr_host = "cmr.earthdata.nasa.gov",
    temporal = datetimeRange
)
pprint(f'Got {len(results)} results')
```

```
'Got 100 results'
```

[6]:
```
collectionName = results[0]['Collection']['ShortName'] # get the collection short name
collectionDate = results[0]['Collection']['Temporal']['RangeDateTime'][
↪'BeginningDateTime'] # get the collection start time

pprint(
    f'Collection {collectionName} was acquired starting at {collectionDate}',␣
↪width=100)
```

```
'Collection GLDAS_NOAH025_3H was acquired starting at 2000-01-01T00:00:00.000Z'
```

It appears the first result correctly matches with the beginning and ending temporal search parameters. Keep in mind that the results are limited to 100 so the final collection returned may not match the end date that was searched for.

Searching by Spatial Filter

Here we will illustrate how to search for collections by a spatial filter. There are a couple of spatial filters available to search by in the CMR including point, line, polygon, and bounding box. In this example, we will explore filtering with a bounding box which is a sequence of four latitude and longitude values in the order of `[W,S,E,N]`.

[7]:
```
collectionDomain = '-42,10,42,20' # specify bounding box to search by

results = maap.searchCollection(
    cmr_host = "cmr.earthdata.nasa.gov",
    bounding_box = collectionDomain
)
pprint(f'Got {len(results)} results')
```

```
'Got 100 results'
```

[8]:
```
collectionName = results[3]['Collection']['ShortName'] # get a collection short name
collectionGeometry = results[3]['Collection']['Spatial']['HorizontalSpatialDomain'][
↪'Geometry'] # grab the spatial information from collection

pprint(f'Collection {collectionName} was acquired within the following geometry: ',␣
↪width=100)
pprint(collectionGeometry)
```

```
'Collection gov.noaa.nodc:0000029 was acquired within the following geometry: '
{'BoundingRectangle': {'EastBoundingCoordinate': '-16.25',
                       'NorthBoundingCoordinate': '46.263167',
                       'SouthBoundingCoordinate': '0.766667',
                       'WestBoundingCoordinate': '-124.041667'},
 'CoordinateSystem': 'CARTESIAN'}
```

We can see from the first collection that the spatial coordinates of the collection intersect our search box.

3.1.4 Searching for Granules in NASA's Operational CMR using maap-py

These examples will walk through the MAAP API functionality of searching granules within a collection based on specific parameters. Granules are individual files from a sensor where a group of granules make a collection within CMR. The granules are the raw data that will be used for processing.

We begin by importing the `maap` and `pprint` packages. Then invoke the `MAAP` constructor, setting the `maap_host` argument to `'api.maap-project.org'`.

[1]:
```
# import the MAAP package
from maap.maap import MAAP

# import printing package to help display outputs
from pprint import pprint

# invoke the MAAP constructor using the maap_host argument
maap = MAAP(maap_host='api.maap-project.org')
```

Here we view the specific arguments and keywords for the `maap.searchGranule` function.

[2]:
```
help(maap.searchGranule)
```

```
Help on method searchGranule in module maap.maap:

searchGranule(limit=20, **kwargs) method of maap.maap.MAAP instance
    Search the CMR granules

    :param limit: limit of the number of results
    :param kwargs: search parameters
    :return: list of results (<Instance of Result>)
```

As we can see from the result, `maap.searchGranule` accepts a limit keyword which limits the number of results from CMR. `maap.searchGranule()` also accepts any additional search parameters that are included in CMR. For a list of accepted parameters, please refer to the CMR Search Granules API reference.

It is important to note that *the default limit on results from the MAAP API is 20*. To increase the number of results we will specify a variable and use it in later queries.

[3]:
```
# get at max 500 results from CMR
MAX_RESULTS = 500
```

In this example we will explore search options including:

1. Searching by collection concept ID
2. Searching by temporal filter
3. Searching by spatial filter

4. Using the results from one search as inputs into another
5. Searching by additional attributes

For the next couple of examples, we will focus on the ICESat-2/ATLAS Land and Vegetation Height dataset.

Searching by Collection Short Name, Version

Here we will search by a short name and version which should uniquely identify a collection CMR. HOWEVER, some datasets exist both in the cloud and on-prem, so in the following example we actually get **2** results.

[4]:
```
atl08_collections = maap.searchCollection(
    short_name='ATL08',
    version='005',
    cmr_host='cmr.earthdata.nasa.gov'
)
len(atl08_collections)
```

[4]:
```
2
```

If you inspect the results, you will see the second result has distribution information which points to an S3 bucket location. You can see this information with the follow code: `atl08_collections[1]['Collection']['DirectDistributionInformation']`.

A simpler solution to finding just the cloud-hosted dataset is to add the `cloud_hosted="true"` parameter to our search.

[5]:
```
atl08_collections = maap.searchCollection(
    short_name='ATL08',
    version='005',
    cmr_host='cmr.earthdata.nasa.gov',
    cloud_hosted="true"
)
len(atl08_collections)
```

[5]:
```
1
```

Now we can look up the collection concept id to find only granules in the cloud-hosted ATL08 v005 dataset.

[6]:
```
COLLECTION_ID = atl08_collections[0]['concept-id']

results = maap.searchGranule(
    concept_id=COLLECTION_ID,
    cmr_host='cmr.earthdata.nasa.gov',
    limit=MAX_RESULTS)
pprint(f'Got {len(results)} results')
```

```
'Got 500 results'
```

We were able to get 500 results! There are most likely more than 500 granules in search results, but remember we limited the results to 500 granules. The result from the MAAP API is a list of granules where each element in the list is the metadata for that particular granule.

Now let's look at the metadata for the first result.

[7]:
```
# print the first granule's metadata
# we use the depth parameter to set the layer of metadata detail in the results, with␣
↪(1) having the least detail
# (1) displays the collection concept ID, concept ID, format, and revision ID
```

(continues on next page)

(continued from previous page)

```
# adjust the depth to a larger value (6) if you would like to view all of the metadata
pprint(results[0], depth=2)
```

```
{'Granule': {'AssociatedBrowseImageUrls': {...},
             'Collection': {...},
             'DataGranule': {...},
             'GranuleUR': 'ATL08_20181014001049_02350102_005_01.h5',
             'InsertTime': '2021-11-14T23:43:07.741Z',
             'LastUpdate': '2021-11-14T23:43:07.741Z',
             'OnlineAccessURLs': {...},
             'OnlineResources': {...},
             'OrbitCalculatedSpatialDomains': {...},
             'Spatial': {...},
             'Temporal': {...}},
 'collection-concept-id': 'C2153574670-NSIDC_CPRD',
 'concept-id': 'G2166182816-NSIDC_CPRD',
 'format': 'application/echo10+xml',
 'revision-id': '1'}
```

There is a lot of information in the metadata so let's break it down...

The `Granule` key has all of the granule information including attributes, browse imagery URLs, spatial, and temporal information. The `collection-concept-id` should match what you searched by and be the same for each granule. Lastly the granule specific `concept-id` is a unique identifier for this granule. This information can be used to further refine search results from CMR, specifically the granule information.

Searching by Temporal Filter

Here we will combine a search from earlier using the Collection Concept ID with a temporal filter to fine tune our results using the `temporal` keyword in our search.

The temporal keyword takes datetime information in a specific format. The date format used is `YYYY-MM-DDThh:mm:ssZ` and temporal search criteria may be either a single date or a date range. If one date is provided then it can be inferred as start or end date. To define a start date and return all granules after the date, put a comma after the date (`YYYY-MM-DDThh:mm:ssZ,`). To define an end date and return all granules prior to the data, put a comma before the date (`,YYYY-MM-DDThh:mm:ssZ`). Lastly, to get a date range, provide the start date and end date separated by a comma (`YYYY-MM-DDThh:mm:ssZ,YYYY-MM-DDThh:mm:ssZ`).

In this example we will search for one month of data.

[8]:
```
date_range = '2018-12-01T00:00:00Z,2018-12-31T23:59:59Z' # specify a date range to␣
↪search for data for Dec. 2018

results = maap.searchGranule(
    temporal=date_range,
    concept_id=COLLECTION_ID,
    limit=MAX_RESULTS,
    cmr_host="cmr.earthdata.nasa.gov"
)
pprint(f'Got {len(results)} results')
```

```
'Got 500 results'
```

[9]:
```
granuleFilename = results[0]['Granule']['DataGranule']['ProducerGranuleId'] # get the␣
↪granule file name
granuleDate = results[0]['Granule']['Temporal']['RangeDateTime']['BeginningDateTime']
↪# get the granule start time
```

(continues on next page)

(continued from previous page)

```
pprint(f'Granule {granuleFilename} was acquired starting at {granuleDate}',width=100)
```

```
'Granule ATL08_20181201001339_09680103_005_01.h5 was acquired starting at 2018-12-
↪01T00:13:48.477Z'
```

It looks like the first result correctly matches with the beginning temporal search parameter. Keep in mind that the results are limited to 500 so the final granule returned may not match the end date that was searched for.

Searching by Spatial Filter

Here we will illustrate how to search for granules by a spatial filter. There are a couple of spatial filters available to search by in CMR including point, line, polygon, and bounding box. The most simple to use is the bounding box which is a sequence of four latitude and longitude values in the order of `[W,S,E,N]`. In this example, we are going to search for data over Gabon using the `bounding_box` keyword.

[10]:
```
granule_bbox = '8.79799563969,-3.97882659263,14.4254557634,2.32675751384' # specify␣
↪bounding box to search by

COLLECTION_ID = 'C1000000240-LPDAAC_ECS' # Collection title: "NASA Shuttle Radar␣
↪Topography Mission Global 1 arc second V003"

results = maap.searchGranule(
    concept_id=COLLECTION_ID,
    bounding_box=granule_bbox,
    cmr_host="cmr.earthdata.nasa.gov"
)
pprint(f'Got {len(results)} results')
```

```
'Got 20 results'
```

[11]:
```
granule_filename = results[0]['Granule']['DataGranule']['ProducerGranuleId'] # get␣
↪the granule file name
geometry = results[0]['Granule']['Spatial']['HorizontalSpatialDomain']['Geometry'] #␣
↪grab the spatial information from granule

pprint(f'Granule {granule_filename} was acquired within the following geometry: ',␣
↪width=100)
pprint(geometry)
```

```
'Granule S03E012.SRTMGL1.hgt.zip was acquired within the following geometry: '
{'BoundingRectangle': {'EastBoundingCoordinate': '13.00027778',
                       'NorthBoundingCoordinate': '-1.99972222',
                       'SouthBoundingCoordinate': '-3.00027778',
                       'WestBoundingCoordinate': '11.99972222'}}
```

We can see from the first granule that the spatial coordinates of the granule intersect our search box.

The MAAP API provides rich functionality to interact with the CMR instance within the MAAP platform. Users can search datasets programmatically by many parameters and even combine parameters such as spatial and temporal filters to refine results.

3.1.5 Searching for and Compiling a List of Granule IDs for Batch Processing

While using the MAAP ADE, you may wish to create a list of granule IDs to be used for batch processing, granules being the individual files from a sensor that are used for processing. For this example, we will imagine a scenario that

we wish to produce a biomass estimate for a single HLS tile, and then expand that estimate over a larger area. In order to produce this expanded estimate, we will create a list of the HLS files which fall within a certain area.

We start by importing the `maap` and `pprint` packages and creating a new MAAP class.

[1]:
```
# import the MAAP package
from maap.maap import MAAP

# import printing package to help display outputs
from pprint import pprint

# create MAAP class
maap = MAAP()
```

We can use the `searchGranule` function to search for "HLS Landsat Operational Land Imager Surface Reflectance and TOA Brightness Daily Global 30m v2.0" granules. Click here for more information about searching for granules with the `searchGranule` function. Since the default limit on results from the MAAP API is 20, we specify a variable to use in our search query.

[2]:
```
# get at max 1000 results from CMR
MAX_RESULTS = 1000
```

To filter our search to HLS data, we create a variable with the collection ID of the HLS collection. Using the collection concept ID is the preferred method to filter by a collection, as it is a unique identifier which avoids ambiguity.

[3]:
```
COLLECTION_ID = 'C2021957657-LPCLOUD' # specifying the collection ID for the HLS␣
↪dataset
```

Next, we search for granules using the collection ID and a spatial filter. We can use a bounding box as our spatial filter. The bounding box is a sequence of four latitude and longitude values in the order of [W,S,E,N]. For this example, let's search for granules from the HLS data using a bounding box for the country Peru.

[4]:
```
collection_bbox = '-81.4109425524,-18.3479753557,-68.6650797187,-0.0572054988649' #␣
↪specify bounding box to search by

# getting results from granule search using the bounding box, collection ID, and␣
↪results limit
results = maap.searchGranule(
    cmr_host='cmr.earthdata.nasa.gov',
    bounding_box=collection_bbox,
    concept_id=COLLECTION_ID,
    limit=MAX_RESULTS
)
pprint(f'Got {len(results)} results') # print the number of results
```

```
'Got 1000 results'
```

We were able to get 1000 results. Each element in the list `results` contains the metadata for one of the granules returned by the search. Within this metadata is the key `concept-id`, which is the unique identifier for each granule. To create a list of granule IDs, we create a new list and use a for loop to add the `concept-id` from each element of `results` into the new list.

[5]:
```
granuleID_list = [] # create list for granule IDs

for result in results: # loop through each element (granule) in the list `results`
    granuleID_list.append(result['concept-id']) # add the concept id for each granule␣
↪to `granuleID_list`
```

(continues on next page)

(continued from previous page)

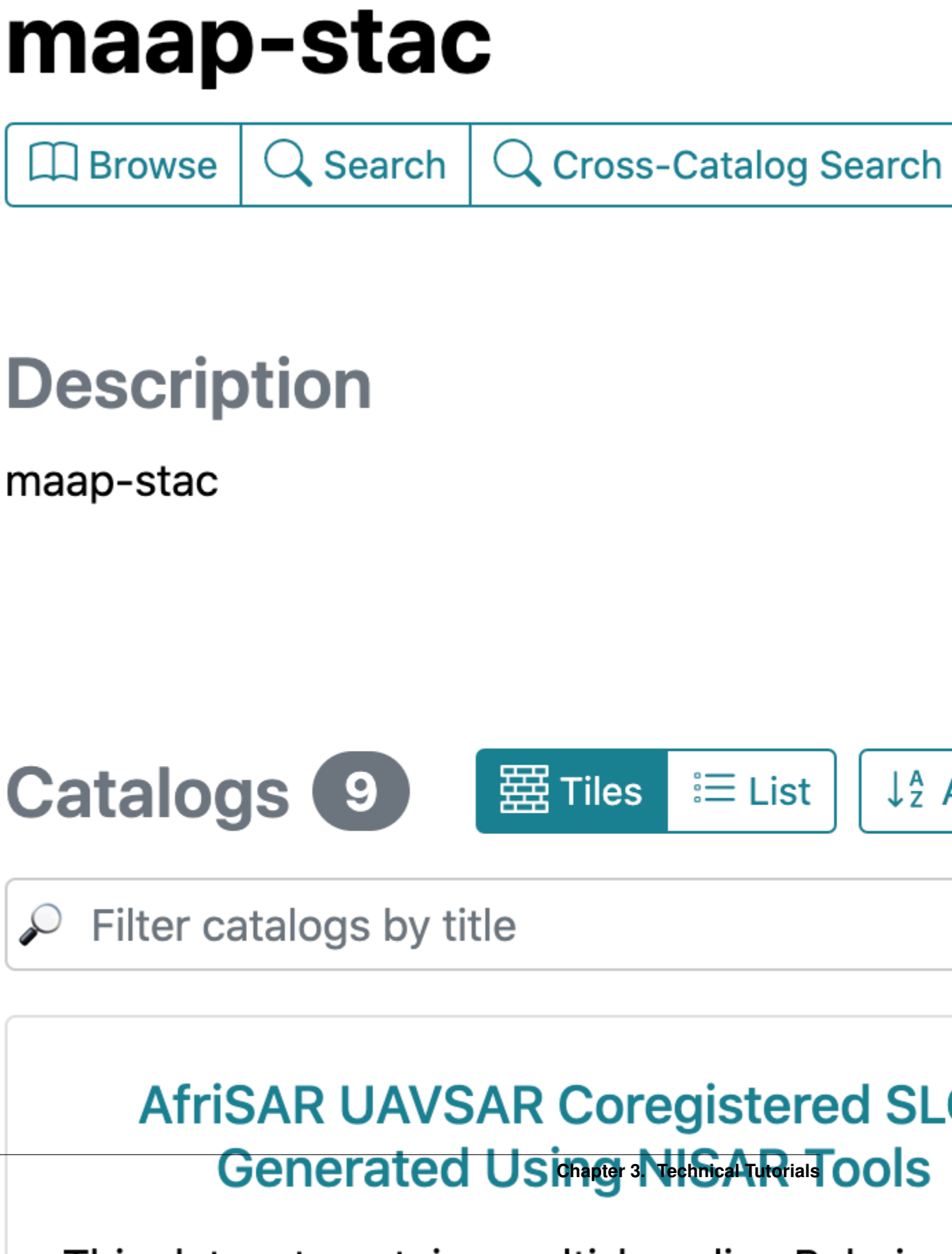
```
# You can uncomment the line below to see the result
# print(granuleID_list)
```

Now we have a list of all the granule IDs for granules in the Landsat 8 collection that fall within the bounding box for the country Peru within `granuleID_list`.

3.1.6 Searching the STAC Catalog

This tutorial provides a basic introduction to searching the MAAP STAC catalog (https://stac.maap-project.org/) using `pystac-client`.

Another method of searching the STAC catalog is via the STAC browser.


maap-stac
Browse
Search
Cross-Catalog Search
Description
maap-stac
Catalogs 9
Tiles
List
Filter catalogs by title
AfriSAR UAVSAR Coregistered SL
Generated Using NISAR Tools

About the STAC Catalog

At this time, the STAC catalog provides discovery of a subset of MAAP datasets. These datasets were selected because MAAP CMR analytics indicated selected datasets were being searched for the most. The data files have not been moved at all in the process of publishing datasets to STAC.

Data will continue to be added to the STAC catalog with priority given to datasets which are known to be in-use by MAAP UWG members through CMR metrics, S3 metrics, direct collaboration with data team members and by request.

Prerequisites

- pystac-client
- rioxarray (for opening a raster as an xarray dataset)

Authorship

- Author: Aimee Barciauskas
- Date: December 13, 2022
- Resources used: https://pystac-client.readthedocs.io/en/stable/tutorials/pystac-client-introduction.html

[1]:
```
%%capture
!pip install -U pystac-client
```

[2]:
```
from pystac_client import Client
```

STAC Client

We first connect to an API by retrieving the root catalog, or landing page, of the API with the Client.open function.

[3]:
```
# STAC API root URL
URL = 'https://stac.maap-project.org/'
cat = Client.open(URL)
cat
```

[3]:
```
<Client id=stac-fastapi>
```

CollectionClient

As with a static catalog the get_collections function will iterate through the Collections in the Catalog. Notice that because this is an API it can get all the Collections through a single call, rather than having to fetch each one individually.

[4]:
```
for collection in cat.get_all_collections():
    print(collection)
```

```
<CollectionClient id=GlobCover_09>
<CollectionClient id=AfriSAR_UAVSAR_Coreg_SLC>
<CollectionClient id=BIOSAR1>
<CollectionClient id=AfriSAR_AGB_Maps_1681>
<CollectionClient id=Landsat8_SurfaceReflectance>
<CollectionClient id=ABLVIS1B>
<CollectionClient id=AfriSAR_UAVSAR_Ungeocoded_Covariance>
<CollectionClient id=GlobCover_05_06>
<CollectionClient id=Global_PALSAR2_PALSAR_FNF>
<CollectionClient id=GEDI02_A>
```

(continues on next page)

(continued from previous page)

```
<CollectionClient id=AFLVIS2>
<CollectionClient id=AFRISAR_DLR>
<CollectionClient id=icesat2-boreal>
<CollectionClient id=AfriSAR_UAVSAR_Normalization_Area>
<CollectionClient id=GEDI_CalVal_Lidar_Data_Compressed>
<CollectionClient id=GEDI_CalVal_Field_Data>
<CollectionClient id=ESACCI_Biomass_L4_AGB_V3_100m_2010>
<CollectionClient id=ESACCI_Biomass_L4_AGB_V3_100m_2018>
<CollectionClient id=ESACCI_Biomass_L4_AGB_V3_100m_2017>
<CollectionClient id=AfriSAR_UAVSAR_Geocoded_SLC>
<CollectionClient id=Global_PALSAR2_PALSAR_Mosiac>
<CollectionClient id=AFRISAR_DLR2>
<CollectionClient id=AfriSAR_UAVSAR_KZ>
<CollectionClient id=GEDI_CalVal_Lidar_Data>
<CollectionClient id=GEDI02_B>
<CollectionClient id=ABoVE_UAVSAR_PALSAR>
<CollectionClient id=Global_Forest_Change_2000-2017>
```

[5]:
```
collection = cat.get_collection('ESACCI_Biomass_L4_AGB_V3_100m_2010')
collection
```

[5]:
```
<CollectionClient id=ESACCI_Biomass_L4_AGB_V3_100m_2010>
```

STAC Items

The main functions for getting items return iterators, where pystac-client will handle retrieval of additional pages when needed. Note that one request is made for the first ten items, then a second request for the next ten.

[6]:
```
items = collection.get_items()

# flush stdout so we can see the exact order that things happen
def get_ten_items(items):
    for i, item in enumerate(items):
        print(f"{i}: {item}")
        if i == 9:
            return

print('First page', flush=True)
get_ten_items(items)
```

```
First page
0: <Item id=S50W080_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0>
1: <Item id=S50W080_ESACCI-BIOMASS-L4-AGB-MERGED-100m-2010-fv3.0>
2: <Item id=S50W070_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0>
3: <Item id=S50W070_ESACCI-BIOMASS-L4-AGB-MERGED-100m-2010-fv3.0>
4: <Item id=S50W060_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0>
5: <Item id=S50W060_ESACCI-BIOMASS-L4-AGB-MERGED-100m-2010-fv3.0>
6: <Item id=S40W080_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0>
7: <Item id=S40W080_ESACCI-BIOMASS-L4-AGB-MERGED-100m-2010-fv3.0>
8: <Item id=S40W070_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0>
9: <Item id=S40W070_ESACCI-BIOMASS-L4-AGB-MERGED-100m-2010-fv3.0>
```

Discover the URL of one item using xarray

[7]:
```
item = collection.get_item('S50W080_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-2010-fv3.0')
item.assets['data'].href
```

[7]:
```
'https://bmap-catalogue-data.oss.eu-west-0.prod-cloud-ocb.orange-business.com/esacci/
↪ESACCI_Biomass_L4_AGB_V3_100m/2010/S50W080_ESACCI-BIOMASS-L4-AGB_SD-MERGED-100m-
↪2010-fv3.0.tif'
```

3.2 Visualize

3.2.1 Visualizing Web Map Tile Service (WMTS) Layers

At this time, there are two collections (AFLVIS2 and AfriSAR_UAVSAR_Coreg_SLC) for which we can visualize WMTS layers. WMTS layers can be visualized using the Common Mapping Client (CMC). The CMC is a starter-kit for web-based mapping applications which utilizes several common mapping funtionalities. This example demonstrates how to visualize WMTS layers using `ipyCMC`, a Jupyter Lab widget for the CMC.

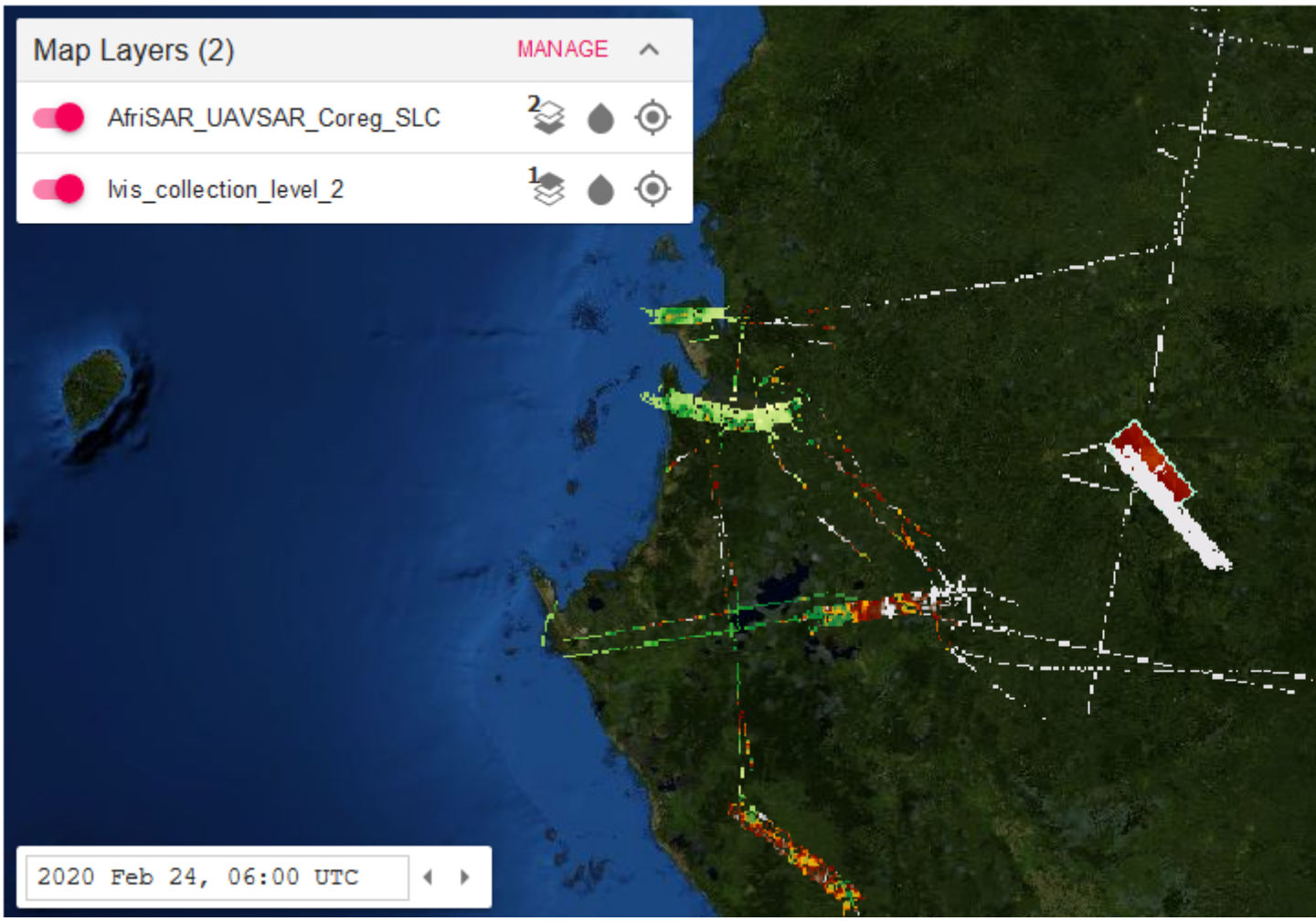
First, we import the `ipycmc` package.

[1]:
```
# Import the ipycmc module
import ipycmc
```

We utilize the CMC widget to visualize data in the MAAP Algorithm Development Environment (ADE).

[2]:
```
# utilize the CMC widget
w = ipycmc.MapCMC()
w
```

```
MapCMC()
```



The CMC user interface provides a variety of tools for analyzing data. In the top left corner, the Map Layers drop-down menu allows us to turn map layers on and off and provides tools for managing layer positioning, setting layer opacity, and zooming to a layer. At the lower right end of the display are tools for displaying the data, adjusting the zoom level, and selecting data. The 'Switch to 3D map' button allows us to switch between 2D and 3D maps. The 'Home' button zooms to the global extent and the 'Zoom In' and 'Zoom Out' buttons zoom the data to appear nearer or farther away. The 'Select Area' button allows us to select an area using a variety of geometrical methods, the 'Select Basemap' button allows us to select from a list of available basemaps, and the 'Fullscreen' button toggles full screen mode. At the bottom right, the cursor location on the map is displayed in Latitude/Longitude decimal degrees. At the bottom left of the user interface, there is a text box allowing us to change the date and time and arrows to change the day.

Visualizing Collections

To visualize the available collections, we utilize the `load_layer_config()` function, handling the url as a WMTS xml file.

[5]:
```
# visualize the available set of layers
w.load_layer_config("https://api.maap-project.org/api/wmts/GetCapabilities", "wmts/xml
↪")
```

Once the layers load, the number displayed in the Map Layers drop-down menu should increase. Pressing the downward arrow on the drop-down menu displays the loaded layers. Turn the layers on and off by pressing the toggle to the

left of the layer names and zoom to a layer by pressing the crosshair button.

Visualizing Single Granules

We can also visualize a single granule which is within one of the available collections by using the granule UR with the `load_layer_config()` function and handling the url as a WMTS file.

[6]:
```
# visualize a single granule using the granule UR
w.load_layer_config("https://api.maap-project.org/api/wmts/GetCapabilities?short_
↪name=AFLVIS2&granule_ur=SC:AFLVIS2.001:138348905", "wmts/xml")
```

This example here shows how we can add a layer by providing a granule UR. These granule URs can be extracted from the metadata by searching using CMR and the MAAP API. Please see the search granule example for specifics on how to search for granules and extract the UR - https://docs.maap-project.org/en/latest/technical_tutorials/search/granules.html.

3.2.2 Visualizing MAAP STAC Dataset with MosaicJSON

Authors: Samuel Ayers (UAH), Alex Mandel (DevSeed), Aimee Barciauskas (DevSeed)

Date: July 23, 2021

Description: In this notebook, we visualize SRTM Cloud-Optimized GeoTIFFs (COGs) from MAAP's STAC using a generated mosaic from MAAP's TiTiler.

Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

Note About the Data

The NASA Shuttle Radar Topographic Mission (SRTM) has provided digital elevation data (DEMs) for over 80% of the globe. This data is currently distributed free of charge by USGS and is available for download from the National Map Seamless Data Distribution System, or the USGS FTP site.

At MAAP, we've converted this elevation data into Cloud-Optimized GeoTIFFs (COGs) so they can be efficiently queried and visualized. These COGs are available in the MAAP STAC.

Additional Resources

- USGS EROS Archive - Shuttle Radar Topography Mission (SRTM)
- TiTiler API Documentation
- Github - MosaicJSON
- Working with MosaicJSON - TiTiler (DevSeed Example)

Importing and Installing Packages

To be able to run this notebook you'll need the following requirements:

- rasterio
- folium
- requests
- tqdm
- rio-tiler (2.0b8) (Optional)
- cogeo-mosaic (Optional)

If the packages below are not installed already, uncomment the following cell:

[ ]:
```
# !pip install rasterio folium requests tqdm
# !pip install rio-tiler cogeo-mosaic --pre
```

[ ]:
```
import requests
from pprint import pprint

from rio_tiler.io import COGReader
from rasterio.features import bounds as featureBounds

from folium import Map, TileLayer, GeoJson
```

Fetch SRTM COG STAC items

[2]:
```
stac_endpoint = "https://stac.maap-project.org/search"

stac_json = {
    "collections": ["SRTMGL1_COD"],
    "bbox": [4,42,16,48],
    "limit": 120
}

headers = {
            "Content-Type": "application/json",
            "Accept": "application/geo+json, application/json",
        }

# Read Items

r_stac = requests.post(stac_endpoint, headers=headers, json=stac_json)
items = r_stac.json()
```

Map the data bounds:

[3]:
```
geojson = {'type': 'FeatureCollection', 'features': items['features']}

bounds = featureBounds(geojson)
zoom_start = 5

m = Map(
    tiles="OpenStreetMap",
```

(continues on next page)

(continued from previous page)

```
    location=((bounds[1] + bounds[3]) / 2,(bounds[0] + bounds[2]) / 2),
    zoom_start=zoom_start
)

geo_json = GeoJson(
    data=geojson,
    style_function=lambda x: {
        'opacity': 1, 'dashArray': '1', 'fillOpacity': 0, 'weight': 1
    },
)
geo_json.add_to(m)
m
```

[3]:
```
<folium.folium.Map at 0x7f4c0e88f050>
```

Create Mosaic

We're using the TiTiler deployed by MAAP

[4]:
```
titiler_endpoint = "https://titiler.maap-project.org"  # MAAP titiler endpoint
```

To begin, we'll pull our COG:

[5]:
```
%time
from rio_tiler.io import COGReader

first_cog = geojson['features'][0]['assets']['cog_default']['href']
with COGReader(first_cog) as cog:
    info = cog.info()
```

```
CPU times: user 2 µs, sys: 2 µs, total: 4 µs
Wall time: 7.87 µs
```

Next, we will create the mosaic using the geojson from above. SRTMGL1 COGs have a "cog default" asset type, so we can create a mosaic of these type="image/tiff" assets. We can get access to these assets by using the accessor function.

[6]:
```
from cogeo_mosaic.mosaic import MosaicJSON

mosaicdata = MosaicJSON.from_features(geojson.get('features'), minzoom=6,␣
↪maxzoom=info.maxzoom, accessor=lambda feature : feature['assets']['cog_default'][
↪'href'])
```

Now we will upload the mosaicjson to the TiTiler:

[7]:
```
r = requests.post(
    url=f"{titiler_endpoint}/mosaics",
    headers={
        "Content-Type": "application/vnd.titiler.mosaicjson+json",
    },
    json=mosaicdata.dict(exclude_none=True)).json()

pprint(r)
```

```
{'id': '90b3f48f-d34d-4ee0-9e32-13c27c6325a3',
 'links': [{'href': 'https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3',
```

(continues on next page)

(continued from previous page)

```
            'rel': 'self',
            'title': 'Self',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3/mosaicjson',
            'rel': 'mosaicjson',
            'title': 'MosaicJSON',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3/tilejson.json',
            'rel': 'tilejson',
            'title': 'TileJSON',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3/tiles/{z}/{x}/{y}',
            'rel': 'tiles',
            'title': 'Tiles',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3/WMTSCapabilities.xml',
            'rel': 'wmts',
            'title': 'WMTS',
            'type': 'application/json'}]}
```

The response from the post request gives endpoints for different services (eg. mosaicjson, tilejson, tiles, wmts, etc). We're fetching the `tilejson` endpoint.

[8]:
```
tilejson_endpoint = list(filter(lambda x: x.get("rel") == "tilejson", dict(r)["links
↪"]))
```

Display Tiles

NOTE: You have to zoom to "minzoom" to load the data.

SECOND NOTE: The important bit is the "tiles" endpoint returned from `f"{titiler_endpoint}/mosaics`. This endpoint (e.g. `https://titiler.maap-project.org/mosaics/4199126b-9313-435a-b4b5-17802716b7b1/tiles/{z}/{x}/{y}`) could be used in any map client which can tile `x,y,z` layers.

[9]:
```
r_te = requests.get(
    tilejson_endpoint[0].get('href')
).json()

pprint(r_te)

tiles = TileLayer(
    tiles=f"{r_te['tiles'][0]}?rescale=0,4000",
    min_zoom=r_te["minzoom"],
    max_zoom=r_te["maxzoom"],
    opacity=1,
    attr="USGS"
)

tiles.add_to(m)
m
```

```
{'bounds': [2.999861111111111,
            40.999861111111116,
            17.000138888888888,
            49.000138888888889],
 'center': [10.0, 45.0, 6],
 'maxzoom': 12,
 'minzoom': 6,
 'name': '90b3f48f-d34d-4ee0-9e32-13c27c6325a3',
 'scheme': 'xyz',
 'tilejson': '2.2.0',
 'tiles': ['https://titiler.maap-project.org/mosaics/90b3f48f-d34d-4ee0-9e32-
↪13c27c6325a3/tiles/{z}/{x}/{y}@1x'],
 'version': '1.0.0'}
```

[9]:
```
<folium.folium.Map at 0x7f4c0e88f050>
```

3.2.3 Interval Color Mapping

Author(s): Aimee Barciauskas (DevSeed)

Date: March 3, 2023

Description: In this tutorial, we will pull from a SpatioTemporal Asset Catalog (STAC) collection containing cloud optimized geotiffs (COG) to create a colormap using CSS RGBA values with the COG's data values. We will then use the custom colormap to visualize the Global Aboveground Biomass (AGB) density estimates.

Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

Additional Resources

- Predefined Color Maps
- CSS RGBA Colors
- Using Python Zip Function

Importing and Installing Packages

We will begin by installing any packages we need and importing the packages that we will use.

Prerequisites

- branca
- folium

[ ]:
```
# %pip install folium
# %pip install branca
```

[1]:
```
import branca
from folium import Map, TileLayer
import json
from matplotlib import cm
import requests
```

Discover Data from STAC

[2]:
```
stac_endpoint = "https://stac.maap-project.org"
titiler_endpoint = "https://titiler.maap-project.org"
collection = "NASA_JPL_global_agb_mean_2020"
item = "SAmerica"

items_response = requests.get(f"{stac_endpoint}/collections/{collection}/items/{item}
↪").json()
url = items_response['assets']['mean']['href']
url
```

[2]:
```
's3://nasa-maap-data-store/file-staging/nasa-map/NASA_JPL_global_agb_mean_2020/global_
↪008_06dc_agb_mean_prediction_2020_mosaic_veg_gfccorr_scale1_SAmerica_cog.tif'
```

Get Data Values Using `/statistics` Endpoint

[3]:
```
# You can use gdalinfo /vsis3/nasa-maap-data-store/file-staging/nasa-map/NASA_JPL_
↪global_agb_mean_2020/global_008_06dc_agb_mean_prediction_2020_mosaic_veg_gfccorr_
↪scale1_SAmerica_cog.tif -stats
# or you can get metadata from titiler.
stats_response = requests.get(
    f"{titiler_endpoint}/cog/statistics",
    params = {
        "url": url
    }
).json()
```

[4]:
```
bins = stats_response['b1']['histogram'][1]
bin_ranges = [[bins[i], bins[i+1]] for i in range(len(bins)-1)]
bin_ranges
```

[4]:
```
[[-1166.0, -882.5],
 [-882.5, -599.0],
 [-599.0, -315.5],
 [-315.5, -32.0],
 [-32.0, 251.5],
 [251.5, 535.0],
 [535.0, 818.5],
 [818.5, 1102.0],
 [1102.0, 1385.5],
 [1385.5, 1669.0]]
```

Pick a Color Map and Create a Linear Mapping

[5]:
```
# There are many pre-defined colormaps supported by matplotlib.
# Some are listed below but a complete list be found here: https://matplotlib.org/3.1.
↪0/tutorials/colors/colormaps.html
# You may define custom color maps, but using the predefined ones makes life easier.
sequential_cmaps = [
    'Greys', 'Purples', 'Blues', 'Greens', 'Oranges', 'Reds',
    'YlOrBr', 'YlOrRd', 'OrRd', 'PuRd', 'RdPu', 'BuPu',
    'GnBu', 'PuBu', 'YlGnBu', 'PuBuGn', 'BuGn', 'YlGn'
]

sequential_cmaps2 = [
    'binary', 'gist_yarg', 'gist_gray', 'gray', 'bone', 'pink',
    'spring', 'summer', 'autumn', 'winter', 'cool', 'Wistia',
    'hot', 'afmhot', 'gist_heat', 'copper']
```

[9]:
```
# Here we create a list of pairs, each pair containing a data value interval range␣
↪(aka "bin")
# with a color value in RGBA (see https://www.w3schools.com/css/css3_colors.asp)
# First we create a list of colors
rgbas = [[int(value) for value in rgb] for rgb in cm.ScalarMappable(cmap='gist_earth
↪').to_rgba(x=bins[:-1], bytes=True)]
# Then we use python's zip function to pair rgba values with data values (https://www.
↪w3schools.com/python/ref_func_zip.asp)
color_map = list(zip(bin_ranges, rgbas))
# some tweaking may be necessary
color_map
```

[9]:
```
[([-1166.0, -882.5], [0, 0, 0, 255]),
 ([-882.5, -599.0], [18, 48, 119, 255]),
 ([-599.0, -315.5], [37, 102, 124, 255]),
 ([-315.5, -32.0], [54, 135, 111, 255]),
 ([-32.0, 251.5], [67, 151, 77, 255]),
 ([251.5, 535.0], [123, 167, 82, 255]),
 ([535.0, 818.5], [169, 179, 91, 255]),
 ([818.5, 1102.0], [191, 164, 100, 255]),
 ([1102.0, 1385.5], [221, 186, 167, 255]),
 ([1385.5, 1669.0], [253, 250, 250, 255])]
```

[7]:
```
# Let's also create a legend using the RGBA values and bins so our map visualization␣
↪can be interpreted!
legend = branca.colormap.StepColormap(rgbas, index=bins, vmin=round(bins[0], 2),␣
↪vmax=round(bins[-1], 2))
```

Preview the data

[10]:
```
# Create a json string of the colormap, so it can be passed as a parameter to titiler
↪'s API.
cmap = json.dumps(color_map)

# We fetch tilejson from titiler endpoint, to build a better map with appropriate␣
↪bounds and zoom level
tilejson_response = requests.get(
    f"{titiler_endpoint}/cog/tilejson.json",
    params = {
        "url": url,
```

(continues on next page)

(continued from previous page)

```
        "colormap": cmap
    }
).json()

bounds = tilejson_response["bounds"]
m = Map(
    location=((bounds[1] + bounds[3]) / 2,(bounds[0] + bounds[2]) / 2),
    zoom_start=tilejson_response["minzoom"] + 1
)

# We add a TileLayer using the tilejson_response "tiles" value which is the XYZ␣
↪endpoint of titiler.
aod_layer = TileLayer(
    tiles=tilejson_response["tiles"][0],
    opacity=1,
    attr="SAmerica"
)
aod_layer.add_to(m)

# Finally, we add the legend.
legend.caption = 'Global Aboveground Biomass (AGB) Density Estimates in Mg/ha␣
↪(megagram per hectare)'
legend.add_to(m)

m
```

[10]:
```
<folium.folium.Map at 0x10877c890>
```

3.2.4 Algorithm Development Environment (ADE) Visualization Example

MosaicJSON

A common challenge in visualizing spatial data is data is stored across many files representing small spatial extents.

MosaicJSON is a specification created by DevelopmentSeed which aims to be an open standard for representing metadata about a spatial mosaic of many COG files.

> MosaicJSON can be seen as a cloud friendly virtual raster (see GDAL's VRT) enabling spatial and temporal indexing for a list of Cloud-Optimized GeoTIFF.

Ref: https://github.com/developmentseed/mosaicjson-spec

Visualizing COGs generated in the MAAP Algorithm Development Environment (ADE)

This notebook visualizes COGs generated in the ADE using the python Common Mapping Client**, an open source project of NASA and JPL.**

Any Cloud Optimized GeoTIFF (COG), or group of COGs, in an ADE workspace can be visualized in a dynamic map by using a tiling service hosted in MAAP.

Steps: 1. Make a list of TIFFs in your workspace to use as a single layer 2. Generate a MosaicJSON file from this list of files (or a GeoJSON index) 3. Combine the MosaicJSON with other tiler visualization parameters to register a layer with your visualization tool.

[11]:
```
import glob
import ipycmc
import os
from pprint import pprint
import requests
import urllib

#!pip install cogeo_mosaic
from cogeo_mosaic.mosaic import MosaicJSON
from cogeo_mosaic.backends import MosaicBackend
```

Build a list of files

You can either make a list of file paths, or create a GeoJSON layer with a column containing the file paths. The paths need to be **S3** paths currently.

[12]:
```
# Local Path to your COGs
dps_output = "/projects/shared-buckets/alexdevseed/landsat8/viz/"

# titiler endpoint
titiler_endpoint = f"https://titiler.maap-project.org"

# Search for files to include, use recursive if nested folders (common in DPS output)
files = glob.glob(os.path.join(dps_output, "Landsat*_dps.tif"), recursive=False)

def local_to_s3(url):
    ''' A Function to convert local paths to s3 urls'''
    return url.replace("/projects/shared-buckets", "s3://maap-ops-workspace/shared")

tiles = [local_to_s3(file) for file in files]
print(tiles)
```

```
['s3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/Landsat8_30542_comp_cog_
↪2015-2020_dps.tif', 's3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/
↪Landsat8_30543_comp_cog_2015-2020_dps.tif', 's3://maap-ops-workspace/shared/
↪alexdevseed/landsat8/viz/Landsat8_30822_comp_cog_2015-2020_dps.tif', 's3://maap-ops-
↪workspace/shared/alexdevseed/landsat8/viz/Landsat8_30823_comp_cog_2015-2020_dps.tif
↪']
```

How to find the S3 path

You might be wondering how to find the S3 path to begin with. Right now the easiest way is to right click on a file in the file explorer on the left panel, and **Get Presigned S3 Url**.

It will look something like `https://maap-ops-workspace.s3.amazonaws.com/shared/alexdevseed/landsat8/viz/Copernicus_30438_covars_cog_topo_stack.tif?AWSAccessKeyId...`

The first part of the url is the bucket name: `maap-ops-workspace`. After the next `/`, it then matches to the local path.

Future versions of MAAP will include functions to do this part for you...

Make a mosaic

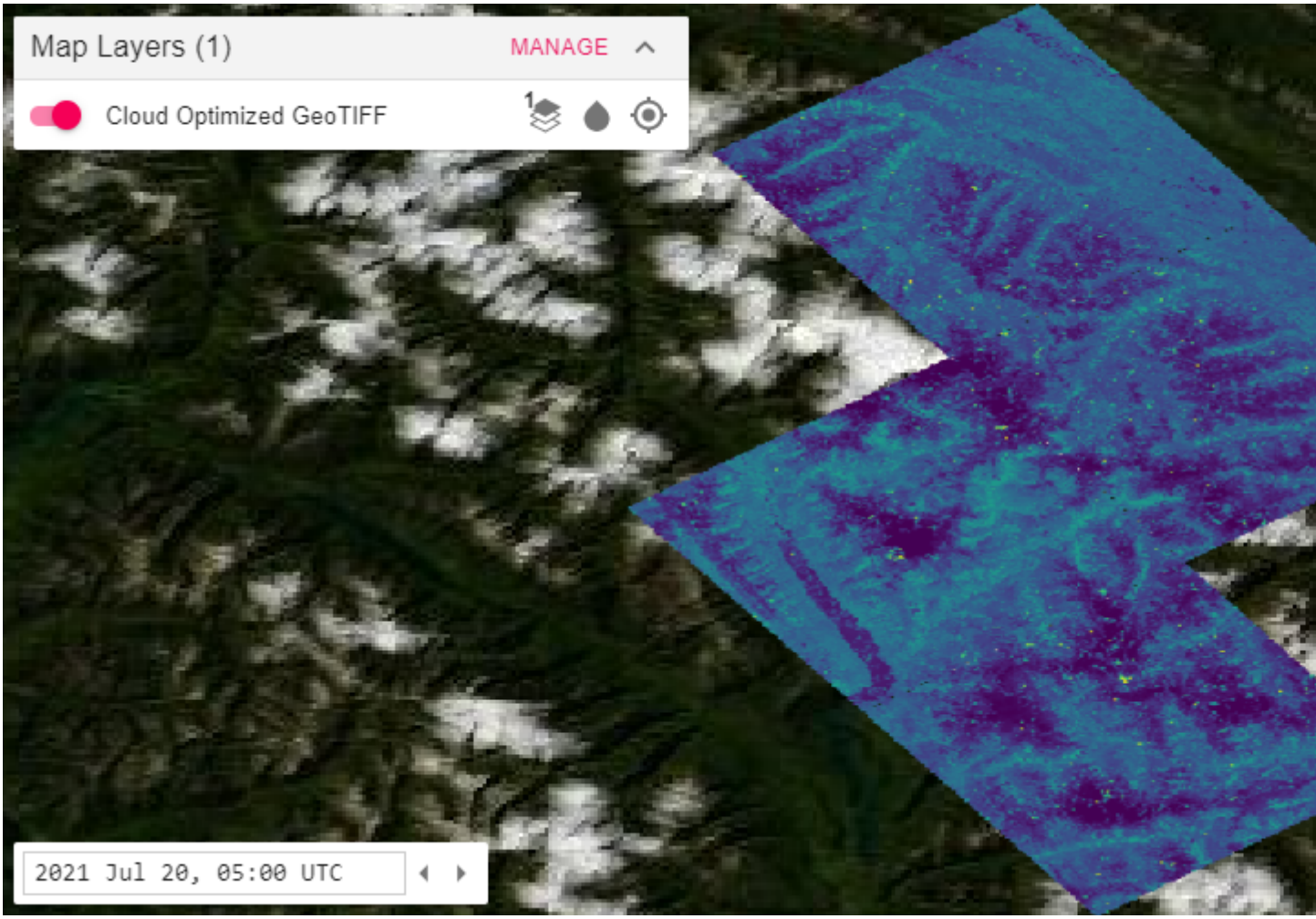
[3]:
```
# Now take the list of S3 paths and generate a mosaic
# TODO: if you have a lot of files (more than 100), creating a GeoJSON index and␣
↪using from_features will be more efficient.

mosaicdata = MosaicJSON.from_urls(tiles, minzoom=9, maxzoom=16)
```

[13]:
```
r = requests.post(
    url=f"{titiler_endpoint}/mosaics",
    headers={
        "Content-Type": "application/vnd.titiler.mosaicjson+json",
    },
    json=mosaicdata.dict(exclude_none=True)).json()

mosaicid = r['id']
```

Make a Map

[14]:
```
w = ipycmc.MapCMC()
w
```

```
MapCMC()
```



[18]:
```
# Build a WMTS call
"""
All of this is subject to change in a future version
The important parameters for users:
  url : the S3 path to the MosaicJSON file,
  bidx (band number),
  rescale (required if using non Byte data type),
  colormap_name or colormap

Other parameters are possible, see https://titiler.maap-project.org/docs#/MosaicJSON/
↪wmts_mosaicjson_WMTSCapabilities_xml_get
"""

wmts_url = f"https://titiler.maap-project.org/mosaics/{mosaicid}/WMTSCapabilities.xml"
params = {
    "tile_format":"png",
    "tile_scale":"1",
    "pixel_selection":"first",
    "TileMatrixSetId":"WebMercatorQuad",
    "bidx":"6", # Select which band to use
    "resampling_method":"nearest",
    "rescale":"0,1", # Values in data are from 0 to 1
    "return_mask":"true",
```

(continues on next page)

(continued from previous page)

```
    "colormap_name":"viridis" # Any colormap from matplotlib will work
}

wmts_call = "?".join([wmts_url, urllib.parse.urlencode(params)])

# Note Jupyter bug add amp; incorrectly when printing the url
wmts_call
```

[18]:
```
'https://titiler.maap-project.org/mosaics/d96e662d-fa69-4de1-a36c-4cf144c3e8e9/
↪WMTSCapabilities.xml?tile_format=png&tile_scale=1&pixel_selection=first&
↪TileMatrixSetId=WebMercatorQuad&bidx=6&resampling_method=nearest&rescale=0%2C1&
↪return_mask=true&colormap_name=viridis'
```

[16]:
```
# This adds a new layer to the map above, call Cloud Optimized GeoTIFF
w.load_layer_config(wmts_call, "wmts/xml")
```

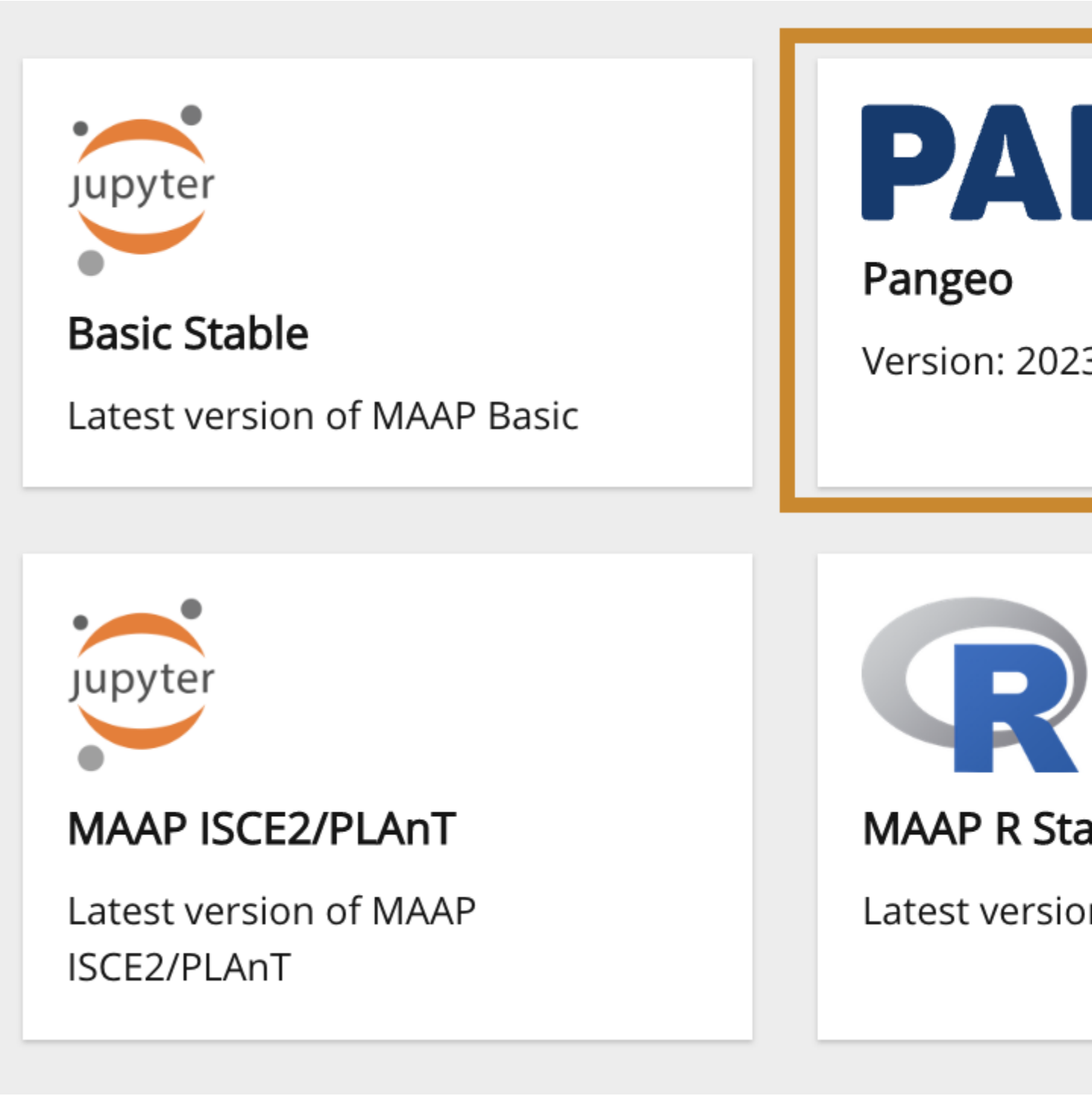
3.2.5 Visualizing STAC Data Layers using stac_ipyleaflet (beta)

This project - **stac_ipyleafet** - leverages the ipyleaflet mapping library and is intended to reduce the code needed to visualize data from the *MAAP STAC*. The beta version supports Cloud-Optimized GeoTIFFs (COGs).

Currently, this package works within this Algorithm Development Environment.

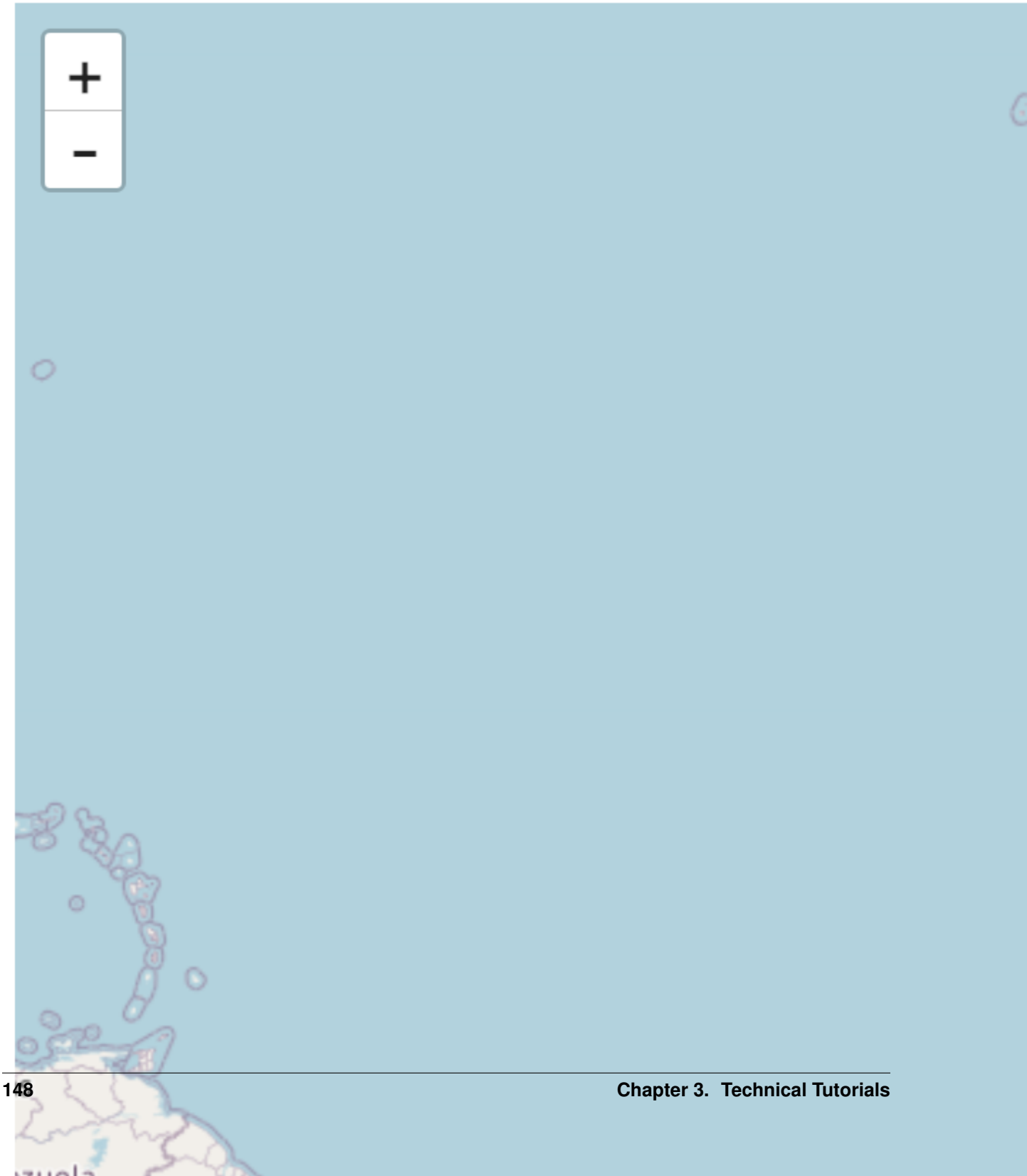
Package Import

1. Open a new **PANGEO workspace**



2. Create a **Notebook**
3. Within the first cell, **Import** the package and **create a map instance** - `m`

```
import stac_ipyleaflet
m = stac_ipyleaflet.StacIpyleaflet()
m
```



The **stac ipyleaflet** notebook's user interface consists of a map and a custom set of tools to aid in the discovery and visualization of STAC datasets, along with Biomass Layers and pre-determined Basemaps.

Map Navigation

To move the map

- press and hold a mouse-click, then drag the map

To adjust the map's scale (zoom extent) - use 1 of 4 methods:

- click the Zoom In / Out buttons in the top left-corner (this will maintain the center)
- use your mouse's scroll-wheel - hovering over an area of interest
- double-click within the map on an area of interest
- while pressing the `shift` key on your keyboard, press and hold a mouse-click, then drag to draw a rectangle around the area of interest

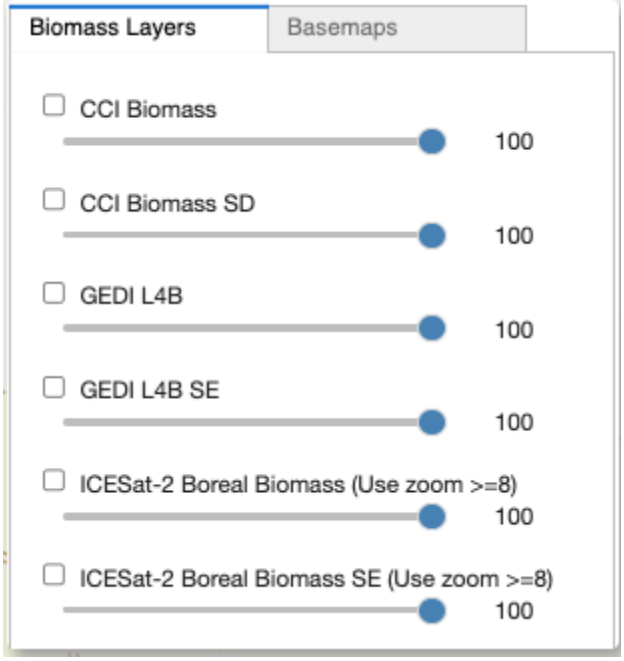
---

Layers Tool

Pressing the Layers button at the top opens the **Layers widget** that consists of 2 tabs. This tool currently allows users to: - View one or more Biomass Layers at the same time to see different combinations. - Choose between common **Basemap Layers** that are known favorites. - Have full control over the **opacity** of any layer or basemap for fine-tuning how the map looks.

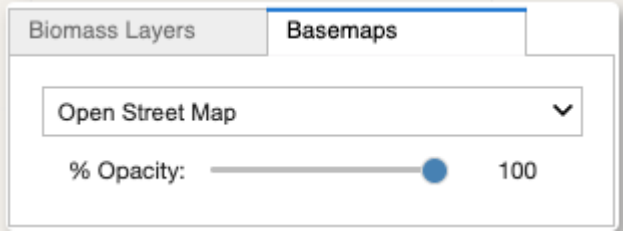
The Layers Tab

- Toggle each layer's visibility by using its checkbox
- Adjust each layer's opacity by moving its slider



The Basemaps Tab

- Select a basemap from the dropdown
- Adjust the basemap's opacity by moving its slider

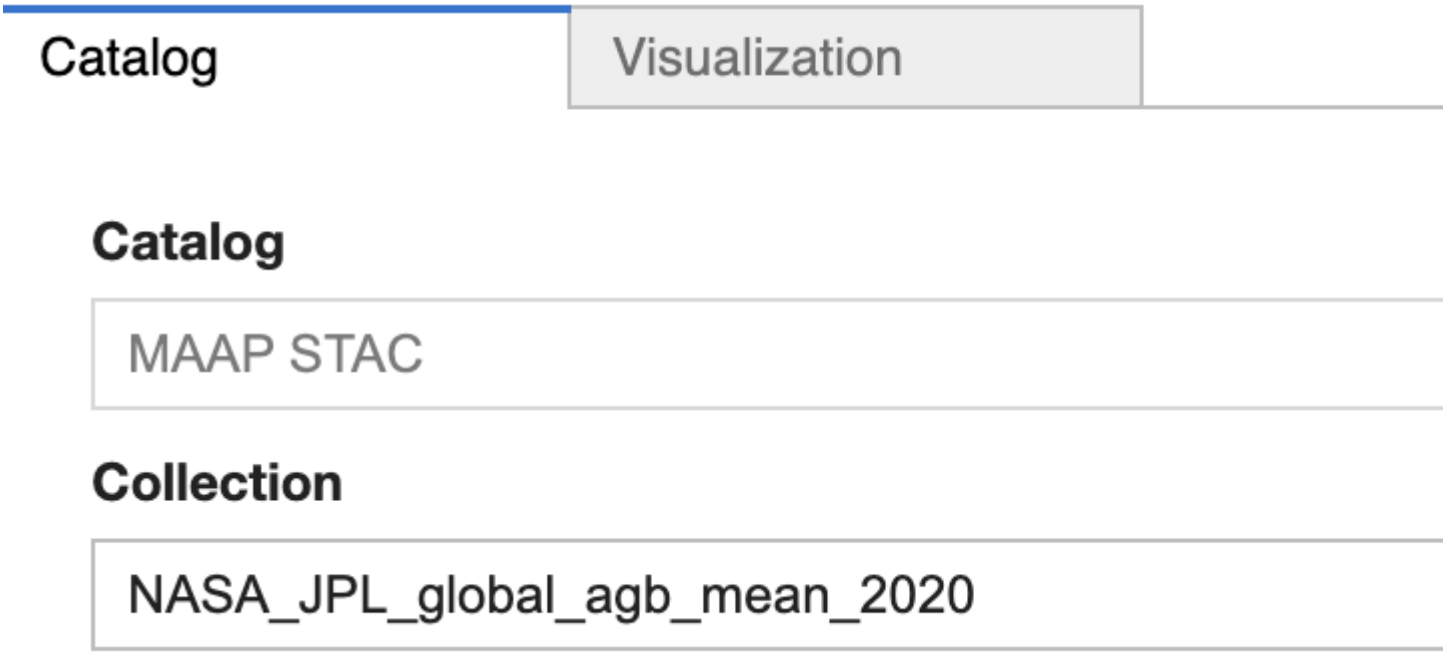


STAC Discovery Tool

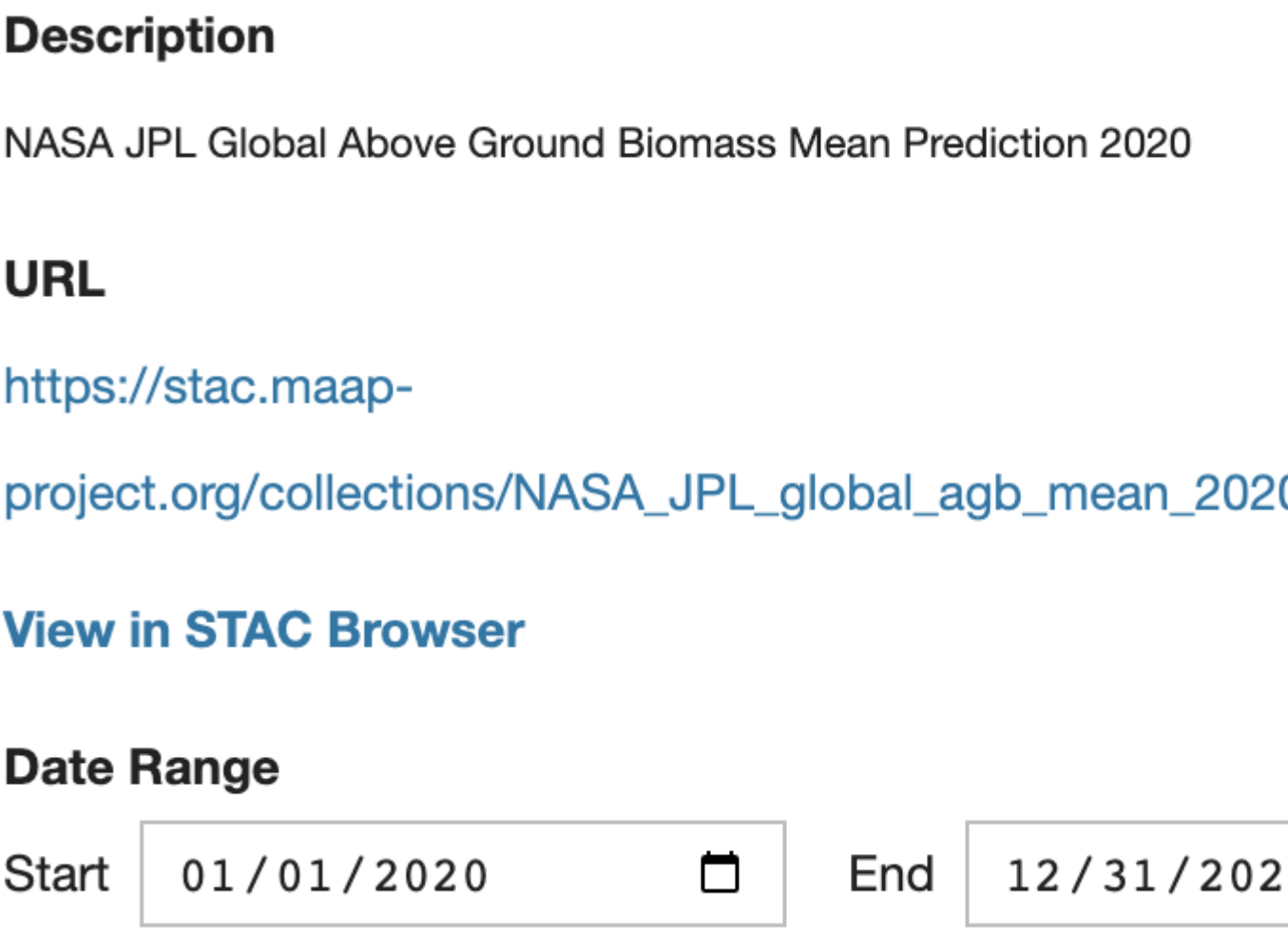
Pressing the STAC Data button at the top opens the **STAC widget** that consists of 2 tabs. This tool currently allows users to: - **Connect** to the MAAP STAC to access collections of mission data. - **Discover** items per the selected collection, including description, available dates, & direct URL. - **Identify** valid COG datasets. - **Add COG tiles** dynamically to the map. - **Customize** the tiles by changing the selected color palette for the selected item.

The Catalog Data Discovery Tab

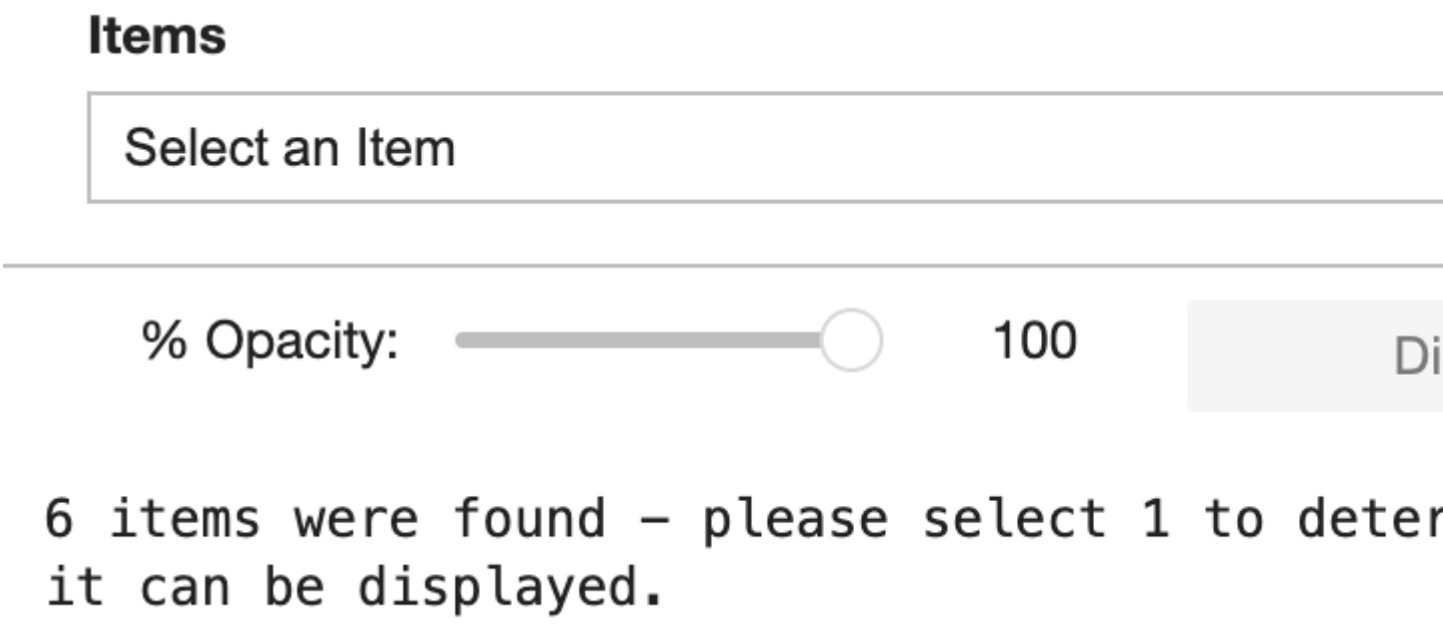
- Select a Collection within the default STAC library.



- Browse through the Collection's details.

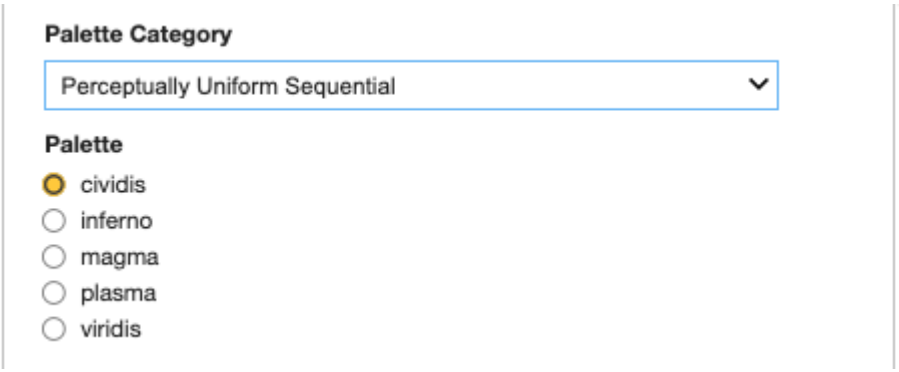


- Select an item from the collection to check if it is a valid COG. If it is, the Display button will become active (available) to add the selected item to the map. The displayed STAC layer's opacity can be adjusted by moving its slider.



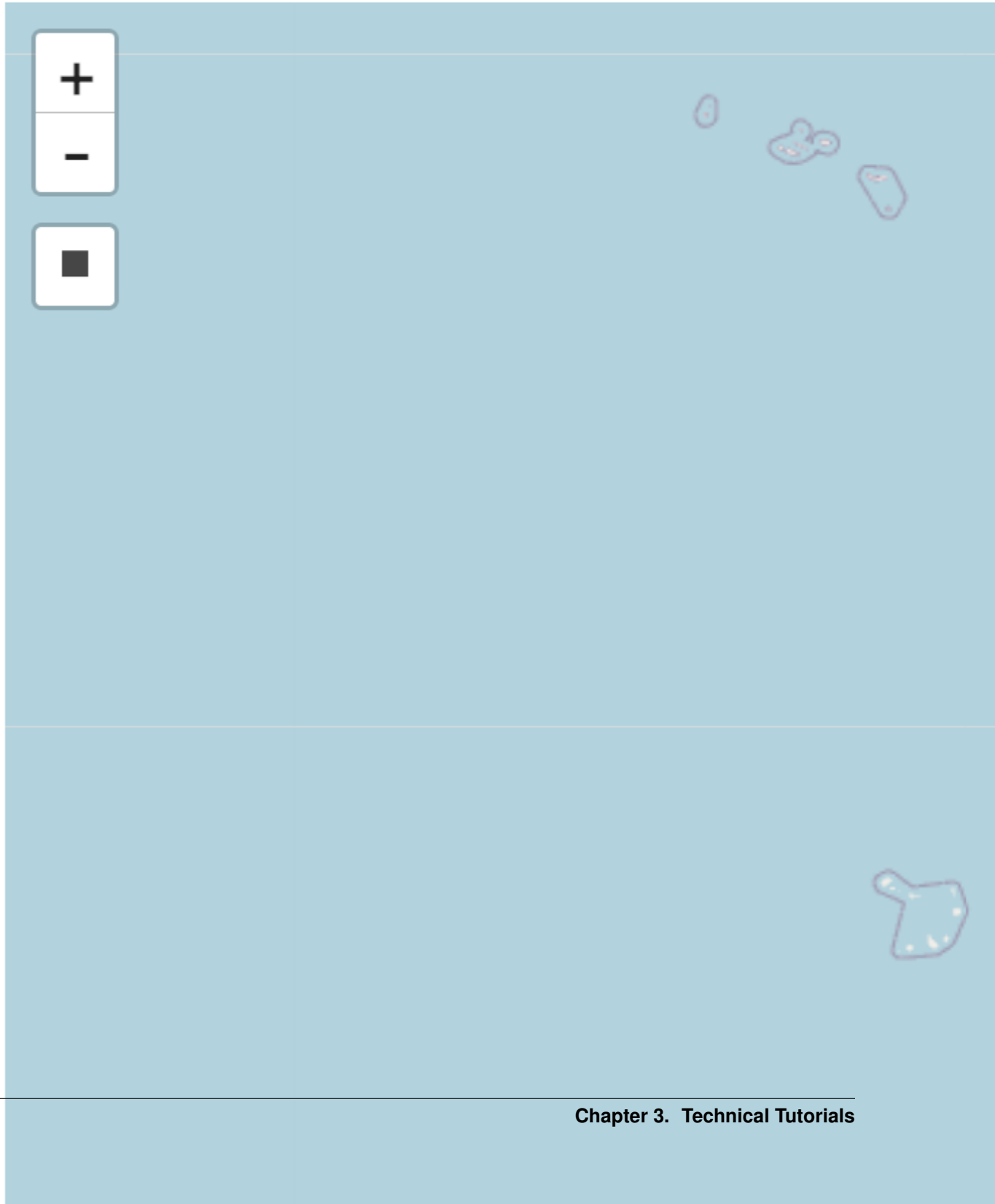
The Visualization Tab

- Select a category from the dropdown.
- Select an item from the corresponding color palettes.
- Press the Display button to update the data on the map.



Draw an Area of Interest & Get Coordinates

1. **Activate the Draw Tools** (click on the top Draw button).
2. **Use your mouse** to click, hold and draw a polygon over the map - releasing to finish.



3. **Note** the AOI's **Coordinates & BBox** within the open window.
4. **Print** the AOI's bbox from within a cell:

```
[3]: m.aoi_bbox

[3]: (-21.049905, 24.670176, -7.513855, 29.5
```

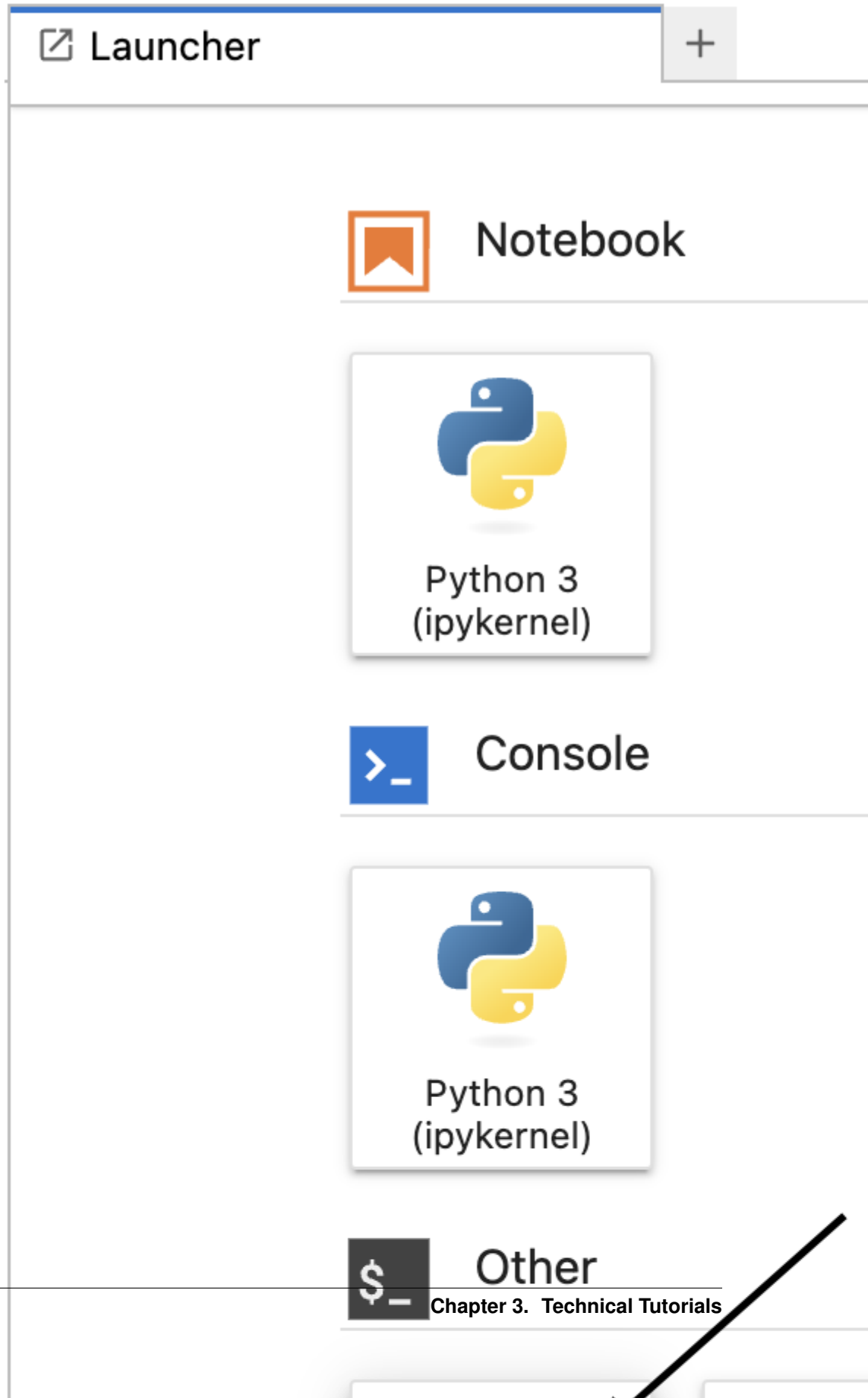
•

5. **Clear** the AOI polygon graphic & coordinates by clicking the **Clear AOI Polygon** Button.

Package Install (outside of a Pangeo)

NOTE: This has only been tested in the **MAAP Basic Stable** environment.

To install the `stac_ipyleaflet` package:



2. Copy, paste and RUN the following statement:

```
pip install git+https://github.com/MAAP-Project/stac_ipyleaflet.
git#egg-info=stac_ipyleaflet
```

A successful installation will end in the following:

```
Successfully built stac-ipyleaflet
Installing collected packages: stac-ipyleaflet
Successfully installed stac-ipyleaflet-0.2.0
```

3.3 Access

3.3.1 Accessing Data from the MAAP

Authors: Samuel Ayers (UAH), Brian Satorius (JPL)

Date: May 26, 2022 (Revised July 2023)

Description: In this example, we demonstrate how to access data from the MAAP using the 'getData' method of the maap-py library. At this time, this procedure is the same for user-contributed data added to the store.

Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

Additional Resources

- Search Granules Tutorial
- Additional Search Attributes

Importing Packages

We import the `os` module, import the `MAAP` package, and create a new MAAP class.

[1]:
```
# import os module
import os

# import the MAAP package
from maap.maap import MAAP

# create MAAP class
maap = MAAP()
```

For this example, the additional bounding_box attribute is used to search for granules within the Mondah Forest Gabon research site spatial area. For more information about searching for granules in MAAP, please see https://docs.maap-project.org/en/latest/technical_tutorials/search/granules.html.

[2]:
```
SHORTNAME = "AFLVIS2"
BBOX = '9.316216,0.538705,9.422509,0.616939'

# search for granules with SHORTNAME
results = maap.searchGranule(
    short_name=SHORTNAME,
    bounding_box=BBOX
)
results[0]
```

[2]:

```
{'concept-id': 'G1549416185-NSIDC_ECS',
 'collection-concept-id': 'C1549378743-NSIDC_ECS',
 'revision-id': '2',
 'format': 'application/echo10+xml',
 'Granule': {'GranuleUR': 'SC:AFLVIS2.001:138348928',
  'InsertTime': '2018-09-24T11:01:33.892Z',
  'LastUpdate': '2018-09-24T11:02:09.869Z',
  'Collection': {'DataSetId': 'AfriSAR LVIS L2 Geolocated Surface Elevation Product␣
↪V001'},
  'DataGranule': {'SizeMBDataGranule': '130.082',
   'ProducerGranuleId': 'LVIS2_Gabon2016_0304_R1808_054746.TXT',
   'DayNightFlag': 'UNSPECIFIED',
   'ProductionDateTime': '2018-08-23T18:32:04.000Z',
   'LocalVersionId': '001'},
  'Temporal': {'RangeDateTime': {'BeginningDateTime': '2016-03-04T15:12:26.391000Z',
    'EndingDateTime': '2016-03-04T15:19:18.916000Z'}},
  'Spatial': {'HorizontalSpatialDomain': {'Geometry': {'GPolygon': {'Boundary': {
↪'Point': [{'PointLongitude': '9.27281',
         'PointLatitude': '0.57138'},
        {'PointLongitude': '9.33743', 'PointLatitude': '0.57138'},
        {'PointLongitude': '9.42716', 'PointLatitude': '0.57138'},
        {'PointLongitude': '9.5169', 'PointLatitude': '0.57138'},
        {'PointLongitude': '9.60664', 'PointLatitude': '0.57138'},
        {'PointLongitude': '9.66048', 'PointLatitude': '0.56777'},
        {'PointLongitude': '9.73945', 'PointLatitude': '0.56416'},
        {'PointLongitude': '9.73586', 'PointLatitude': '0.54613'},
        {'PointLongitude': '9.64972', 'PointLatitude': '0.54974'},
        {'PointLongitude': '9.55998', 'PointLatitude': '0.54974'},
        {'PointLongitude': '9.47024', 'PointLatitude': '0.54974'},
        {'PointLongitude': '9.3805', 'PointLatitude': '0.54974'},
        {'PointLongitude': '9.32307', 'PointLatitude': '0.54613'},
        {'PointLongitude': '9.2764', 'PointLatitude': '0.54974'},
        {'PointLongitude': '9.27281', 'PointLatitude': '0.56416'},
        {'PointLongitude': '9.27281', 'PointLatitude': '0.56777'}]}}}}},
  'Platforms': {'Platform': {'ShortName': 'B-200',
    'Instruments': {'Instrument': {'ShortName': 'LVIS',
      'Sensors': {'Sensor': {'ShortName': 'LVIS'}}}}}},
  'Campaigns': {'Campaign': {'ShortName': 'AfriSAR'}},
  'AdditionalAttributes': {'AdditionalAttribute': [{'Name': 'SIPSMetGenVersion',
     'Values': {'Value': '2.5'}},
    {'Name': 'ThemeID', 'Values': {'Value': 'AfriSAR'}},
    {'Name': 'AircraftID', 'Values': {'Value': 'N529NA'}}]},
  'OnlineAccessURLs': {'OnlineAccessURL': [{'URL': 'https://n5eil01u.ecs.nsidc.org/
↪DP8/ICEBRIDGE/AFLVIS2.001/2016.03.04/LVIS2_Gabon2016_0304_R1808_054746.TXT',
     'MimeType': 'application/octet-stream'}]},
  'OnlineResources': {'OnlineResource': {'URL': 'https://n5eil01u.ecs.nsidc.org/DP8/
↪ICEBRIDGE/AFLVIS2.001/2016.03.04/LVIS2_Gabon2016_0304_R1808_054746.TXT.xml',
    'Type': 'METADATA',
```


(continues on next page)

(continued from previous page)

```
    'MimeType': 'text/xml'}},
  'Orderable': 'true',
  'Visible': 'true'}}
```

We assign a variable (in this case, `data_file`) to the first result of our search from the cell above.

A data directory is then set, and if the directory does not already exist, it is created. The file from our search is then downloaded into the file system in this directory. Here, the function getData() is downloading the data using the access URLs in the CMR granule and downloading it directly to the path provided.

[3]:
```
# grab first result
data_file = results[0]

# set data directory
dataDir = './data'

# check if directory exists -> if directory doesn't exist, directory is created
if not os.path.exists(dataDir):
    os.mkdir(dataDir)

# extract the link to the resource
data = data_file.getData(dataDir)
```

We can now see that the data directory has been created and the data file is downloaded into the directory. The downloaded file remains in the data directory until the user deletes it.

Accessing Data from AWS Requester Pays Buckets

Some data is cloud available but in requester pays buckets. In this example, we use Rasterio, Boto3, and MAAP's `aws.requester_pays_credentials()` function to retrieve data within the `usgs-landsat` requester pays bucket.

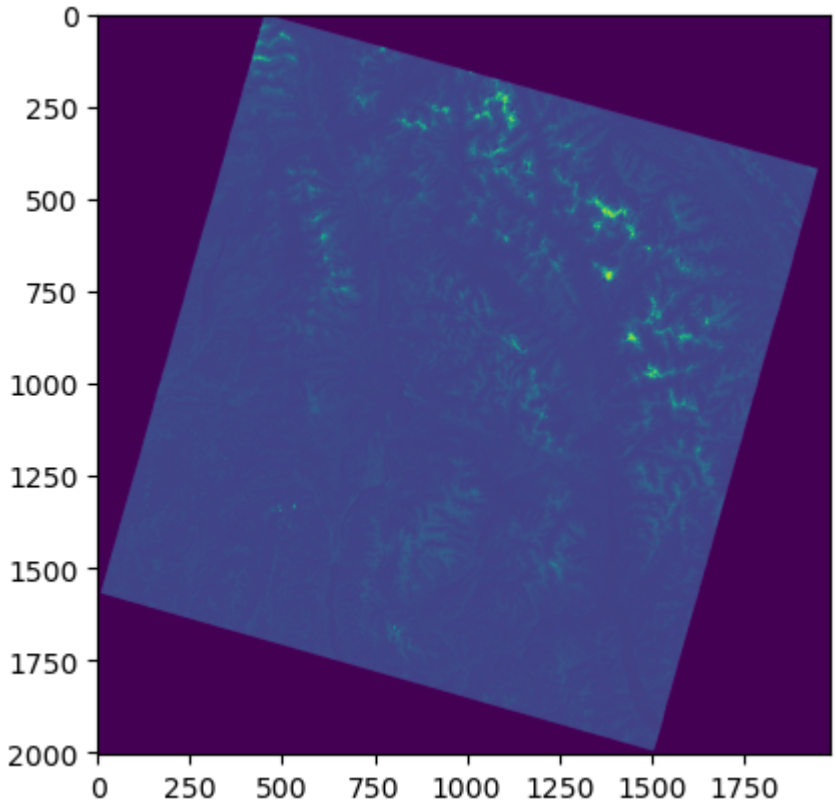
[4]:
```
import boto3
import rasterio as rio

from maap.maap import MAAP
from rasterio.plot import show
from rasterio.session import AWSSession


maap = MAAP(maap_host='api.maap-project.org')
credentials = maap.aws.requester_pays_credentials()

boto3_session = boto3.Session(
    aws_access_key_id=credentials['aws_access_key_id'],
    aws_secret_access_key=credentials['aws_secret_access_key'],
    aws_session_token=credentials['aws_session_token']
)
```

[5]:
```
aws_session = AWSSession(boto3_session, requester_pays=True)
file_s3 = 's3://usgs-landsat/collection02/level-2/standard/oli-tirs/2015/044/025/LC08_
↪L2SP_044025_20150812_20200908_02_T1/LC08_L2SP_044025_20150812_20200908_02_T1_SR_B2.
↪TIF'
with rio.Env(aws_session):
    with rio.open(file_s3, 'r') as src:
```

(continues on next page)

(continued from previous page)

```
        # list of overviews
        oviews = src.overviews(1)
        # get second item from list to retrieve a thumbnail
        oview = oviews[1]
        # read first band of file and set shape of new output array
        thumbnail = src.read(1, out_shape=(1, int(src.height // oview), int(src.width␣
↪// oview)))
# now display the thumbnail
show(thumbnail)
```



[5]:
```
<Axes: >
```

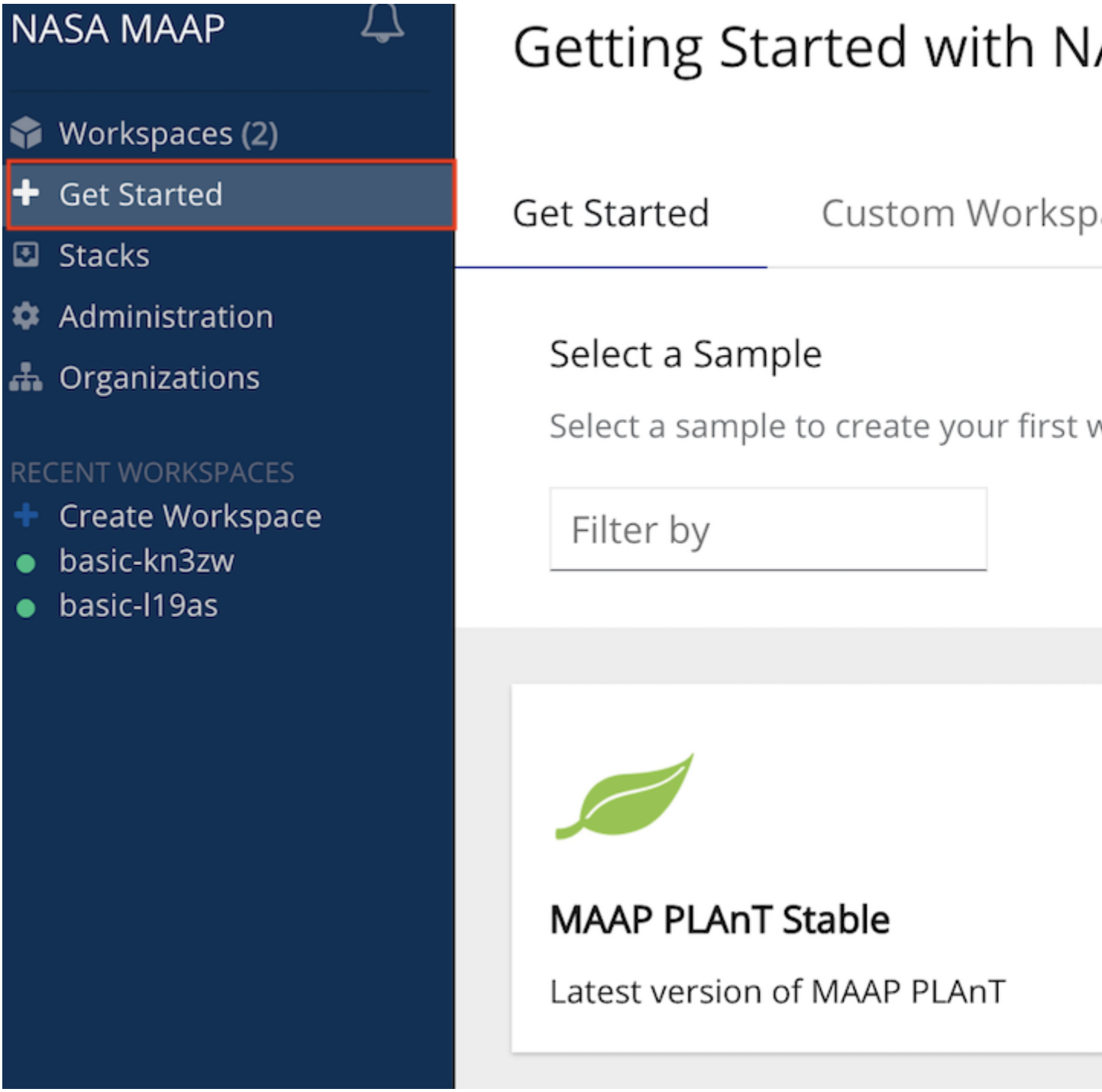
You may adjust the expiration time of the AWS credentials generated by `maap.aws.requester_pays_credentials()`:

[ ]:
```
# Credential expiration time in seconds (defaults to 12 hours)
maap.aws.requester_pays_credentials(expiration=3600)
```

3.3.2 Accessing data provided by NASA's Distributed Active Archive Centers (DAACs)

It is possible to download data provided by DAACs, including data which is not cataloged by the MAAP's CMR, using the NASA MAAP ADE. This data is hosted externally from the MAAP but can be accessed using the NASA MAAP ADE's authentication systems.

In order to do this, we start by creating a Jupyter workspace within the NASA MAAP ADE. Using the left-hand navigation, select "+ Get Started" and then select the "Jupyter - MAAP Basic Stable" workspace.



Alternatively, you can create a workspace using the "Workspaces" interface. See Create Workspace for more information.

Accessing data from Jupyter Notebooks in your workspace

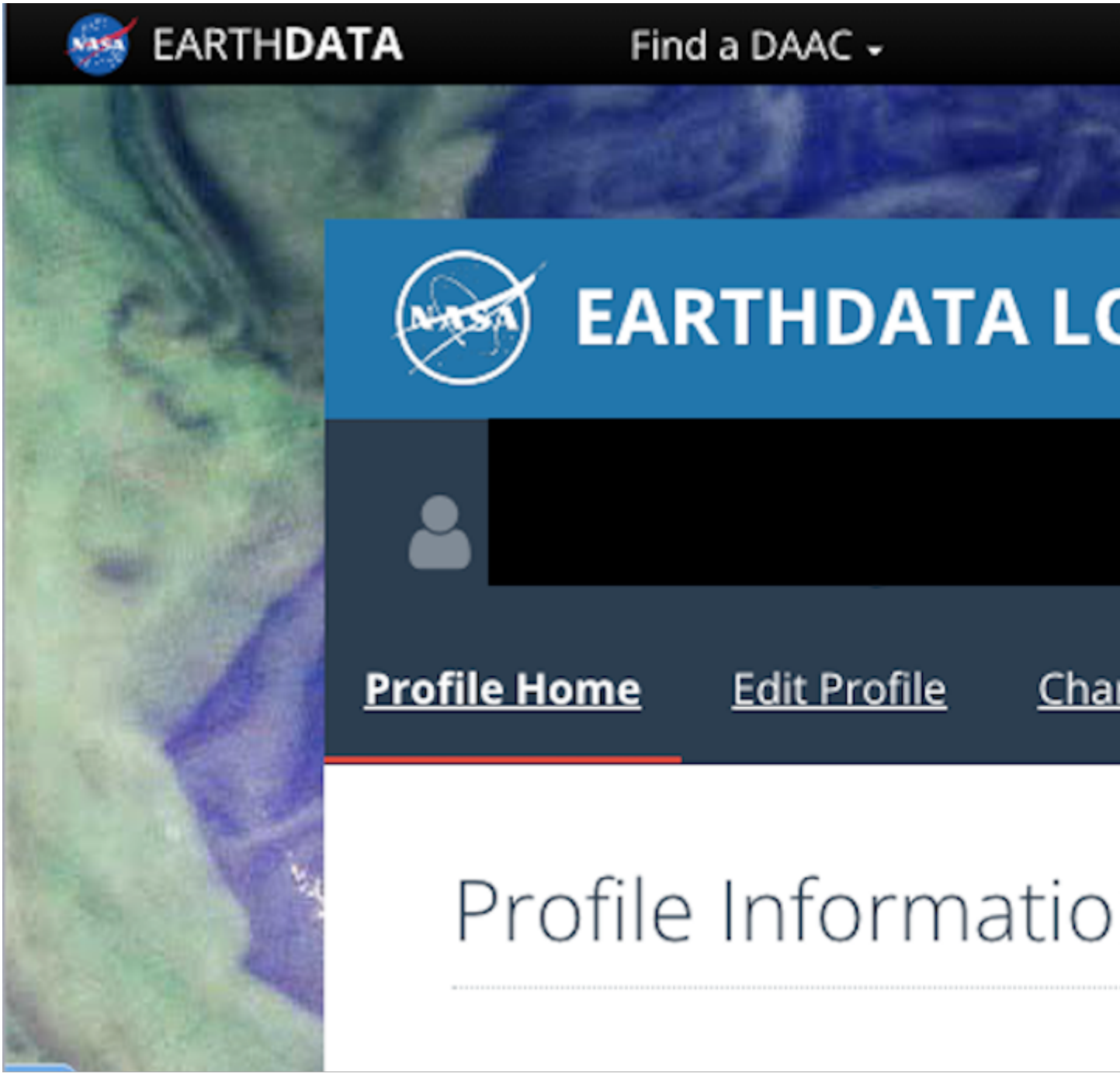
Within your Jupyter Notebook, start by importing the `maap` package. Then invoke the `MAAP` constructor, setting the `maap_host` argument to `'api.ops.maap-project.org'`.

[1]:
```
# import the maap package
from maap.maap import MAAP
# invoke the MAAP constructor using the maap_host argument
maap = MAAP(maap_host='api.ops.maap-project.org')
```

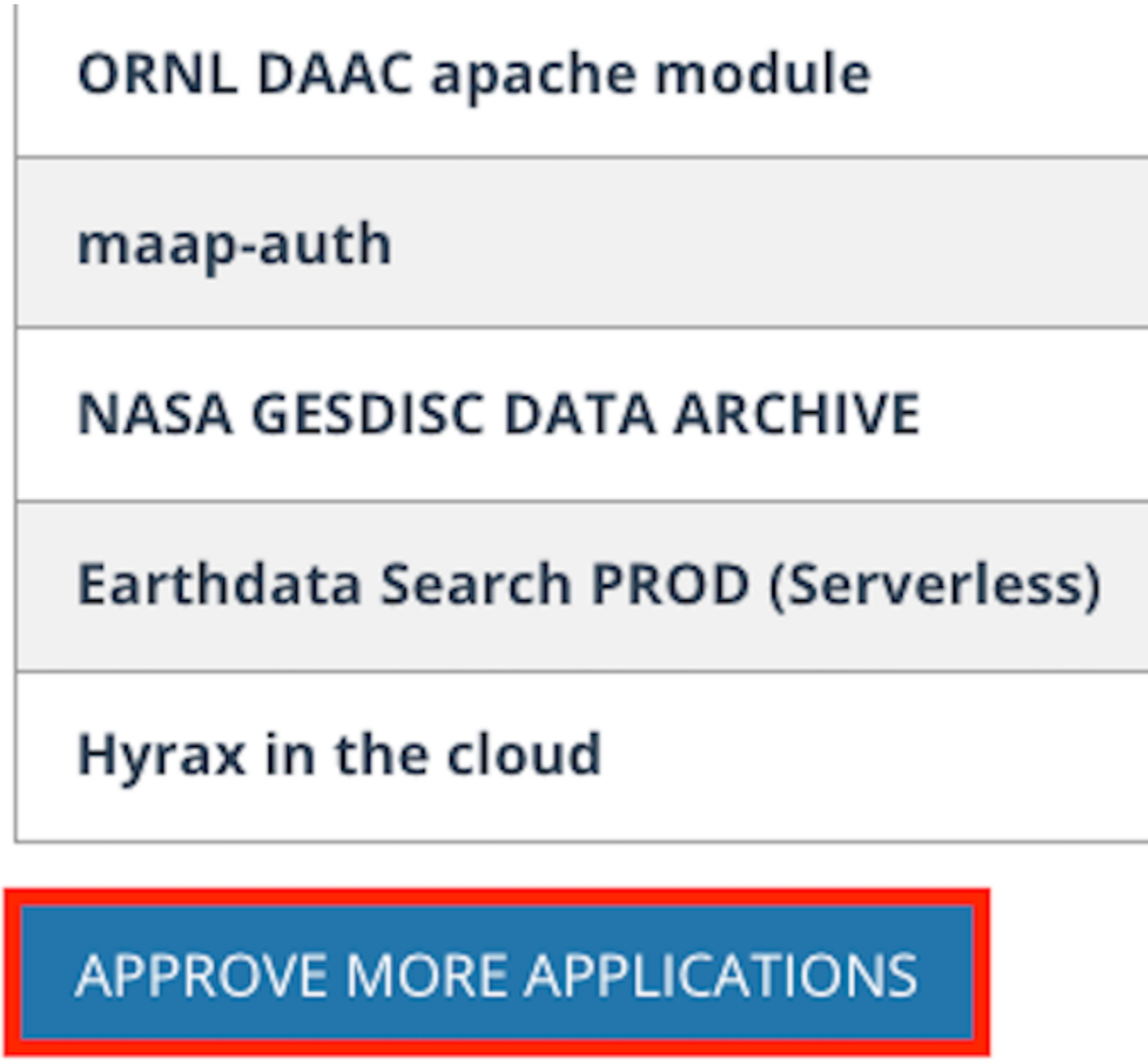
Granting Earthdata Login access to your target DAAC application

In order to access external DAAC data from the NASA MAAP ADE, MAAP uses your Earthdata Login profile to send a data request to the desired DAAC application.

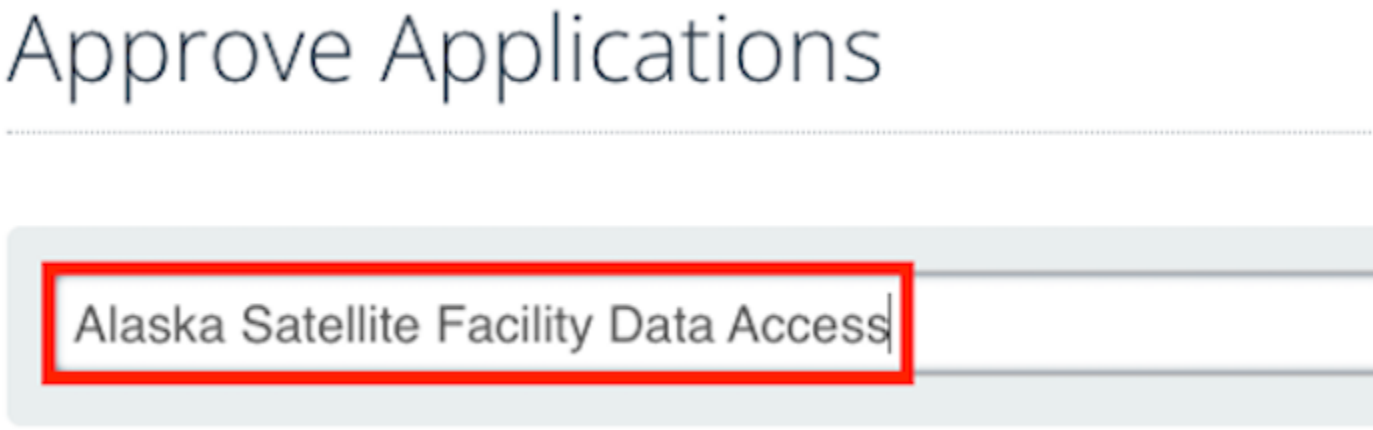
Some DAAC applications (such as 'Alaska Satellite Facility Data Access') must be authorized before you can use them. Login or register at https://urs.earthdata.nasa.gov/ in order to see the applications that you have authorized. From the profile page, click on the 'Applications' tab and select 'Authorized Apps' from the drop-down menu.



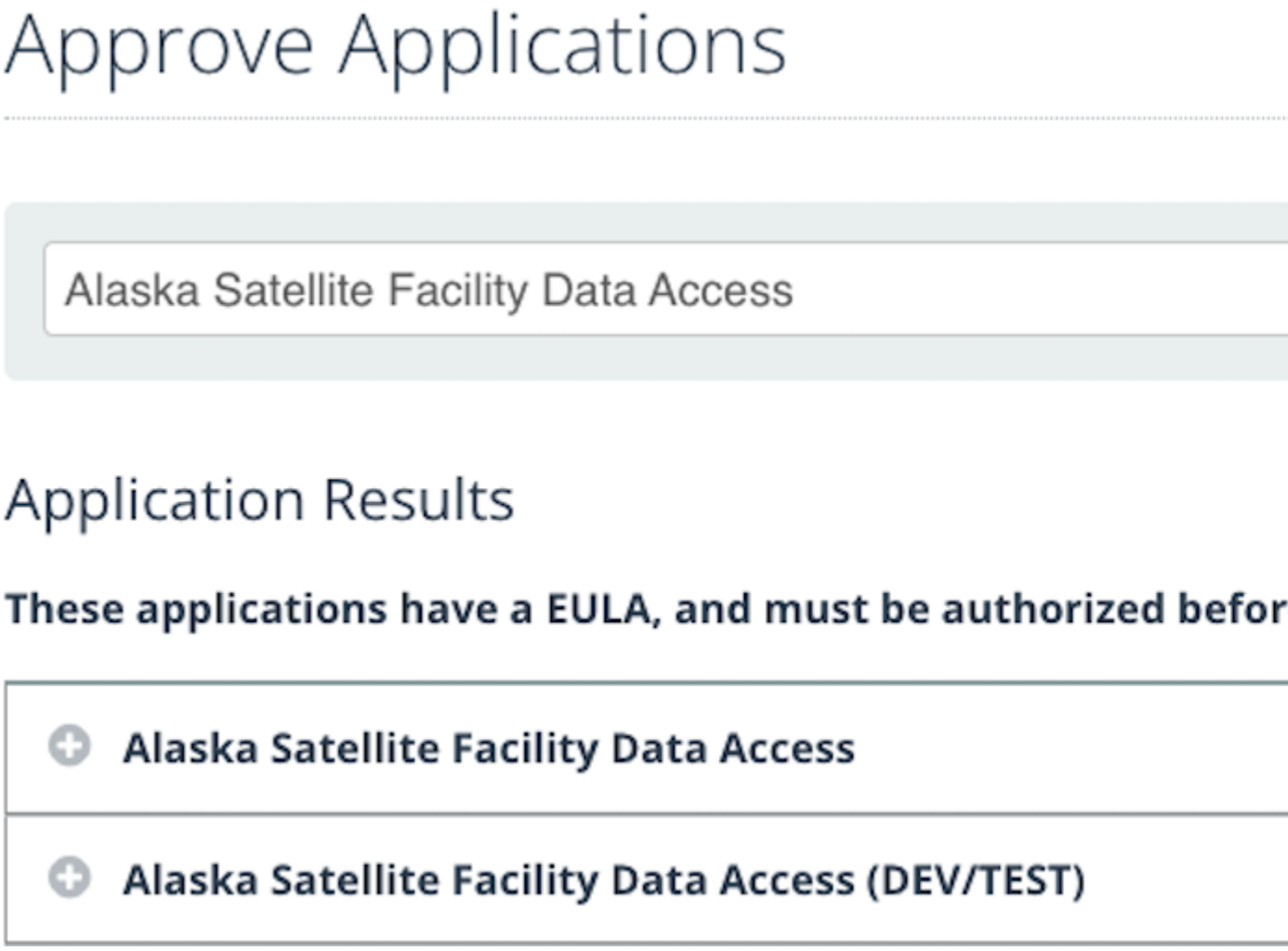
This takes you to the Approved Applications page which lists the applications you have authorized. To add more applications, scroll down to the bottom of the page and click the 'APPROVE MORE APPLICATIONS' button which takes you to the Application search page.



Enter the desired application name within the search box and click the 'SEARCH' button. After this, a list of search results appears.

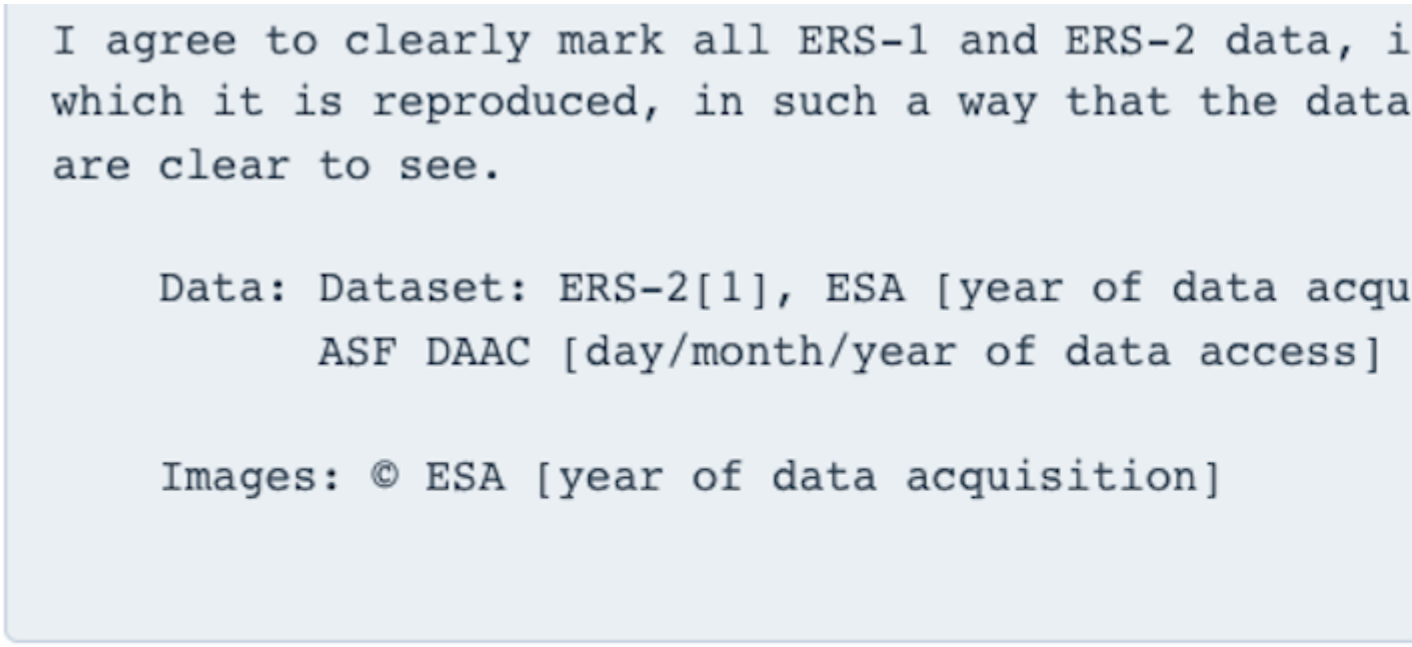


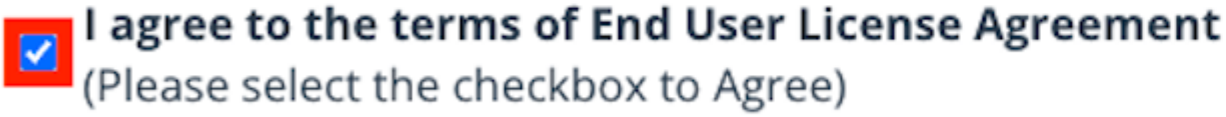
Once you find the desired application, click the 'AUTHORIZE' button next to the name.



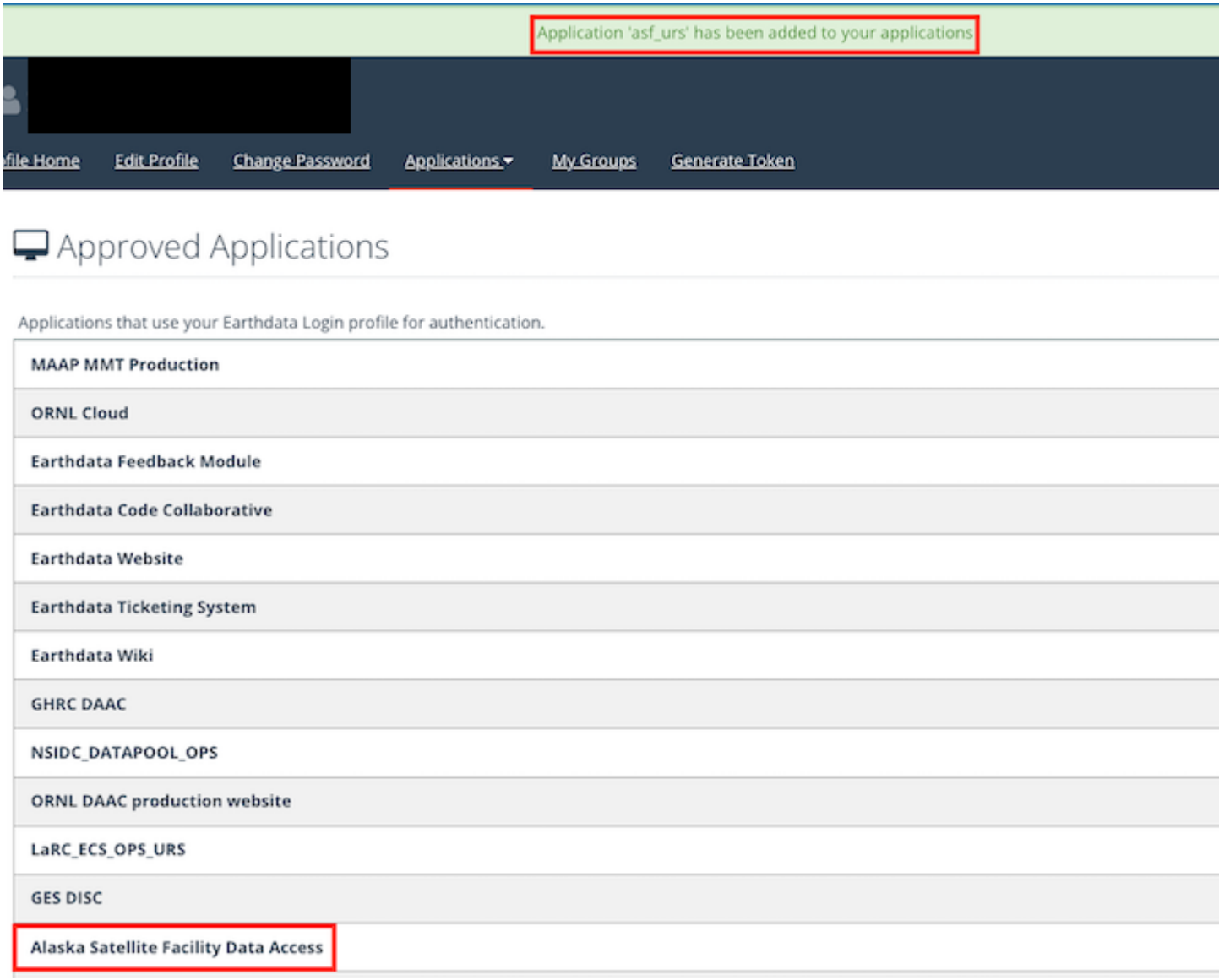
You are then presented with its End User License Agreement. In order to have authorization, you need to select the 'I

agree to the terms of End User License Agreement' checkbox and then click the 'AGREE' button.





After this is done, you are then shown the Approved Applications page again and the desired application should now be listed.



Note that if Earthdata Login access is not granted to your target DAAC application, the following example will result in a 401-permission error.

Accessing Sentinel-1 Granule Data from the Alaska Satellite Facility (ASF)

Search for a granule using the `searchGranule` function (for more information on searching for granules, see Searching for Granules in MAAP). Then utilize the `getData` function, which downloads granule data if it doesn't already exist locally. We can use `getData` to download the first result from our granule search into the file system and assign it to a variable (in this case `download`). Note that you will need to authorize the 'Alaska Satellite Facility Data Access' application before downloading any results from our search (see the above section for more information concerning authorizing applications).

[2]:
```
# search for granule data using the short_name argument
results = maap.searchGranule(short_name='SENTINEL-1A_DP_GRD_HIGH')
# download first result
download = results[0].getData()
```

Note that we can then use the `print` function to see the file name and directory.

[3]:
```
# print file directory
print(download)
```

```
./S1A_S3_GRDH_1SDH_20140615T034444_20140615T034512_001055_00107C_8977.zip
```

Accessing Harmonized Landsat Sentinel-2 (HLS) Level 3 Granule Data from the Land Processes Distributed Active Archive Center (LP DAAC)

We use a similar approach in order to access HLS Level 3 granule data. Note that this data is not cataloged by the MAAP's CMR but we can use `searchGranule`'s `cmr_host` argument to specify a CMR instance external to MAAP.

[4]:
```
# search for granule data using CMR host name and short name arguments
results = maap.searchGranule(
    cmr_host='cmr.earthdata.nasa.gov',
    short_name='HLSL30')
# download first result
download = results[0].getData()
```

As in the previous example, we can use the `print` function to see the file name and directory.

[5]:
```
# print file directory
print(download)
```

```
./HLS.L30.T59WPT.2013101T001445.v2.0.B09.tif
```

3.3.3 Accessing Global Ecosystem Dynamics Investigation (GEDI) Level 3 Granule Data

In this example, we demonstrate how to access GEDI Level 3 granule data on the MAAP ADE.

Within your Jupyter Notebook, start by importing the `maap` package. Then invoke the `MAAP` constructor, setting the `maap_host` argument to `'api.maap-project.org'`.

[8]:
```
# import os module
import os
# import the maap package to handle queries
from maap.maap import MAAP
# invoke the MAAP constructor using the maap_host argument
maap = MAAP(maap_host='api.maap-project.org')
```

Search for a granule using the `searchGranule` function (for more information on searching for granules, see Searching for Granules in MAAP). Note that we can use `searchGranule`'s `cmr_host` argument to specify `cmr.earthdata.nasa.gov` as the CMR instance.

[11]:
```
# search for granule data using CMR host name and collection concept ID arguments
results = maap.searchGranule(
    cmr_host='cmr.earthdata.nasa.gov',
    collection_concept_id='C2153683336-ORNL_CLOUD'
)
```

Let's view the list of `GranuleURs` within our `results`:

[12]:
```
# show all of the items
[item['Granule']['GranuleUR'] for item in results]
```

[12]:
```
['GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2020287_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2020287_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2021104_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2020287_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2021104_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2021104_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2021216_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2021216_002_02.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2021216_002_02.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_mean_2019108_2022019_002_03.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_counts_2019108_2022019_002_03.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_mean_2019108_2022019_002_03.tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2022019_002_03.
↪tif',
 'GEDI_L3_LandSurface_Metrics_V2.GEDI03_rh100_stddev_2019108_2022019_002_03.tif']
```

For this example, we are interested in downloading `GEDI03_elev_lowestmode_stddev_2019108_2021104_002_02.tif`.

[13]:
```
# select item
results[5]['Granule']['GranuleUR']
```

[13]:
```
'GEDI_L3_LandSurface_Metrics_V2.GEDI03_elev_lowestmode_stddev_2019108_2021104_002_02.
↪tif'
```

Now utilize the `getData` function, which downloads granule data if it doesn't already exist locally. We can use `getData` to download the third result from our granule search into the file system and assign its local path to a variable (in this case `download`).

[16]:
```
# download granule item
local_dir = '/projects/local_data'  # download directory (absolute path or relative␣
↪to current directory)
os.makedirs(local_dir, exist_ok=True) # create directories, as necessary
download = results[5].getData(local_dir)  # default download directory is current␣
↪directory, if no directory is given
```

We can then use the `print` function to see the file name and directory.

[17]:
```
# print path to downloaded file
print(download)
```

```
/projects/local_data/GEDI03_elev_lowestmode_stddev_2019108_2021104_002_02.tif
```

3.3.4 Accessing Global Ecosystem Dynamics Investigation (GEDI) Level 4A Granule Data

In this example, we demonstrate how to access GEDI Level 4A granule data on the MAAP ADE.

Within your Jupyter Notebook, start by importing the `maap` package. Then invoke the `MAAP` constructor, setting the `maap_host` argument to `'api.maap-project.org'`.

[7]:
```
# import os module
import os
# import the maap package to handle queries
from maap.maap import MAAP
# invoke the MAAP constructor using the maap_host argument
maap = MAAP(maap_host='api.maap-project.org')
```

Search for a granule using the `searchGranule` function (for more information on searching for granules, see Searching for Granules in MAAP). Note that we can use `searchGranule`'s `cmr_host` argument to specify `cmr.earthdata.nasa.gov` as the CMR instance. Then utilize the `getData` function, which downloads granule data if it doesn't already exist locally. We can use `getData` to download the first result from our granule search into the file system and assign its local path to a variable (in this case `download`).

[8]:
```
# search for granule data using CMR host name, collection concept ID, and Granule UR␣
↪arguments
results = maap.searchGranule(
    cmr_host='cmr.earthdata.nasa.gov',
    collection_concept_id='C2237824918-ORNL_CLOUD',
    granule_ur='GEDI_L4A_AGB_Density_V2_1.GEDI04_A_2019107224731_O01958_01_T02638_02_
↪002_02_V002.h5')
# download first result
local_dir = '/projects/local_data'  # Download directory (absolute path or relative␣
↪to current directory)
os.makedirs(local_dir, exist_ok=True) # Create directories, as necessary
download = results[0].getData(local_dir)  # Default download directory is current␣
↪directory, if no directory is given
```

We can then use the `print` function to see the file name and directory.

[9]:
```
# print path to downloaded file
print(download)
```

```
/projects/local_data/GEDI04_A_2019107224731_O01958_01_T02638_02_002_02_V002.h5
```

3.3.5 Accessing Cloud Optimized Data

Authors: Samuel Ayers (UAH)

Date: April 28, 2021 (Revised July 2023)

Description: The following is an example that uses Shuttle Radar Topography Mission (SRTM) Cloud Optimized GeoTIFF (COG) data from the MAAP data store, via MAAP STAC search. In this example, we read in elevation data using a bounding box tile.

3.3.6 Run This Notebook

To access and run this tutorial within MAAP's Algorithm Development Environment (ADE), please refer to the "Getting started with the MAAP" section of our documentation.

Disclaimer: it is highly recommended to run a tutorial within MAAP's ADE, which already includes packages specific to MAAP, such as maap-py. Running the tutorial outside of the MAAP ADE may lead to errors.

3.3.7 Additional Resources

- Using pystac-client

3.3.8 Importing and Installing Packages

To be able to run this notebook you'll need the following requirements:

- rasterio
- folium
- geopandas
- rio-cogeo

If the packages below are not installed already, uncomment the following cell

[1]:
```
# %pip install -U folium geopandas rasterio>=1.2.3 rio-cogeo
# %pip install pystac-client
```

[2]:
```
# import the maap package to handle queries
from maap.maap import MAAP

# invoke the MAAP
maap = MAAP()
```

We can use pystac_client `get_collection` to retrieve the desired collection, in this case the SRTMGL1_COD collection and set the result of the function to a variable.

[3]:
```
# Get the SRTMGL1_COD collection
from pystac_client import Client
catalog = 'https://stac.maap-project.org/'
client = Client.open(catalog)

collection = client.get_collection('SRTMGL1_COD')
collection
```

[3]:
```
<CollectionClient id=SRTMGL1_COD>
```

We can narrow the area of interest for the items within the collection by passing bounding box values into the pystac_client `search` function.

[4]:
```
# set bounding box
bbox = "-101.5,45.5,-100.5,46.5"

# retreive STAC items from the collection that are within the bounding box
search = client.search(bbox=bbox, collections=collection, max_items=20)
```

Let's check how many items we are working with.

[5]:
```
# show number of items in search results
len(list(search.items()))
```

[5]:
```
4
```

Inspecting the Results

Now we can work on inspecting our results. In order to do this, we import the `geopandas`,`shapely`, and `folium` packages.

[6]:
```
# import geopandas to work with geospatial data
import geopandas as gpd

# import shapely for manipulation and analysis of geometric objects
from shapely.geometry import shape, box

# import folium to visualize data in an interactive leaflet map
import folium
```

We can use the `gpd.GeoSeries` function to create a GeoSeries of all the shapely polygons created from the geometries of our item results. According to the GeoPandas documentation, a GeoSeries is a "Series [a type of one-dimensional array] object designed to store shapely geometry objects". We use 'EPSG:4326' (WGS 84) for the coordinate reference system then we can check the GeoSeries.

[7]:
```
# create GeoSeries of all polygons from granule results with WGS 84 coordinate␣
↪reference system
geometries = gpd.GeoSeries(
    [shape(item.geometry) for item in search.items()],
    crs='EPSG:4326'
)

# check GeoSeries
geometries
```

[7]:
```
0    POLYGON ((-102.00014 45.99986, -100.99986 45.9...
1    POLYGON ((-101.00014 45.99986, -99.99986 45.99...
2    POLYGON ((-102.00014 44.99986, -100.99986 44.9...
3    POLYGON ((-101.00014 44.99986, -99.99986 44.99...
dtype: geometry
```

Now we create a list from our bounding box values. Then we use the `centroid` function to get the centroid of our bounding box and set to a point. Next we use `folium.Map` to create a map. For the map's parameters, let's set the centroid point coordinates as the location, "cartodbpositron" for the map tileset, and 7 as the starting zoom level. With our map created, we can create a dictionary containing style information for our bounding box. Then we can use `folium.GeoJson` to create GeoJsons from our `geometries` GeoSeries and add them to the map. We also use the `folium.GeoJson` function to create a GeoJson of a polygon created from our bounding box list, name it, and add our style information. Finally, we check the map by displaying it.

[8]:
```
# create list of bounding box values
bbox_list =  [float(value) for value in bbox.split(',')]

# get centroid point from bounding box values
center = box(*bbox_list).centroid

# create map with folium with arguments for lat/lon of map, map tileset, and starting␣
↪zoom level
m = folium.Map(
    location=[center.y,center.x],
```

(continues on next page)

(continued from previous page)

```
    tiles="cartodbpositron",
    zoom_start=7,
)
# create style information for bounding box
bbox_style = {'fillColor': '#ff0000', 'color': '#ff0000'}

# create GeoJson of `geometries` and add to map
folium.GeoJson(geometries, name="tiles").add_to(m)

# create GeoJson of `bbox_list` polygon and add to map with specified style
folium.GeoJson(
    box(*bbox_list),
    name="bbox",
    style_function=lambda x:bbox_style
).add_to(m)

# display map
m
```

[8]:
```
<folium.folium.Map at 0x7f50e9d3b5d0>
```

Creating a Mosaic

Let's check the raster information contained in our item results. In order to do this, we import some more packages. - To read and write files in raster format, import `rasterio`. - From `rasterio` we import `merge` to copy valid pixels from an input into an output file - `AWSSession` to set up an Amazon Web Services (AWS) session - `show` to display images and label axes - Import `boto3` in order to work with AWS - From `matplotlib`, we want to import `imshow` which allows us to display images from data - Import `numpy` to work with multi-dimensional arrays and `numpy.ma` to work with masked arrays. - From `pyproj`, import `Proj` for converting between geographic and projected coordinate reference systems and `Transformer` to make transformations.

[9]:
```
# import rasterio for reading and writing in raster format
import rasterio as rio

# copy valid pixels from input files to an output file.
from rasterio.merge import merge

# set up AWS session
from rasterio.session import AWSSession

# display images, label axes
from rasterio.plot import show

# import boto3 to work with Amazon Web Services
import boto3

# display images from data
from matplotlib.pyplot import imshow

# import numpy to work with multi-dimensional arrays
import numpy as np

# import numpy.ma to work with masked arrays
import numpy.ma as ma
```

(continues on next page)

(continued from previous page)

```
# convert between geographic and projected coordinates and make transformations
from pyproj import Proj, Transformer
```

Finally, we import `os` and run some code in order to speed up Geospatial Data Abstraction Library (GDAL) reads from Amazon Simple Storage Service (S3) buckets by skipping sidecar (connected) files.

[10]:
```
# speed up GDAL reads from S3 buckets by skipping sidecar files
import os
os.environ['GDAL_DISABLE_READDIR_ON_OPEN'] = 'EMPTY_DIR'
```

Now that we have the necessary packages, let's get a list of S3 urls to the granules. To start, set up an AWS session. The S3 urls are contained within the `search.items()` assets, so we loop through the items in our results to get the S3 urls and add them to a new list. We can then use the `sort` function to sort the S3 urls in an acending order. Then we can check our S3 url list.

[11]:
```
# set up AWS session
aws_session = AWSSession(boto3.Session())

# get the S3 urls to the granules
file_S3 = [item.assets["cog_default"].href for item in search.items()]

# sort list in ascending order
file_S3.sort()

# check list
file_S3
```

[11]:
```
['s3://nasa-maap-data-store/file-staging/nasa-map/SRTMGL1_COD___001/N45W101.SRTMGL1.
↪tif',
 's3://nasa-maap-data-store/file-staging/nasa-map/SRTMGL1_COD___001/N45W102.SRTMGL1.
↪tif',
 's3://nasa-maap-data-store/file-staging/nasa-map/SRTMGL1_COD___001/N46W101.SRTMGL1.
↪tif',
 's3://nasa-maap-data-store/file-staging/nasa-map/SRTMGL1_COD___001/N46W102.SRTMGL1.
↪tif']
```

We can now check to see that we can read the AWS files and display a thumbnail. Pass the boto3 session to `rio.Env`, which is the GDAL/AWS environment for `rasterio`. Use the `rio.open` function to read in one of the Cloud Optimized GeoTIFFs.

Now let's use the `overviews` command to get a list of overviews. Overviews are versions of the data with lower resolution, and can thus increase performance in applications. Let's get the first overview from our list for retrieving a thumbnail.

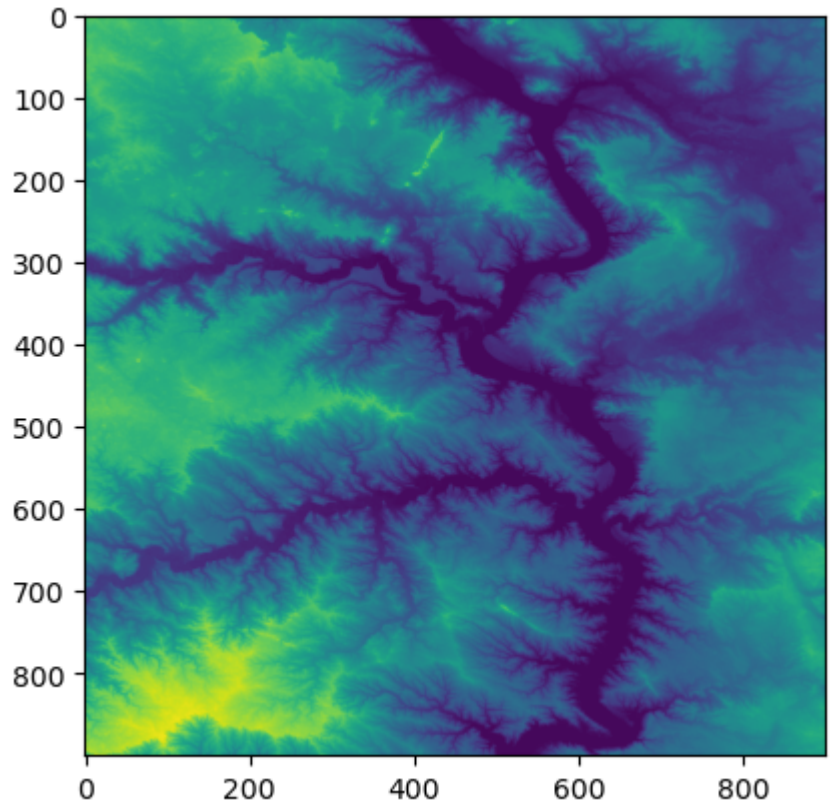
Retrieve a thumbnail by reading the first band of our file and setting the shape of the new output array. The shape can be set with a tuple of integers containing the number of datasets as well as the `height` and `width` of the file divided by our integer from the overview list. Now use the `show` function to display the thumbnail.

[12]:
```
# prove that we can read the AWS files
# for more information - https://automating-gis-processes.github.io/CSC/notebooks/L5/
↪read-cogs.html
with rio.Env(aws_session):
    with rio.open(file_S3[0], 'r') as src:
        # list of overviews
        oviews = src.overviews(1)
        # get second item from list to retrieve a thumbnail
        oview = oviews[1]
```

(continues on next page)

(continued from previous page)

```
        # read first band of file and set shape of new output array
        thumbnail = src.read(1, out_shape=(1, int(src.height // oview), int(src.width␣
↪// oview)))
# now display the thumbnail
show(thumbnail)
```



[12]:
```
<Axes: >
```

Since we verified that we can read the AWS files and display a thumbnail, we can create a mosaic from all of the rasters in our `file_S3` list. To do this, again pass the boto3 session to `rio.Env`. Then create a list which contains all of the read in Cloud Optimized GeoTIFFs (this may take a while).

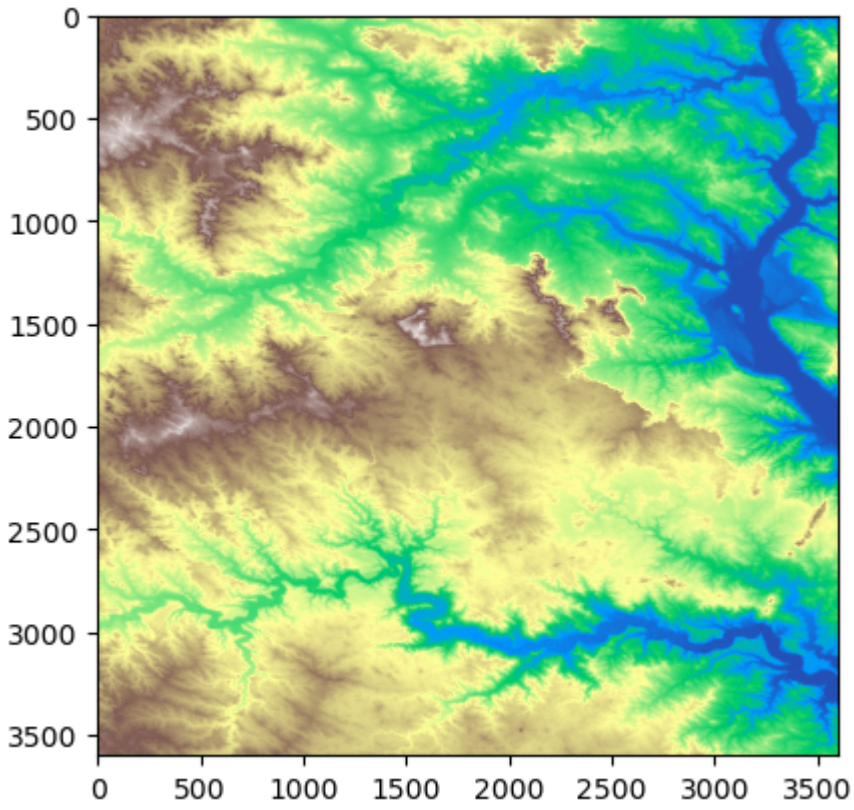
[13]:
```
# create a mosaic from all the images
with rio.Env(aws_session):
    sources = [rio.open(raster) for raster in file_S3]
```

Now we can use the `merge` function to merge these source files together using our list of bounding box values as the bounds. `merge` copies the valid pixels from input rasters and outputs them to a new raster.

[14]:
```
# merge the source files
mosaic, out_trans = merge(sources, bounds = bbox_list)
```

Lastly, we use ma.masked_values for masking all of the NoData values, allowing the mosaic to be plotted correctly. ma.masked_values returns a MaskedArray in which a data array is masked by an approximate value. For the parameters, let's use our mosaic as the data array and the integer of the "nodata" value as the value to mask by. Now we can use `show` to display our masked raster using `matplotlib` with a "terrain" colormap.

[15]:
```
# mask the NoData values so it can be plotted correctly
masked_mosaic = ma.masked_values(mosaic, int(sources[0].nodatavals[0]))

# display the masked mosaic
show(masked_mosaic, cmap = 'terrain')
```



[15]:
```
<Axes: >
```

3.3.9 Accessing EDAV Data via Web Coverage Service

This example demonstrates how to retrieve raster data from EDAV using a Web Coverage Service (WCS). A WCS lets you access coverage data with multiple dimensions online. The downloaded data is subsetted from data hosted on the MAAP and is available with the full resolution and values.

Setting up the Environment

We start by installing the libraries that are used to query the WCS connection point, and then to load, explore, and plot the raster data. We use `rasterio` for reading and writing raster formats, `rio-cogeo` for creating and validating Cloud Optimized GEOTIFF (COG) data, and `owslib` for interacting with Open Geospatial Consortium (OGC) services.

[1]:
```
# install libraries
# %pip is a magic command that installs into the current kernel
# -q means quiet (to give less output)
%pip install -q rasterio
```

(continues on next page)

(continued from previous page)

```
%pip install -q rio-cogeo
%pip install -q owslib
```

After installing the libraries, note that you may see multiple messages that you need to restart the kernel. Then import `rasterio`, `show` from `rasterio.plot` to display images with labeled axes, and `WebCoverageService` from `owslib.wcs` to program with an OGC web service.

[2]:
```
# import rasterio
import rasterio as rio
# import show
from rasterio.plot import show
# import WebCoverageService
from owslib.wcs import WebCoverageService
```

Querying the WCS

Now we can configure the WCS source, use the `getCoverage` function to request a file in GeoTIFF format, and save what is returned to our workspace.

[3]:
```
# configure the WCS source
EDAV_WCS_Base = "https://edav-wcs.adamplatform.eu/wcs"
wcs = WebCoverageService(f'{EDAV_WCS_Base}?service=WCS', version='2.0.0')
```

[4]:
```
# request imagery to download
response = wcs.getCoverage(
    identifier=['test_afrisar_onera_ClopeTB10_biomass_COG'], # coverage ID
    format='image/tiff', # format what the coverage response will be returned as
    filter='false', # define constraints on query
    scale=1, # resampling factor (1 full resolution, 0.1 resolution degraded of a␣
↪factor of 10)
    subsets=[('Long',11.54,11.8),('Lat',-0.3,0.0)] # subset the image by lat / lon
)

# save the results to file as a tiff
results = "EDAV_example.tif"
with open(results, 'wb') as file:
    file.write(response.read())
```

We can use `gdalinfo` to provide information about our raster dataset to make sure the data is valid and contains spatial metadata.

[5]:
```
# gives information about the dataset
!gdalinfo {results}
```

```
Driver: GTiff/GeoTIFF
Files: EDAV_example.tif
Size is 2836, 3403
Coordinate System is:
GEOGCRS["WGS 84",
    DATUM["World Geodetic System 1984",
        ELLIPSOID["WGS 84",6378137,298.257223563,
            LENGTHUNIT["metre",1]]],
    PRIMEM["Greenwich",0,
        ANGLEUNIT["degree",0.0174532925199433]],
    CS[ellipsoidal,2],
```

(continues on next page)

(continued from previous page)

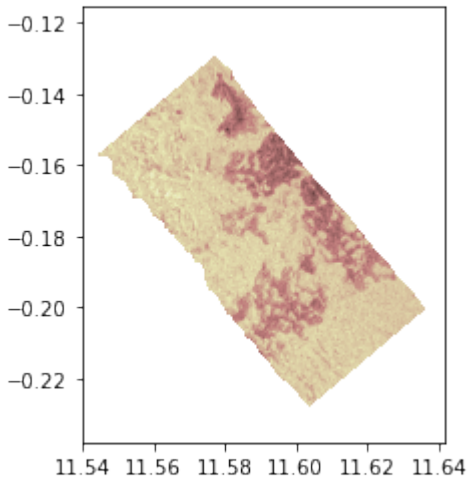
```
        AXIS["geodetic latitude (Lat)",north,
            ORDER[1],
            ANGLEUNIT["degree",0.0174532925199433]],
        AXIS["geodetic longitude (Lon)",east,
            ORDER[2],
            ANGLEUNIT["degree",0.0174532925199433]],
    ID["EPSG",4326]]
Data axis to CRS axis mapping: 2,1
Origin = (11.539968877500000,-0.115643000000000)
Pixel Size = (0.000035932500000,-0.000035932500000)
Metadata:
  AREA_OR_POINT=Area
Image Structure Metadata:
  COMPRESSION=LZW
  INTERLEAVE=BAND
Corner Coordinates:
Upper Left  (  11.5399689,  -0.1156430) ( 11d32'23.89"E,  0d 6'56.31"S)
Lower Left  (  11.5399689,  -0.2379213) ( 11d32'23.89"E,  0d14'16.52"S)
Upper Right (  11.6418734,  -0.1156430) ( 11d38'30.74"E,  0d 6'56.31"S)
Lower Right (  11.6418734,  -0.2379213) ( 11d38'30.74"E,  0d14'16.52"S)
Center      (  11.5909212,  -0.1767821) ( 11d35'27.32"E,  0d10'36.42"S)
Band 1 Block=2836x1 Type=Float32, ColorInterp=Gray
  NoData Value=0
```

Reading the Data

We can now use `rio.open` with our `results` path string and return an opened dataset object. We can set a variable (`rast`) to what is read from this dataset object. Then, we utilize the function `show` to display the raster using Matplotlib.

[6]:
```
# take path and return opened dataset object, set variable to read dataset object
with rio.open(results, 'r') as src:
        rast = src.read()
```

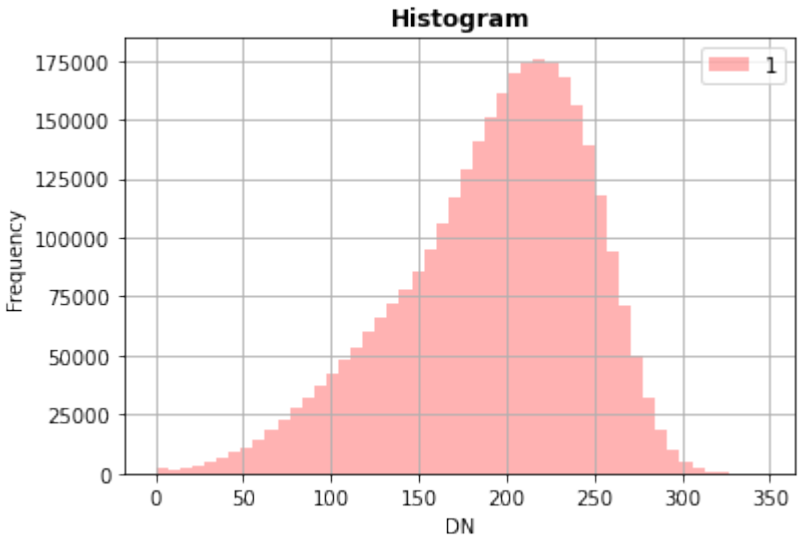
[7]:
```
# make a plot
show(rast, transform=src.transform, cmap='pink')
```



[7]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f27bc568c90>
```

We now have a visual of our raster. Let's import and employ the `show_hist` function from `rasterio.plot` to generate a histogram of the raster.

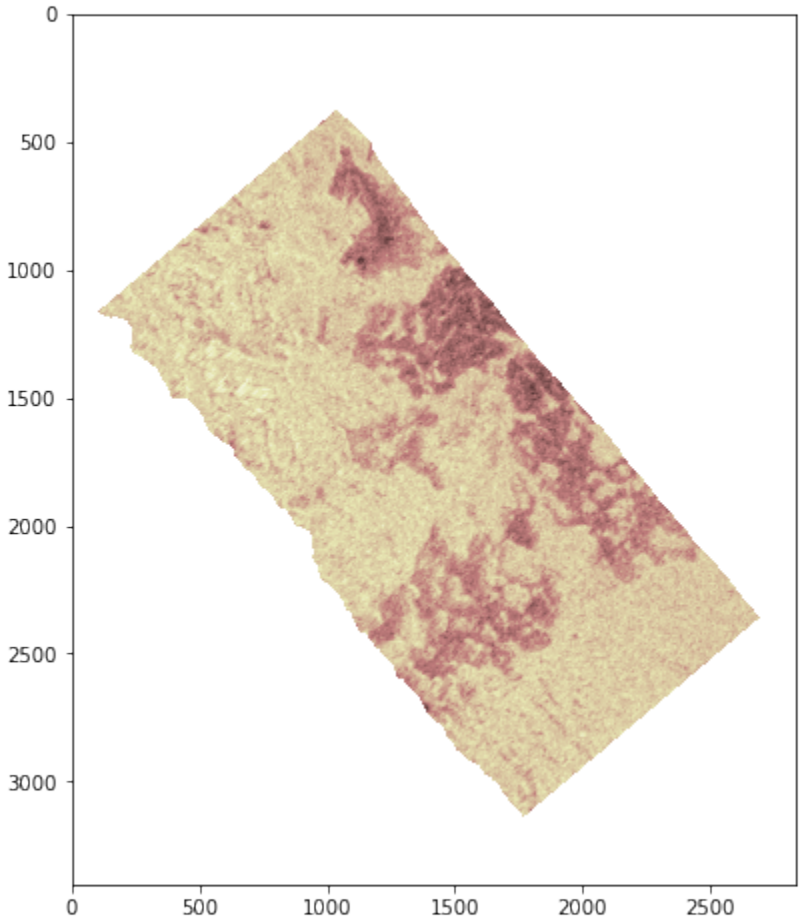
[8]:
```
# import show_hist
from rasterio.plot import show_hist
# create histogram
show_hist(rast,
          bins=50,# number of bins to compute histogram across
          alpha=.3,# transparancy
          title="Histogram"# figure title
         )
```



We can also generate a plot using Matplotlib. Let's import `matplotlib.pyplot` and `numpy` and make a new plot. To do this, use the `plt.subplots` function to return a figure and a single "Axes" instance. Then remove single-dimensional entries from the shape of our array using `np.squeeze` and display the data as an image using `imshow`. Now, we can set the norm limits for image scaling using the `set_clim` function.

[9]:
```
# import matplotlib.pyplot
import matplotlib.pyplot as plt
# import numpy
import numpy as np
```

[10]:
```
# set figure and single "Axes" instance
fig, ax = plt.subplots(1, figsize=(8,8))
# remove single-dimensional entries from the shape of the variable rast
#     and display the image
edavplot = ax.imshow(np.squeeze(rast), cmap='pink')
```



Newer Method `rioxarray`

Another way to work with raster data is with the `rasterio` "xarray" extension. Let's install and import `rioxarray` and create a plot using the `open_rasterio` and `plot` functions.

[11]:
```
# install rasterio xarray extension
%pip install -q rioxarray
```

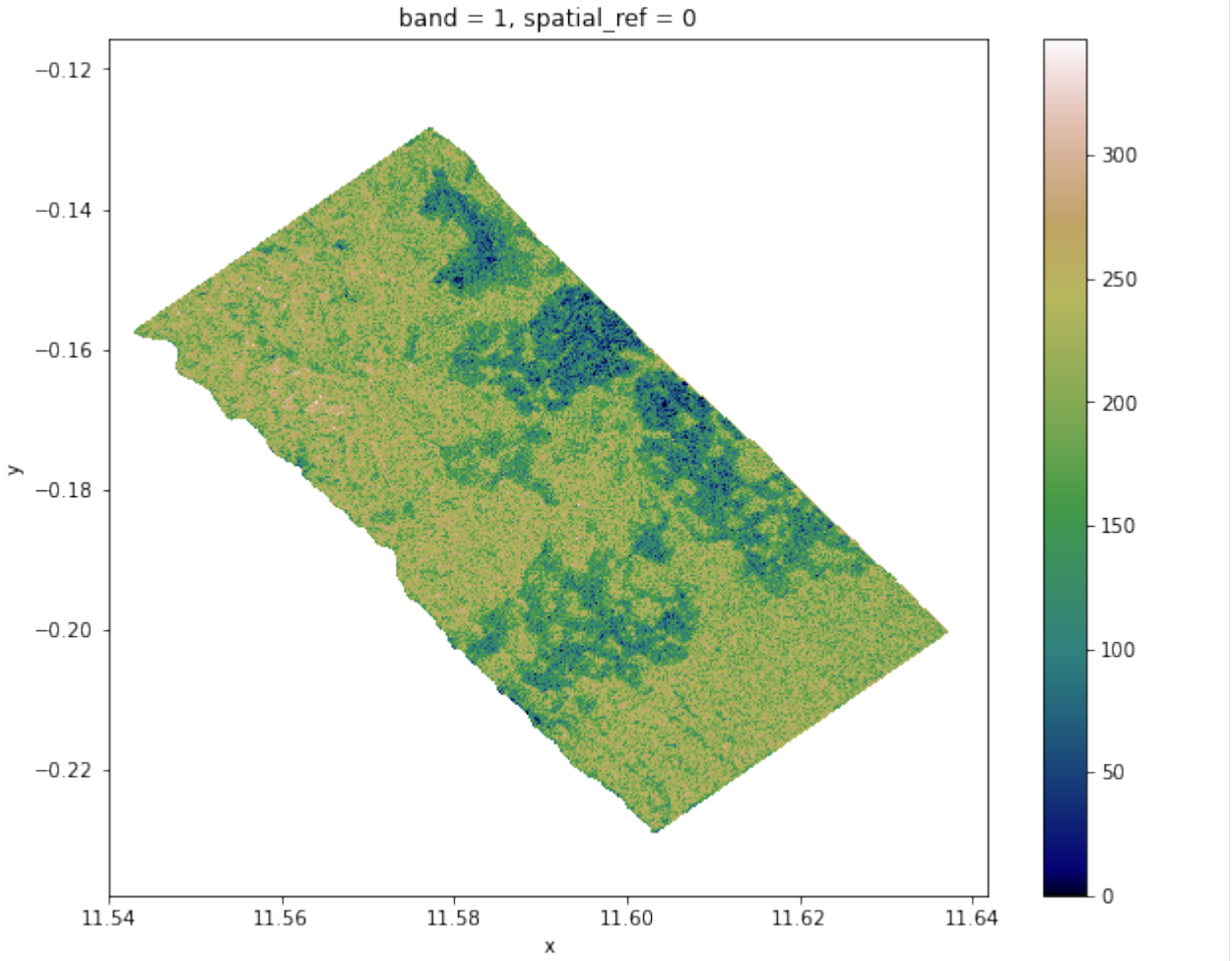
```
WARNING: Running pip as the 'root' user can result in broken permissions and␣
↪conflicting behaviour with the system package manager. It is recommended to use a␣
↪virtual environment instead: https://pip.pypa.io/warnings/venv
WARNING: You are using pip version 22.0.3; however, version 23.0.1 is available.
You should consider upgrading via the '/opt/conda/bin/python -m pip install --upgrade␣
↪pip' command.
Note: you may need to restart the kernel to use updated packages.
```

[12]:
```
# import rasterio xarray extension
import rioxarray
```

[13]:
```
# opens results with rasterio to set dataarray
edav_x = rioxarray.open_rasterio(results)
```

[14]:
```
# plot dataarray
edav_x.plot(cmap="gist_earth", figsize=(10,8))
```

[14]:
```
<matplotlib.collections.QuadMesh at 0x7f275e7e5310>
```



References

- WCS Adapted from: Jan Verbesselt, Jorge Mendes de Jesus, Aldo Bergsma, Dainius Masiliūnas, David Swinkels, Corné Vreugdenhil. Handling Raster data with Python - 2020-01-20 https://geoscripting-wur.github.io/PythonRaster/
- OWSLib https://github.com/geopython/OWSLib
- rioxarray https://corteva.github.io/rioxarray

3.4 Query

3.4.1 How to Query Data from the MAAP via Python Client

Supported collections can be subsetted through the MAAP Query Service. At the time of writing (03/22/2021), the GEDI Calibration/Validation Field Survey Dataset is the only valid dataset for this service. However, more data will be made available for querying as the MAAP team continues to develop expanded services for the platform. Users can interact with the service through the MAAP Python client.

First, we import the `json` module, import the `MAAP` package, and create a new MAAP class.

[18]:
```
# import the json module
import json
# import the MAAP package
from maap.maap import MAAP
# create MAAP class
maap = MAAP()
```

How to Use `maap.executequery()`

We use the `executeQuery()` function to return a response object, containing the server's response to our HTTP request. This object can be used to view the response headers, access the raw data of the response, or parse the response as a JavaScript Object Notation (JSON). JSON is a data-interchange format, designed to be easy for humans to read and write.

`executeQuery` Parameters

- `src` - a dictionary-like object specifying the dataset to query. Currently, the only option is as follows:

```
{
  "Collection": {
    "ShortName": "GEDI Cal/Val Field Data_1",
    "VersionId": "2"
  }
}
```

- `query` - a dictionary-like object specifying the parameters for query. A dictionary can include `bbox`, `where`, `fields`, and `table`:
  - `bbox` - a list of floating point numbers identifying a bounding box of geographic coordinates.
  - `where` - a dictionary-like object which maps fields to required values within a query.
  - `fields` - a list of fields to return, a subset of all fields available for the corresponding dataset.
  - `table` - the name of the table to query. At this time, the only valid options are "tree" or "plot" corresponding to tables in the GEDI Cal/Val database.
- `poll_results` - a parameter which must be `True` to use the `timeout` parameter.
- `timeout` - the waiting period for a response. This indicates the maximum number of seconds to wait for a response. Note that timeout has a default value of '180' and requires that the `poll_results` parameter be `True`. Depending on the request, it may be necessary to modify the timeout to make sure the server has enough time to process the request.

- `wait_interval` - number of seconds to wait between each poll for results. `wait_interval` is only used if `poll_results=True` (default 0.5). -`max_redirects` - the maximum number of redirects to follow when scheduling an execution (default 5).

Query Searching for a Project Name

In this example, we create a dictionary containing a `Collection` key, which contains entries for `ShortName` and `VersionId`. This is used later in the `executeQuery()` function. For this example, we use the GEDI Calibration/Validation Field Survey Dataset collection. More information about dictionaries in Python can be found here.

[19]:
```
# create dictionary with a "Collection" key containing short name and version ID␣
↪entries:
collection  = {
  "Collection": {
      "ShortName": "GEDI Cal/Val Field Data_1",
      "VersionId": "2"
  }
}
```

We also create a function (in this example named `fetch_results`) which utilizes the `executeQuery` function to return results of queries. Within this function, we use the `executeQuery` function to get a response object. Within the `executeQuery` function, our `collection` dictionary is assigned to `src`, a dictionary-like object specifying the dataset to query. We also have the query used in the argument assigned to `query`, a dictionary-like object which specifies the parameters for the query. We set `timeout` to the timeout used in the argument (default is 180 seconds) and set `poll_results` to `True` in order to set the maximum waiting period for a response. We can check the 'Content-Type' header of our response to see the content type of the response. In the following code, the 'Content-Type' header is checked to determine if it is JSON or not, in order to set an appropriate variable to return.

[20]:
```
def fetch_results(query={}, timeout=180):
    """
    Function which utilizes the `executeQuery` function to return the results of␣
↪queries.
    """
    # use the executeQuery() function to get a response object
    response = maap.executeQuery(
      # dictionary-like object specifying the dataset to query
      src = collection,
      # dictionary-like object specifying the parameters for query
      query = query,
      # must be True to use the timeout parameter
      poll_results = True,
      # max waiting period for a response in seconds
      timeout = timeout
    )
    # if the 'Content-Type' is json, creates variable with json version of the␣
↪response
    if (response.headers.get("Content-Type") == "application/json"):
        data = response.json()
    # if the 'Content-Type' is not json, creates variable with unicode content of the␣
↪response
    else:
        data = response.text
    # returns `data` as json string
    return json.loads(data)
```

Now that we have our collection dictionary and our function to return the results of queries, let's use a print statement to display the first project name from a query which utilizes the `bbox` and `fields` parameters within `query`. The `bbox`

parameter is a GeoJSON-compliant bounding box ([minX, minY, maxX, maxY]) which is used to filter data spatially. GeoJSON is a format for encoding geographic data structures. More information about the bounding box can be found in the standard specification of the GeoJSON format, located here - https://tools.ietf.org/html/rfc7946#section-5. The `fields` parameter is a list of fields to return in the query response. In this case, we assign 'project' to `fields`.

[21]:
```
# prints the first project name in the results of the query as a json string
print(json.dumps(fetch_results({"bbox": [9.31, 0.53, 9.32, 0.54], "fields": ["project
↪"]})[0], indent=2))
```

```
{
  "project": "gabon_mondah"
}
```

Inspecting a Single Observation

In the previous example, we displayed the project name for a result from our query, but let's say we wished to see all of the fields and their associated values for a result. In this example, we make another query, this time only specifying the bounding box. The print statement displays the variables for a single observation. A list of the variables and their units and descriptions can be found here.

[22]:
```
# prints the fields with values for the first result in the results of the query as a␣
↪json string
print(json.dumps(fetch_results({"bbox": [9.315, 0.535, 9.32, 0.54]})[0], indent=2))
```

```
{
  "project": "gabon_mondah",
  "plot": "NASA11",
  "subplot": "1",
  "survey": "AfriSAR_ESA_2016",
  "private": 0,
  "date": "2016-02-01",
  "region": "Af",
  "vegetation": "TropRF",
  "map": 3083.93471636915,
  "mat": 25.6671529098763,
  "pft.modis": "Evergreen Broadleaf trees",
  "pft.name": null,
  "latitude": 0.538705025207016,
  "longitude": 9.31982893597376,
  "p.sample": 0,
  "p.stemmap": 0,
  "p.origin": "C",
  "p.orientation": -2.18195751718555,
  "p.shape": "R",
  "p.majoraxis": 100,
  "p.minoraxis": 100,
  "p.geom": "POLYGON ((535537.75 59601.25, 535627.75 59590.5, 535642.25 59489.25,␣
↪535543 59498.5, 535537.75 59601.25, 535537.75 59601.25))",
  "p.epsg": 32632,
  "p.area": 10000,
  "p.mindiam": 0.01,
  "sp.geom": "POLYGON((535537.510093 59494.109258, 535537.526710 59519.109253, 535562.
↪526704 59519.092636, 535562.510088 59494.092642, 535537.510093 59494.109258))",
  "sp.ix": 1,
  "sp.iy": 4,
  "sp.shape": "R",
  "sp.area": 625,
```

(continues on next page)

(continued from previous page)


```
  "sp.mindiam": 0.01,
  "pai": null,
  "lai": null,
  "cover": null,
  "dft": "EA",
  "agb": null,
  "agb.valid": null,
  "agb.lower": null,
  "agb.upper": null,
  "agbd.ha": null,
  "agbd.ha.lower": null,
  "agbd.ha.upper": null,
  "sn": null,
  "snd.ha": null,
  "sba": null,
  "sba.ha": null,
  "swsg.ba": null,
  "h.t.max": null,
  "sp.agb": null,
  "sp.agb.valid": null,
  "sp.agbd.ha": null,
  "sp.agbd.ha.lower": null,
  "sp.agbd.ha.upper": null,
  "sp.sba.ha": null,
  "sp.swsg.ba": null,
  "sp.h.t.max": null,
  "l.project": "jpl_mondah",
  "l.instr": null,
  "l.epsg": 32632,
  "l.date": null,
  "tree.date": "2016-02-10",
  "family": "Myristicaceae",
  "species": "Coelocaryon sp.",
  "pft": null,
  "wsg": 0.49353216374269,
  "wsg.sd": 0.0941339098036177,
  "tree": "5501",
  "stem": "1",
  "x": 535539.544644182,
  "y": 59496.0746845487,
  "z": null,
  "status": 1,
  "allom.key": 2,
  "a.stem": 0.0437435361085843,
  "h.t": 9.46667,
  "h.t.mod": 17.0934257579035,
  "d.stem": 0.236,
  "d.stem.valid": 1,
  "d.ht": 1.3,
  "c.w": null,
  "m.agb": 145.005237378992,
  "id": 2077,
  "geom_obj":
↪"0103000020E6100000010000000500000032A8C5A385A32240A1FCE9F75E39E13F32892DA985A32240A012DE4C393BE13F
↪",
  "wwf.ecoregion": "Central African mangroves"
}
```

Query Using Multiple Parameters With `where`

In the output of the previous example, we can see the field `"species"`. Let's say we are interested in finding observations for the `"gabon_mondah"` project within the same bounding box as the previous example which have the same species. We can do this using `where`, a dictionary-like object which maps fields to required values within a query. To help demonstrate how to use `where`, we can create a new function (in this example named `species_query`) to print the number of results as well as the first result, adding on to our previous `fetch_results` function which utilized the `executeQuery` function.

[23]:
```
def species_query(query = {}, timeout = 180):
    """
    Function which utilizes the `fetch_results` (and thereby `executeQuery`) function␣
↪and prints the number of results
    as well as the first result.
    """
    # set `data` to results of query
    data = fetch_results(query = query, timeout = timeout)
    # get the number of results within `data`
    num_results = len(data)
    # if `data` is not null and contains at least one result, the number of results␣
↪and the first result are printed
    if((data is not None) and (num_results > 0)):
        first_result = data[0]
        print(f"Number of results: {num_results}")
        print(f"First result: {json.dumps(first_result, indent=2)}")
    # else prints "No result"
    else:
        print(num_results)
        print("No result")
```

Let's call the `species_query` function. We enter the same bounding box values as in the previous example. This time around, we enter `where` in our query and set the `project` as `gabon_mondah` and the `species` as `Coelocaryon sp.`. We can set a list of fields to return in query response using `fields`. For this example, we can choose to return only the project, family, species, latitude, and longitude values. After completing our query, we can manually set the timeout value (in this example '200').

[24]:
```
# call `species_query` function with bounding box values, where the project is "gabon_
↪mondah",
# the species is "Aucoumea klaineana", fields include "project", "family", "species",
↪"latitude", and "longitude",
# and the timeout value is 200 (use the scrollbar to see the entire function call)
species_query({"bbox": [9.315, 0.535, 9.32, 0.54], "where": {"project": "gabon_mondah
↪", "species": "Coelocaryon sp."}, "fields": ["project", "family", "species",
↪"latitude", "longitude"]}, 200)
```

```
Number of results: 10
First result: {
  "project": "gabon_mondah",
  "family": "Myristicaceae",
  "species": "Coelocaryon sp.",
  "latitude": 0.538705025207016,
  "longitude": 9.31982893597376
}
```

We now see that there are many results and the latitude and longitude coordinates for the first result. To see this information for Aucoumea klaineana, we can copy the code from the above cell, changing the `species` to `Aucoumea klaineana` within the function argument.

[25]:
```
# call `species_query` function with bounding box values, where the project is "gabon_
↪mondah",
# the species is "Coelocaryon sp.", fields include "project", "family", "species",
↪"latitude", and "longitude",
# and the timeout value is 200 (use the scrollbar to see the entire function call)
species_query({"bbox": [9.315, 0.535, 9.32, 0.54], "where": {"project": "gabon_mondah
↪", "species": "Aucoumea klaineana"}, "fields": ["project", "family", "species",
↪"latitude", "longitude"]}, 200)
```

```
Number of results: 49
First result: {
  "project": "gabon_mondah",
  "family": "Burseraceae",
  "species": "Aucoumea klaineana",
  "latitude": 0.538705025207016,
  "longitude": 9.31982893597376
}
```

Using `executeQuery` to Vizualize Plots and Trees

Now that we know how to print the query results of the collection, let's look at examples of vizualizing the information from query results. The first time you run this, you will need to install `folium`. `folium` allows us to visualize data on an interactive map. With `folium` installed, you will then import the `folium` library. Also import `matplotlib.pyplot`, which allows us to generate interactive plots

[26]:
```
# installs folium
!pip install folium
```

```
Requirement already satisfied: folium in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (0.12.1.post1)
Requirement already satisfied: numpy in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from folium) (1.21.5)
Requirement already satisfied: jinja2>=2.9 in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from folium) (3.0.3)
Requirement already satisfied: requests in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from folium) (2.27.1)
Requirement already satisfied: branca>=0.3.0 in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from folium) (0.4.2)
Requirement already satisfied: MarkupSafe>=2.0 in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from jinja2>=2.9->folium) (2.0.1)
Requirement already satisfied: certifi>=2017.4.17 in /Users/philvarner/code/devseed/
↪maap-documentation/.venv/lib/python3.7/site-packages (from requests->folium) (2021.
↪10.8)
Requirement already satisfied: idna<4,>=2.5 in /Users/philvarner/code/devseed/maap-
↪documentation/.venv/lib/python3.7/site-packages (from requests->folium) (3.3)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /Users/philvarner/code/
↪devseed/maap-documentation/.venv/lib/python3.7/site-packages (from requests->
↪folium) (1.26.8)
Requirement already satisfied: charset-normalizer~=2.0.0 in /Users/philvarner/code/
↪devseed/maap-documentation/.venv/lib/python3.7/site-packages (from requests->
↪folium) (2.0.11)
```

[27]:
```
# import folium library and matplotlib.pyplot
import folium
import matplotlib.pyplot as plt
```

Plot Project Plots

Now we can begin plotting all the plots for a given project. For this example, we'll check out the `australia_ausplotsforests` project. We can query the plot table by using our `fetch_results` helper function. This time, we set the `table` parameter. `table` is the name of the table to query. The GEDI Cal/Val database has tables for "tree" and "plot" so let's set `table` to "plot". We can then preview our results.

[28]:
```
# set project and query for all the plots in that project
project = 'australia_ausplotsforests'
results = fetch_results({
    "bbox": [-180, 90, 180, -90],
    "where": {
        "project": project
    },
    "fields": ["latitude", "longitude", "plot"],
    "table":"plot"
}, 1000)
# print first 10 results
results[0:10]
```

[28]:
```
[{'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'},
 {'latitude': -31.2421, 'longitude': 152.4609, 'plot': 'NSFNNC002'}]
```

Create Project Plots Dictionary

Now we can create a dictionary for the project plots from our results.

[29]:
```
# create dictionary for project plots, where each plot has lat/lon info
project_plots = {}
keys = [ 'latitude', 'longitude' ]
for result in results:
    project_plots[result['plot']] = { key: result[key] for key in keys }
```

To center our map more easily and plot trees later on, we can set a variable to the first project plot to be used for centering the map.

[30]:
```
# Select the first plot, just to center the map easily
first_plot = list(project_plots.keys())[0]
first_plot
```

[30]:
```
'NSFNNC002'
```

Map the Plots

Now we can create a map and place the plot points onto it. First we will set variables for the map's location center and starting zoom level. Then we will use the `folium.Map` function to create a map which displays "Stamen Terrain" tiles with our location center and starting zoom level. Next, we will create a loop using the `folium.Marker` function

to add markers to the map at the latitude and longitude for each plot and have popup text displaying each plot's name. Lastly, we display the map in order to interact with it.

[31]:
```
# set location center for map
center = [ project_plots[first_plot]['latitude'], project_plots[first_plot]['longitude
↪'] ]
# set zoom level, note that depending on the project,
# you may wish to increase the zoom level
zoom = 3
# create a map with Folium and Leaflet.js
m = folium.Map(location=center, tiles="Stamen Terrain", zoom_start = zoom)
# Add markers to map for each plot in `project_plots`
for plot in project_plots.items():
    folium.Marker(
        [plot[1]['latitude'], plot[1]['longitude']],
        popup = f"plot: {plot[0]}"
    ).add_to(m)
# display map
m
```

[31]:
```
<folium.folium.Map at 0x125d8ccd0>
```

The map has zoom in and zoom out buttons, can be dragged with a mouse, and displays markers which can be clicked on to display their popup text.

Plot Trees

Now let's look at plotting trees for the first plot of the designated project. We construct another query with our `fetch_results` helper function. This time, we query for fields containing plant height, UTM coordinates, and elevation as well as set the `table` parameter to "tree".

[32]:
```
# query for trees for the first plot of the project
results = fetch_results({
    "bbox": [-180, 90, 180, -90],
    "where": {
        "project": project,
        "plot": first_plot
    },
    # h.t -> total height of plant from ground to highest leaf
    # x -> easting UTM coordinate
    # y -> northing UTM coordinate
    # z -> elevation relative to geoid coordinate
    "fields": [ "h.t", "x", "y", "z"],
    "table": "tree"
}, 1000)
```

Determine the Number and Height of Trees

To see the number of trees in the list of results, we use the `len` function within a formatted print statement. Since not all of the results contain heights, lets check how many trees with height data exist by creating a new variable set to the results with heights and use the `len` function within a formatted print statement again. Also, we can set variables to contain the heights, x and y values contained within our results to be used when plotting the trees.

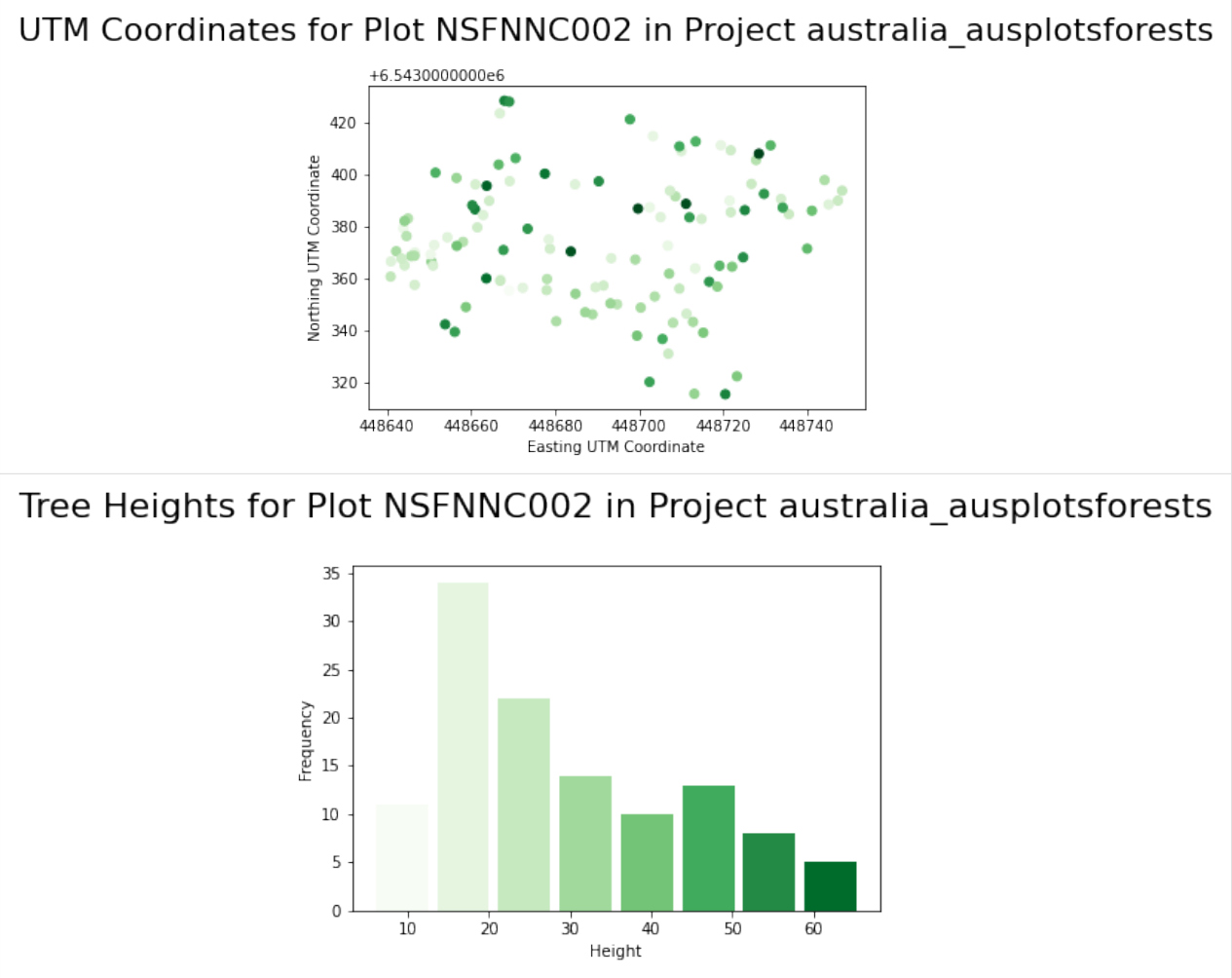
[33]:
```
# print number of trees in results
print(f"Number of trees: {len(results)}")
# print number of trees with heights
heights = [r["h.t"] for r in results if r["h.t"] is not None]
print(f"Number of trees with heights: {len(heights)}")
# set variables to the heights, x and y values contained in the results
hts = [r['h.t'] for r in results]
xs = [r['x'] for r in results]
ys = [r['y'] for r in results]
```

```
Number of trees: 520
Number of trees with heights: 117
```

Plot the Trees

With our variables for tree heights, x and y coordinate values, let's use the `plt.scatter` function to create a scatter plot with a title and labels for the UTM coordinates of the trees. Lastly, let's also create a histogram using `plt.hist` to visualize the distribution of tree heights. Here we also include some code to adjust the display colors and add labels and a title.

[34]:
```
# create scatter plot with x and y values
plt.scatter(xs, ys, c=hts, cmap="Greens")
# add labels and title
plt.xlabel('Easting UTM Coordinate')
plt.ylabel('Northing UTM Coordinate')
plt.title(f"UTM Coordinates for Plot {first_plot} in Project {project}\n",␣
↪fontsize=22)
plt.show()
# adjust display colors
cm = plt.cm.Greens
nbins = 8
# create histogram for tree heights
n, bins, patches = plt.hist(x=heights, bins=nbins, rwidth=0.85)
for i, p in enumerate(patches):
    plt.setp(p, 'facecolor', cm(i/nbins))
# add labels and title
plt.xlabel('Height')
plt.ylabel('Frequency')
plt.title(f"Tree Heights for Plot {first_plot} in Project {project}\n", fontsize=22)
plt.show()
```



3.4.2 Fields Within the Gedi Cal/Val Data

Last updated: 2020-06-30

The Global Ecosystem Dynamics Investigation (GEDI) Forest Structure and Biomass Database (FSBD) is a collection of field and LiDAR datasets developed to serve a central repository for calibration and validation of the GEDI mission data. The GEDI Cal/Val Field Data collection contains the field data contributions to the FSBD. The database has contributions across the globe for a variety of vegetation types from projects dating back to 2003.

The table below describes the variables found in the files:

variable	units	description
plot	NA	name of plot
subplot	NA	subplot identifier
survey	NA	name of survey event
private	NA	data privacy (1=private, 0=public)
date	NA	date of survey event
region	NA	region of plot location: Eu=Europe; As=Asia; Au=Australia/New Zealand; Af=Africa; SEAsia=Sout
vegetation	NA	vegetation type where sampled: Sav = Savannah; TropRF = Tropical rainforest; TempRF = Temperat

Table 1

variable	units	description
map	mm	mean annual rainfall
mat	deg	mean annual temperature
pft.modis	NA	plant functional type from the nearest MCD12Q1 product date
pft.name	NA	plant functional type from field descriptions or other ancillary data
wwf.ecoregion	NA	Terrestrial ecoregion from the WWF Olson et al. (2004) terrestrial ecoregions of the globe
latitude	deg	latitude of location where sampled (-90 to 90 deg South to North)
longitude	deg	longitude of location where sampled (-180 to 180 deg West to East)
p.sample	NA	subplot type: 0=independent fixed area plots; 1=variable area plots; 2=multiple fixed area plots for ag
p.stemmap	NA	stem mapped plot (1=TRUE,0=FALSE)
p.origin	NA	origin of the plot, SW has to be interpreted as the lower left corner relative to the orientation
p.orientation	deg	orientation of plot degrees clockwise from UTM map grid north
p.shape	NA	plot shape (E=ellipse, R=rectangle)
p.majoraxis	m	major axis of plot
p.minoraxis	m	minor axis of plot
p.geom	m	plot geometry WKT string
p.epsg	NA	epsg code of the UTM zone used for the tree coordinates
p.area	m2	plot polygon area
p.mindiam	m	minimum tree diameter measured across all subplots
sp.geom	m	subplot geometry WKT string
sp.ix	NA	subplot track index
sp.iy	NA	subplot pulse index
sp.shape	NA	subplot shape (E=ellipse, R=rectangle)
sp.area	m2	subplot area
sp.mindiam	m	minimum tree diameter measured for the subplot
pai	m2/m2	plant area index of vegetation
lai	m2/m2	leaf area index of vegetation
cover	NA	vertically projected canopy cover (1-Pgap)
dft	NA	dominant plant functional type: EA = evergreen angiosperm; DA = deciduous angiosperm; EG = eve
agb	kg	mass of above-ground standing trees and shrubs
agb.valid	NA	valid above ground mass prediction (1=TRUE,0=FALSE)
agb.lower	kg	lower limit of standard error of mass of above-ground standing trees and shrubs
agb.upper	kg	upper limit of standard error of mass of above-ground standing trees and shrubs
agbd.ha	Mg/ha	mass density of above-ground standing trees and shrubs
agbd.ha.lower	Mg/ha	lower limit of standard error of mass density of above-ground standing trees and shrubs
agbd.ha.upper	Mg/ha	upper limit of standard error of mass density of above-ground standing trees and shrubs
sn	n	plot stem number
snd.ha	n/ha	plot stem number density
sba	m2	plot stand basal area
sba.ha	m2/ha	plot stand basal area per hectare
swsg.ba	g/cm3/m2	basal-area-weighted wood specific gravity
h.t.max	m	maximum total height of plants from ground to highest leaf
sp.agb	kg	subplot mass of above-ground standing trees and shrubs
sp.agb.valid	NA	valid subplot above ground mass prediction (1=TRUE,0=FALSE)
sp.agbd.ha	Mg/ha	subplot mass density of above-ground standing trees and shrubs
sp.agbd.ha.lower	Mg/ha	subplot lower limit of standard error of mass density of above-ground standing trees and shrubs
sp.agbd.ha.upper	Mg/ha	subplot upper limit of standard error of mass density of above-ground standing trees and shrubs
sp.sba.ha	m2/ha	subplot stand basal area per hectare
sp.swsg.ba	g/cm3/m2	subplot basal-area-weighted wood specific gravity
sp.h.t.max	m	subplot maximum total height of plants from ground to highest leaf

Table 1

variable	units	description
l.project	NA	name of the lidar project dataset on UMD servers
l.instr	NA	lidar instrument manufacturer and model
l.epsg	NA	epsg code of the corresponding lidar data
l.date	NA	date of the corresponding lidar acquisition
g.fp	NA	geometry collection WKT string of GEDI lidar footprint center locations
tree.date	NA	date of tree measurement
family	NA	latin name of botanical family
species	NA	latin name of species (genus species)
pft	NA	plant functional type: EA = evergreen angiosperm; DA = deciduous angiosperm; EG = evergreen gy
wsg	g/cm3	the ratio of wood density to water
wsg.sd	g/cm3	standard deviation of WSG sample values
tree	NA	tree identifier
stem	NA	stem identifier
x	m	easting UTM coordinate
y	m	northing UTM coordinate
z	m	elevation relative to geoid coordinate
status	NA	live tree (1=TRUE,0=FALSE)
allom.key	NA	key to allometric LUT
a.stem	m2	area of total stem cross-section at measurement height
h.t	m	total height of plant from ground to highest leaf
h.t.mod	NA	tree heights modelled using local or regional DBH-Ht relationship
d.stem	m	diameter of stem at measurement height
d.stem.valid	NA	valid tree diameter to use in AGB calculation (1=TRUE,0=FALSE)
d.ht	m	height at which stem diameter was measured
c.w	m	diameter or width of crown
m.agb	kg	mass of above-ground components of tree

3.5 Adding Cloud-Optimized GeoTIFFs to the MAAP Dashboard

The following notebook steps through how to add a dataset to the MAAP Dashboard.

Note, there are 2 scenarios:

1. Adding a single Cloud-Optimized GeoTIFF (COG) and,
2. Adding many distinct COGs as a "mosaic" with mosaicJSON.

High-level, the steps are:

1. Inspect your COG(s) to understand the best rescale and colormap name parameters. Optionally create a mosaic.
2. Define a colormap. Colormaps provide mappings of data values to RGB values.
3. Create a PR to the datasets repo to add or update your dataset.

The MAAP dashboard has 3 environments:

1. Developer-in-test (DIT): https://biomass.dit.maap-project.org
2. Staging: https://biomass.staging.maap-project.org
3. Production: https://biomass.maap-project.org

These instructions will guide you towards adding your dataset to `biomass.dit.maap-project.org`. The MAAP Dashboard team will "promote" changes to staging and production periodically (release schedule forthcoming).

3.5.1 Setup

[1]:
```
%%capture
!pip install rasterio rio-cogeo supermercado
```

[3]:
```
# import the MAAP package
import ipycmc
from maap.maap import MAAP

import glob
import json
import os
import matplotlib
import urllib

# create MAAP class
maap = MAAP(maap_host='api.maap-project.org')
```

[4]:
```
project_dir = "/projects/shared-buckets/<your_name>/<project_dir>"
# e.g.
project_dir = "/projects/shared-buckets/alexdevseed/landsat8/viz/"
```

[5]:
```
# Search for files to include, use recursive if nested folders (common in DPS output)
files = glob.glob(os.path.join(project_dir, "*.tif"), recursive=False)
files = [os.path.basename(f) for f in files]
#files
```

[6]:
```
my_tif = files[1]
```

[7]:
```
%%bash -s "$project_dir" "$my_tif"
rio cogeo validate $1/$2
```

```
/projects/shared-buckets/alexdevseed/landsat8/viz/Copernicus_30439_covars_cog_topo_
↪stack.tif is a valid cloud optimized GeoTIFF
```

[8]:
```
%%bash -s "$project_dir" "$my_tif"
gdalinfo $1/$2 -stats
```

```
Driver: GTiff/GeoTIFF
Files: /projects/shared-buckets/alexdevseed/landsat8/viz//Copernicus_30439_covars_cog_
↪topo_stack.tif
       /projects/shared-buckets/alexdevseed/landsat8/viz//Copernicus_30439_covars_cog_
↪topo_stack.tif.aux.xml
Size is 1000, 1000
Coordinate System is:
PROJCRS["unknown",
    BASEGEOGCRS["NAD83",
        DATUM["North American Datum 1983",
            ELLIPSOID["GRS 1980",6378137,298.257222101004,
                LENGTHUNIT["metre",1]]],
        PRIMEM["Greenwich",0,
```

(continues on next page)

(continued from previous page)

```
            ANGLEUNIT["degree",0.0174532925199433]],
        ID["EPSG",4269]],
    CONVERSION["Albers Equal Area",
        METHOD["Albers Equal Area",
            ID["EPSG",9822]],
        PARAMETER["Latitude of false origin",40,
            ANGLEUNIT["degree",0.0174532925199433],
            ID["EPSG",8821]],
        PARAMETER["Longitude of false origin",180,
            ANGLEUNIT["degree",0.0174532925199433],
            ID["EPSG",8822]],
        PARAMETER["Latitude of 1st standard parallel",50,
            ANGLEUNIT["degree",0.0174532925199433],
            ID["EPSG",8823]],
        PARAMETER["Latitude of 2nd standard parallel",70,
            ANGLEUNIT["degree",0.0174532925199433],
            ID["EPSG",8824]],
        PARAMETER["Easting at false origin",0,
            LENGTHUNIT["metre",1],
            ID["EPSG",8826]],
        PARAMETER["Northing at false origin",0,
            LENGTHUNIT["metre",1],
            ID["EPSG",8827]]],
    CS[Cartesian,2],
        AXIS["easting",east,
            ORDER[1],
            LENGTHUNIT["metre",1,
                ID["EPSG",9001]]],
        AXIS["northing",north,
            ORDER[2],
            LENGTHUNIT["metre",1,
                ID["EPSG",9001]]]]
Data axis to CRS axis mapping: 1,2
Origin = (698522.000000000000000,3153304.000000000000000)
Pixel Size = (30.000000000000000,-30.000000000000000)
Metadata:
  AREA_OR_POINT=Area
  OVR_RESAMPLING_ALG=NEAREST
Image Structure Metadata:
  COMPRESSION=DEFLATE
  INTERLEAVE=PIXEL
  LAYOUT=COG
Corner Coordinates:
Upper Left  (  698522.000, 3153304.000) (163d20'28.79"W, 67d29'44.89"N)
Lower Left  (  698522.000, 3123304.000) (163d30'48.19"W, 67d14'14.17"N)
Upper Right (  728522.000, 3153304.000) (162d39'23.92"W, 67d25'44.51"N)
Lower Right (  728522.000, 3123304.000) (162d50' 6.57"W, 67d10'16.36"N)
Center      (  713522.000, 3138304.000) (163d 5'11.87"W, 67d20' 1.17"N)
Band 1 Block=256x256 Type=Float32, ColorInterp=Gray
  Description = elevation
  Min=-1.830 Max=557.863
  Minimum=-1.830, Maximum=557.863, Mean=74.140, StdDev=84.128
  Overviews: 500x500, 250x250
  Metadata:
    STATISTICS_MAXIMUM=557.86291503906
    STATISTICS_MEAN=74.140305286651
    STATISTICS_MINIMUM=-1.8297636508942
```

(continues on next page)

(continued from previous page)

```
    STATISTICS_STDDEV=84.128259469301
    STATISTICS_VALID_PERCENT=100
Band 2 Block=256x256 Type=Float32, ColorInterp=Undefined
  Description = slope
  Min=0.000 Max=40.566
  Minimum=0.000, Maximum=40.566, Mean=2.972, StdDev=4.108
  Overviews: 500x500, 250x250
  Metadata:
    STATISTICS_MAXIMUM=40.566223144531
    STATISTICS_MEAN=2.9720626801483
    STATISTICS_MINIMUM=0
    STATISTICS_STDDEV=4.1079777350157
    STATISTICS_VALID_PERCENT=100
Band 3 Block=256x256 Type=Float32, ColorInterp=Undefined
  Description = tsri
  Min=0.000 Max=1.000
  Minimum=0.000, Maximum=1.000, Mean=0.466, StdDev=0.333
  Overviews: 500x500, 250x250
  Metadata:
    STATISTICS_MAXIMUM=0.99999368190765
    STATISTICS_MEAN=0.46556238474922
    STATISTICS_MINIMUM=4.3117461245856e-06
    STATISTICS_STDDEV=0.33263731194436
    STATISTICS_VALID_PERCENT=100
Band 4 Block=256x256 Type=Float32, ColorInterp=Undefined
  Description = tpi
  Min=-4.774 Max=7.687
  Minimum=-4.774, Maximum=7.687, Mean=-0.000, StdDev=0.349
  Overviews: 500x500, 250x250
  Metadata:
    STATISTICS_MAXIMUM=7.6866292953491
    STATISTICS_MEAN=-4.5027400747145e-05
    STATISTICS_MINIMUM=-4.7744207382202
    STATISTICS_STDDEV=0.34907900787421
    STATISTICS_VALID_PERCENT=100
Band 5 Block=256x256 Type=Float32, ColorInterp=Undefined
  Description = slopemask
  Min=0.000 Max=1.000
  Minimum=0.000, Maximum=1.000, Mean=0.982, StdDev=0.128
  Overviews: 500x500, 250x250
  Metadata:
    STATISTICS_MAXIMUM=1
    STATISTICS_MEAN=0.98238805740667
    STATISTICS_MINIMUM=0
    STATISTICS_STDDEV=0.12799686533701
    STATISTICS_VALID_PERCENT=100
```

3.5.2 Create some parameters for the dynamic tiler.

These parameters will be pased to `titiler_url` for visualization.

Note the values below: `band_min`, `band_max` and `colormap_name` are set as the current defaults for biomass in Mg/hectare for the dashboard. For other datasets, users should inspect the output of the `gdalinfo` for `band_min` and `band_max` and modify the `colormap_name` as makes sense for the current dataset. This notebook includes section on what colormaps are available and how to configure different types of colormaps and legends.

[9]:
```
titiler_url = f"https://titiler.maap-project.org"
band_min = 0
band_max = 400
rescale = f"{band_min},{band_max}"
bidx = 1
colormap_name = 'gist_earth_r'
```

3.5.3 Helper functions

3.5.4 Step 1, Scenario 1: Adding a single COG

1. Upload the file

Only use the following steps if you only have one COG to share to the dashboard. If you want to create a mosaic from multiple COGs, skip to Step 1 Option 2.

Upload the file to S3 and make note of the location.

[10]:
```
def local_to_s3(url):
    ''' A Function to convert local paths to s3 urls'''
    return url.replace("/projects/shared-buckets", "s3://maap-ops-workspace/shared")

location = local_to_s3(f"{project_dir}{my_tif}")
location
```

[10]:
```
's3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/Copernicus_30439_covars_cog_
↪topo_stack.tif'
```

3.5.5 2. Test the COG with the titiler and ipyCMC

[11]:
```
## Build a WMTS call
wmts_url = f"{titiler_url}/cog/WMTSCapabilities.xml"
params = {
    "tile_format": "png",
    "tile_scale": "1",
    "pixel_selection": "first",
    "TileMatrixSetId": "WebMercatorQuad",
    "url": location,
    "bidx": bidx, # Select which band to use
    "resampling_method":"nearest",
    "rescale": rescale,
    "return_mask": "true",
    "colormap_name": colormap_name # Any colormap from matplotlib will work
}

wmts_call = "?".join([wmts_url, urllib.parse.urlencode(params)])

# Note Jupyter bug add amp; incorrectly when printing the url
wmts_call
```

[11]:
```
'https://titiler.maap-project.org/cog/WMTSCapabilities.xml?tile_format=png&tile_
↪scale=1&pixel_selection=first&TileMatrixSetId=WebMercatorQuad&url=s3%3A%2F%2Fmaap-
↪ops-workspace%2Fshared%2Falexdevseed%2Flandsat8%2Fviz%2FCopernicus_30439_covars_cog_
↪topo_stack.tif&bidx=1&resampling_method=nearest&rescale=0%2C400&return_mask=true&
↪colormap_name=gist_earth_r'
```

[12]:
```
w = ipycmc.MapCMC()
w
```

```
MapCMC()
```

[13]:
```
w.load_layer_config(wmts_call, "wmts/xml")
```

3.5.6 Step 1, Scenario 2: Adding data from multiple COGs by creating a mosaic

Many datasets are comprised of many tiles distributed spatially over the globe. In order to visualize them all together, we can use mosaicJSON to create a mosaic for the dynamic tiler API. The dynamic tiler API knows how to read this mosaicJSON and select which tiles to render based on the current zoom, x and y coordinates across spatially distinct COGs.

[14]:
```
%%capture
!pip install cogeo-mosaic
```

[15]:
```
import glob
import os
import urllib

from cogeo_mosaic.mosaic import MosaicJSON
from cogeo_mosaic.backends import MosaicBackend
```

[16]:
```
# Skipping this step since we don't want to upload these large files more than once!
full_path_files = [f'{project_dir}{file}' for file in files]
```

[17]:
```
def local_to_s3(url):
    ''' A Function to convert local paths to s3 urls'''
    return url.replace("/projects/shared-buckets", "s3://maap-ops-workspace/shared")

tiles = [local_to_s3(file) for file in full_path_files]
print(tiles)
```

```
['s3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/Copernicus_30438_covars_cog_
↪topo_stack.tif', 's3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/
↪Copernicus_30439_covars_cog_topo_stack.tif', 's3://maap-ops-workspace/shared/
↪alexdevseed/landsat8/viz/Copernicus_30440_covars_cog_topo_stack.tif', 's3://maap-
↪ops-workspace/shared/alexdevseed/landsat8/viz/Copernicus_30542_covars_cog_topo_
↪stack.tif', 's3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/Copernicus_
↪30543_covars_cog_topo_stack.tif', 's3://maap-ops-workspace/shared/alexdevseed/
↪landsat8/viz/Copernicus_30821_covars_cog_topo_stack.tif', 's3://maap-ops-workspace/
↪shared/alexdevseed/landsat8/viz/Copernicus_30822_covars_cog_topo_stack.tif', 's3://
↪maap-ops-workspace/shared/alexdevseed/landsat8/viz/Copernicus_30823_covars_cog_topo_
↪stack.tif', 's3://maap-ops-workspace/shared/alexdevseed/landsat8/viz/Landsat8_30542_
↪comp_cog_2015-2020_dps.tif', 's3://maap-ops-workspace/shared/alexdevseed/landsat8/
↪viz/Landsat8_30543_comp_cog_2015-2020_dps.tif', 's3://maap-ops-workspace/shared/
↪alexdevseed/landsat8/viz/Landsat8_30822_comp_cog_2015-2020_dps.tif', 's3://maap-ops-
↪workspace/shared/alexdevseed/landsat8/viz/Landsat8_30823_comp_cog_2015-2020_dps.tif
↪']
```

[18]:
```
mosaicdata = MosaicJSON.from_urls(tiles, minzoom=1, maxzoom=16)
```

3.5.7 Using mosaicJSON with titiler

There are 2 options for using mosaicJSON with titiler:

1. (Preferred) Post mosaicJSON to titiler `mosaics` endpoint and use the `mosaicjson/mosaics` endpoint for dynamic tiling.
2. Upload mosaicJSON to S3 and pass the S3 url to the titiler `mosaicjson/tiles` endpoint.

3.5.8 Post mosaicJSON to titiler

[19]:
```
import requests
from pprint import pprint

r = requests.post(
    url=f"{titiler_url}/mosaics",
    headers={
        "Content-Type": "application/vnd.titiler.mosaicjson+json",
    },
    json=mosaicdata.dict(exclude_none=True)).json()

pprint(r)
```

```
{'id': 'a1cc4903-5810-4b38-a637-05c4d10edb75',
 'links': [{'href': 'https://titiler.maap-project.org/mosaics/a1cc4903-5810-4b38-a637-
↪05c4d10edb75',
            'rel': 'self',
            'title': 'Self',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/a1cc4903-5810-4b38-a637-
↪05c4d10edb75/mosaicjson',
            'rel': 'mosaicjson',
            'title': 'MosaicJSON',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/a1cc4903-5810-4b38-a637-
↪05c4d10edb75/tilejson.json',
            'rel': 'tilejson',
            'title': 'TileJSON',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/a1cc4903-5810-4b38-a637-
↪05c4d10edb75/tiles/{z}/{x}/{y}',
            'rel': 'tiles',
            'title': 'Tiles',
            'type': 'application/json'},
           {'href': 'https://titiler.maap-project.org/mosaics/a1cc4903-5810-4b38-a637-
↪05c4d10edb75/WMTSCapabilities.xml',
            'rel': 'wmts',
            'title': 'WMTS',
            'type': 'application/json'}]}
```

[20]:
```
mosaic_link = list(filter(lambda x: x['rel'] == 'tiles', r['links']))[0]['href']
wmts_mosaic_link = list(filter(lambda x: x['rel'] == 'wmts', r['links']))[0]['href']
```

3.5.9 Test your mosaic

[21]:
```
params = {
    "tile_format": "png",
    "bidx": bidx, # Select which band to use
    "resampling_method":"nearest",
    "rescale": rescale,
    "return_mask": "true",
    "colormap_name": colormap_name # Any colormap from matplotlib will work
}

wmts_call = "?".join([wmts_mosaic_link, urllib.parse.urlencode(params)])

w.load_layer_config(wmts_call, "wmts/xml")
```

By default, the image will be displayed in greyscale if no `colormap_name` parameter is passed to the titiler API. Guidance below is provided to help determine what a valid colormap_name might be and how to create a legend for the dashboard.

Dashboard ColorRamps & Legends

When using the dashboard, there 2 components for implementing a color scheme for your map. There is the map render and there is the legend.

> Titiler used for Cloud Optimized Geotiff (COG) rendering accepts any color scheme from the python matplotlib library, and custom color formulas.

- Rio Tiler Colors
- Matplotlib Colors

Available `colormap_name` values for titiler: `above, accent, accent_r, afmhot, afmhot_r, autumn, autumn_r, binary, binary_r, blues, blues_r, bone, bone_r, brbg, brbg_r, brg, brg_r, bugn, bugn_r, bupu, bupu_r, bwr, bwr_r, cfastie, cividis, cividis_r, cmrmap, cmrmap_r, cool, cool_r, coolwarm, coolwarm_r, copper, copper_r, cubehelix, cubehelix_r, dark2, dark2_r, flag, flag_r, gist_earth, gist_earth_r, gist_gray, gist_gray_r, gist_heat, gist_heat_r, gist_ncar, gist_ncar_r, gist_rainbow, gist_rainbow_r, gist_stern, gist_stern_r, gist_yarg, gist_yarg_r, gnbu, gnbu_r, gnuplot, gnuplot2, gnuplot2_r, gnuplot_r, gray, gray_r, greens, greens_r, greys, greys_r, hot, hot_r, hsv, hsv_r, inferno, inferno_r, jet, jet_r, magma, magma_r, nipy_spectral, nipy_spectral_r, ocean, ocean_r, oranges, oranges_r, orrd, orrd_r, paired, paired_r, pastel1, pastel1_r, pastel2, pastel2_r, pink, pink_r, piyg, piyg_r, plasma, plasma_r, prgn, prgn_r, prism, prism_r, pubu, pubu_r, pubugn, pubugn_r, puor, puor_r, purd, purd_r, purples, purples_r, rainbow, rainbow_r, rdbu, rdbu_r, rdgy, rdgy_r, rdpu, rdpu_r, rdylbu, rdylbu_r, rdylgn, rdylgn_r, reds, reds_r, rplumbo, schwarzwald, seismic, seismic_r, set1, set1_r, set2, set2_r, set3, set3_r, spectral, spectral_r, spring, spring_r, summer, summer_r, tab10, tab10_r, tab20, tab20_r, tab20b, tab20b_r, tab20c, tab20c_r, terrain, terrain_r, twilight, twilight_r, twilight_shifted, twilight_shifted_r, viridis, viridis_r, winter, winter_r, wistia, wistia_r, ylgn, ylgn_r, ylgnbu, ylgnbu_r, ylorbr, ylorbr_r, ylorrd, ylorrd_r`

Example 1: Class based known colors

In this example, the raster represents classes of forest with 11 possible values. There are specific colors selected to correspond to each class. We combine the list of colors and the list of classes and format them for the legend parameter the dashboard needs.

https://github.com/MAAP-Project/dashboard-datasets-maap/blob/main/datasets/taiga-forest-classification.json

[ ]:
```
colors = [
    '#5255A3','#1796A3','#FDBF6F','#FF7F00', '#FFFFBF','#D9EF8B','#91CF60','#1A9850',
↪'#C4C4C4','#FF0000', '#0000FF'
]
labels = [
    'Sparse & Uniform',
    'Sparse & Diffuse-gradual',
    'Sparse & Diffuse-rapid',
    'Sparse & Abrupt ',
    'Open & Uniform ',
    'Open & Diffuse-gradual',
    'Open & Diffuse-rapid',
    'Open & Abrupt',
    'Intermediate & Closed',
    'Non-forest edge (dry)',
    'Non-forest edge (wet)'
]

legend = [dict(color=colors[i],label=labels[i]) for i in range(0, len(colors))]
print(json.dumps(legend, indent=2))

# Copy and Paste the output below to your dashboard config.
```

Example 2: Discrete ColorRamp

In this example, the range of values is known, but the color scale has many non-sequential colors. Starting with the premade color list, we create a continuous color ramp that uses the known colors as stops points. Arbitrarly 12 breaks looked decent in the dashboard legend so we split it into 12 discrete colors. Then combine the list of values and colors into the correct json syntax.

https://github.com/MAAP-Project/dashboard-datasets-maap/blob/main/datasets/ATL08.json

[ ]:
```
forest_ht = matplotlib.colors.LinearSegmentedColormap.from_list('forest_ht', ['#636363
↪','#FC8D59','#FEE08B','#FFFFBF','#D9EF8B','#91CF60','#1A9850','#005A32'], 12)
cols = [matplotlib.colors.to_hex(forest_ht(i)) for i in range(forest_ht.N)]

cats = range(0,25, (25//len(cols)))
legend = [[cats[i],cols[i]] for i in range(0, len(cols))]
text = (json.dumps(legend, separators=(',', ': ') ))

print(text.replace('],[',']\n['))

# Copy and Paste the output below to your dashboard config.
```

Example 3: Continuous ColorRamp

In this example, we are using a built in ColorRamp from matplotlib. So we just need to extract enough colors to fill the legend adequately, and convert the colors to hex codes.

https://github.com/MAAP-Project/dashboard-datasets-maap/blob/main/datasets/topo.json

[ ]:
```
cmap_name = 'gist_earth_r'
cmap = matplotlib.cm.get_cmap(cmap_name, 12)
cols = [matplotlib.colors.to_hex(cmap(i)) for i in range(cmap.N)]
print(cols)

# Copy and Paste the output below to your dashboard config.
```

3.5.10 Step 3: Create and submit your dashboard dataset json

[25]:
```
# This example is for a continuous color ramps
dataset_id = "paraguay-estimated-biomass"
dataset_name = "Estimated Biomass in Paraguay"
dataset_type = "raster"
legend_type = "gradient-adjustable"
info = "Estimated biomass within 6km grids."
stops = cols
parameters = f"colormap_name={cmap_name}&rescale={band_min},{band_max}&bidx={bidx}"
```

[26]:
```
# Single COG
tiles_link = f"{{titiler_server_url}}/cog/tiles/{{z}}/{{x}}/{{y}}.png?url={location}&
↪{parameters}"
```

[ ]:
```
tiles_link
```

[ ]:
```
dataset_dict = {
    "id": dataset_id,
    "name": dataset_name,
    "type": dataset_type,
    "swatch": {
      "color": "#6976d7",
      "name": "Moody Blue"
    },
    "source": {
        "type": dataset_type,
        "tiles": [ tiles_link ]
    },
    "legend": {
      "type": legend_type,
       "min": band_min,
       "max": band_max,
      "stops": stops
    },
    "info": info
}
print(json.dumps(dataset_dict, indent=4))
```

3.5.11 Create a PR to the datasets repo

```
git clone git@github.com:MAAP-Project/biomass-dashboard-datasets.git
cd biomass-dashboard-datasets
git checkout -b ab/add-dataset
```

(continues on next page)

(continued from previous page)

```
# select and copy json above
echo <copied_json> >> datasets/paraguay-estimated-biomass.json
```

Add to the datasets list in config.yml

In `config.yml`:

```
DATASETS:
- paraguay-estimated-biomass.json
```

Add to the product or country pilot

In `country_pilots/paraguay/country_pilot.json`:

```
{
    "id": "paraguay",
    "label": "Paraguay",
    //...
    "datasets": [
        {
            "id": "paraguay-forest-mask"
        },
        {
            "id": "paraguay-tree-cover"
        },
        {
            "id": "paraguay-estimated-biomass"
        }
    ]
}
```

3.5.12 Add content to the summary

There should be a `summary.html` file corresponding to the product or country pilot you are working on, for example: `country_pilots/paraguay/summary.html`. Add or modify content in that file as appropriate.

3.5.13 Submit a PR

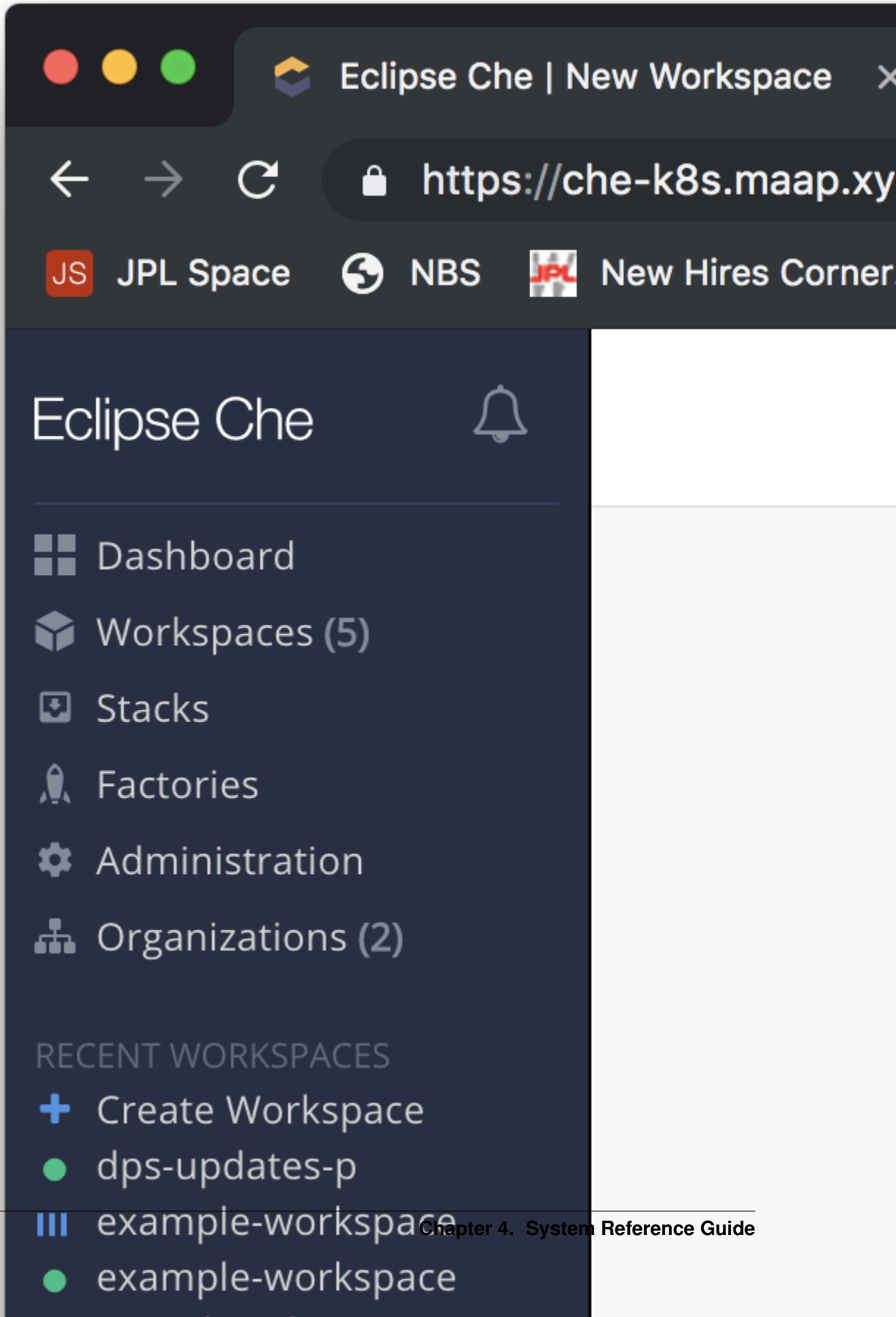
Once you have added the dataset json file and summary content, submit a PR to https://github.com/MAAP-Project/biomass-dashboard-datasets. A member of the data team will review the PR and when it is merged your content will appear in biomass.dit.maap-project.org.

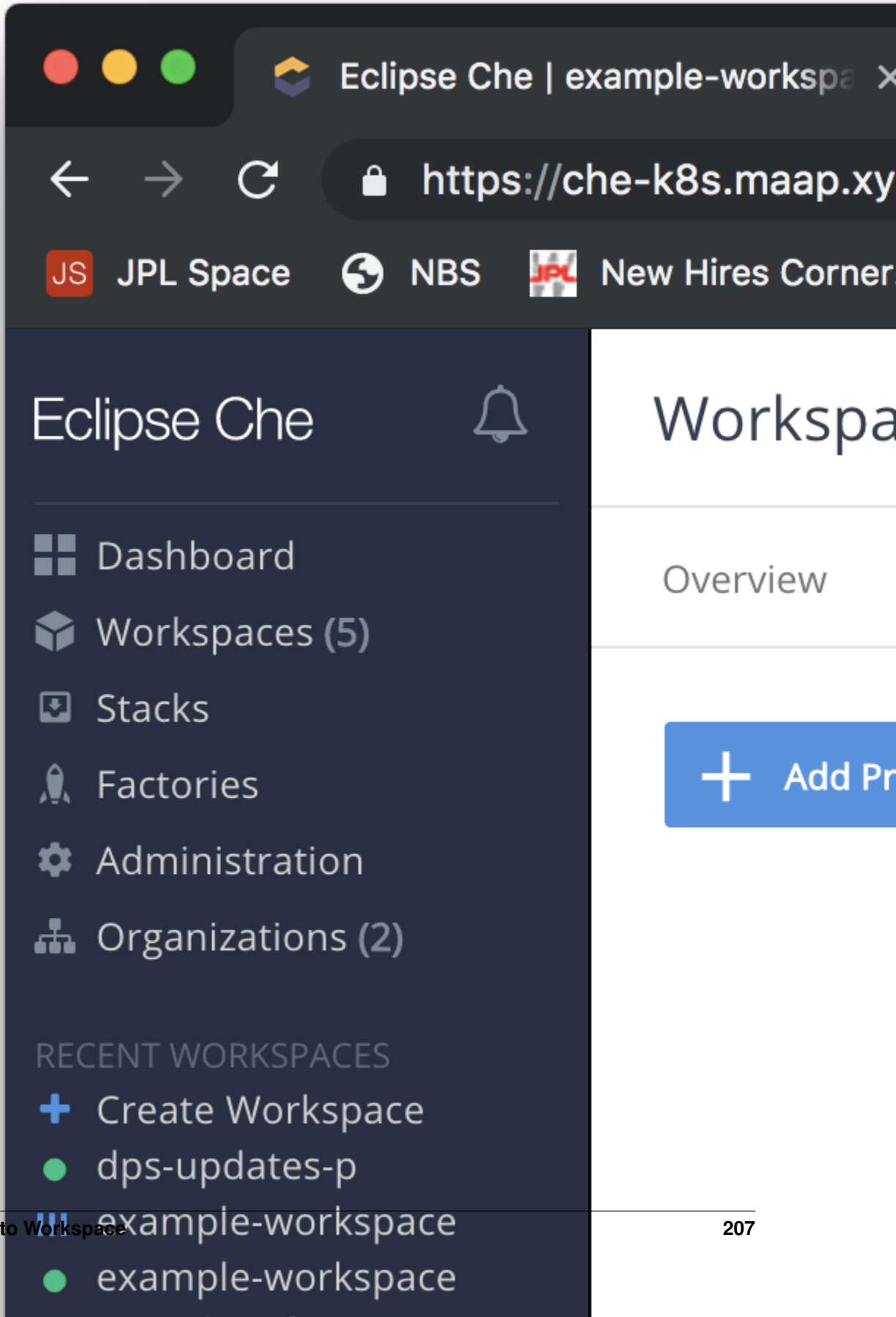
CHAPTER 4

System Reference Guide

4.1 Add Project to Workspace

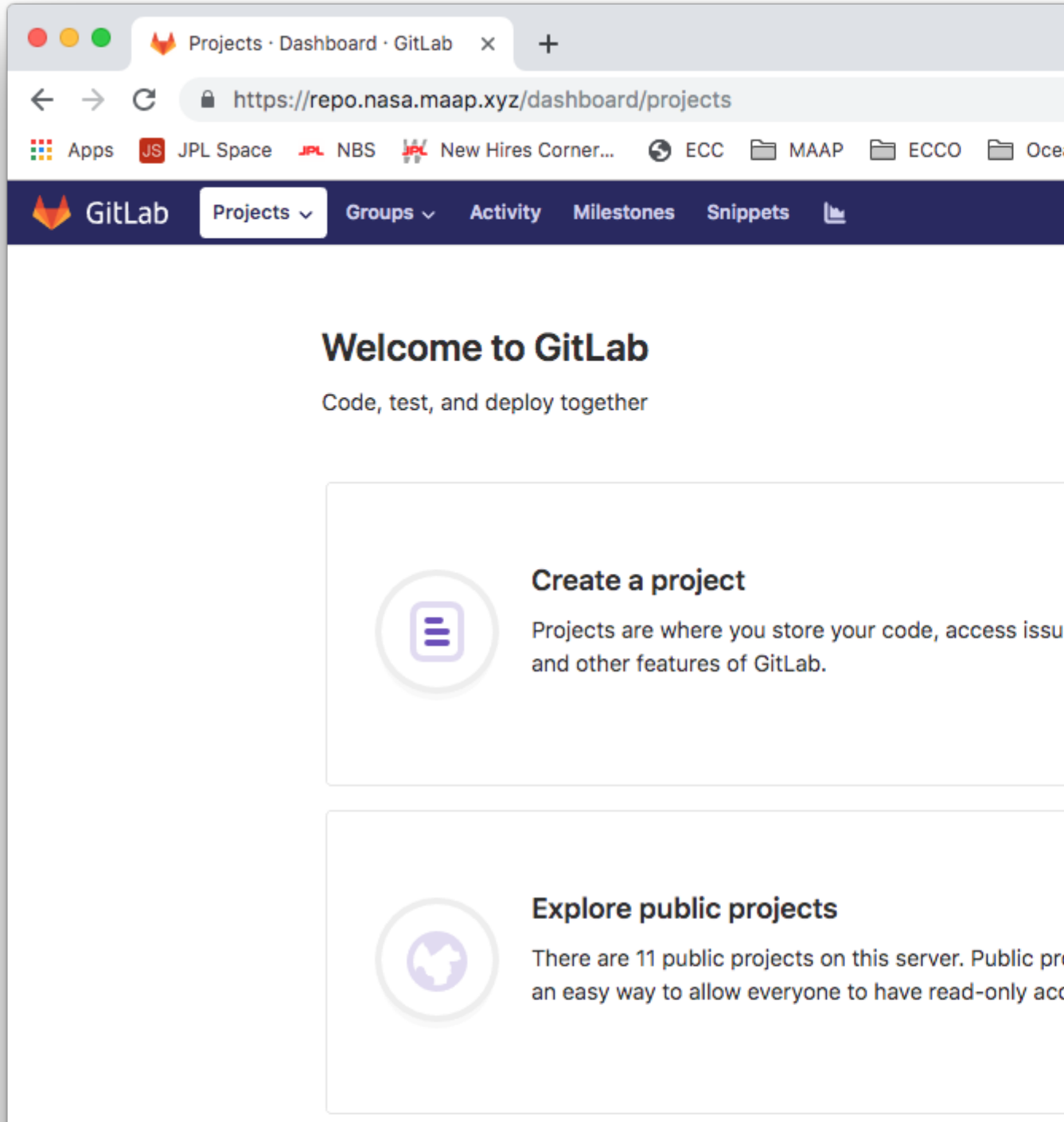
Projects can be added to a workspace during or after workspace creation. If a project is added after a workspace exists, the workspace must be restarted when the projects are added. After restart, your projects will automatically appear in your workspace.


Eclipse Che | New Workspace
https://che-k8s.maap.xy
JS JPL Space
NBS
New Hires Corner
Eclipse Che
Dashboard
Workspaces (5)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
dps-updates-p
example-workspace
example-workspace


Eclipse Che | example-workspa
https://che-k8s.maap.xy
JS JPL Space
NBS
New Hires Corner
Eclipse Che
Dashboard
Workspaces (5)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
dps-updates-p
example-workspace
example-workspace
Workspa
Overview
Add Pr

4.2 Create a Project

The first step in getting started with projects in MAAP is going to MAAP's GitLab and creating a new project. When you create a MAAP account, an account is automatically configured for you on MAAP's GitLab instance. To sign in, click on `Sign in with CAS`, then use your URS or ESA login. This will direct you to your projects.

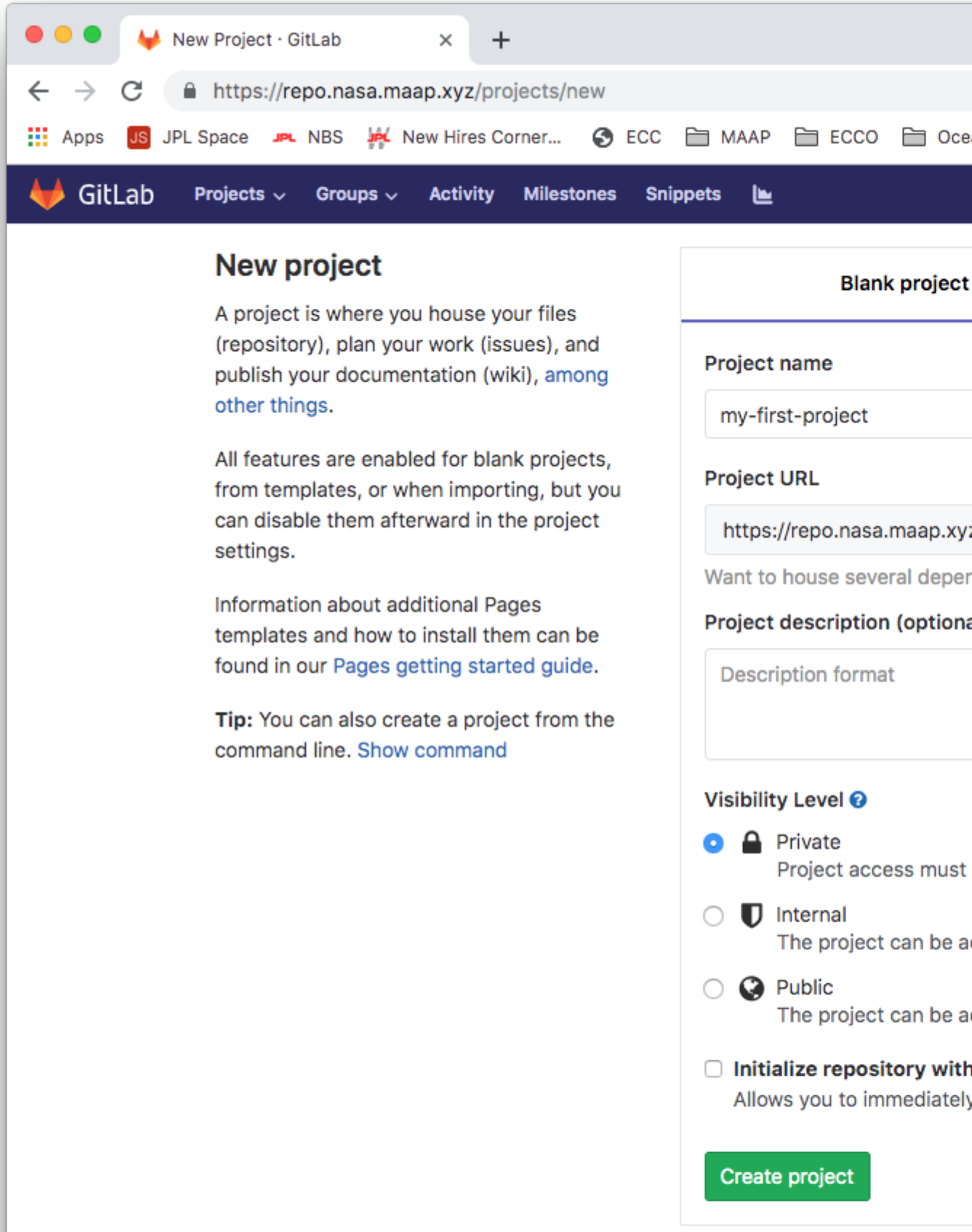

Projects · Dashboard · GitLab
https://repo.nasa.maap.xyz/dashboard/projects
Apps
JPL Space
NBS
New Hires Corner...
ECC
MAAP
ECCO
GitLab
Projects
Groups
Activity
Milestones
Snippets
Welcome to GitLab
Code, test, and deploy together
Create a project
Projects are where you store your code, access issu
and other features of GitLab.
Explore public projects
There are 11 public projects on this server. Public pr
an easy way to allow everyone to have read-only acc

Click on `Create a Project`. Name your project and select its visibility level.

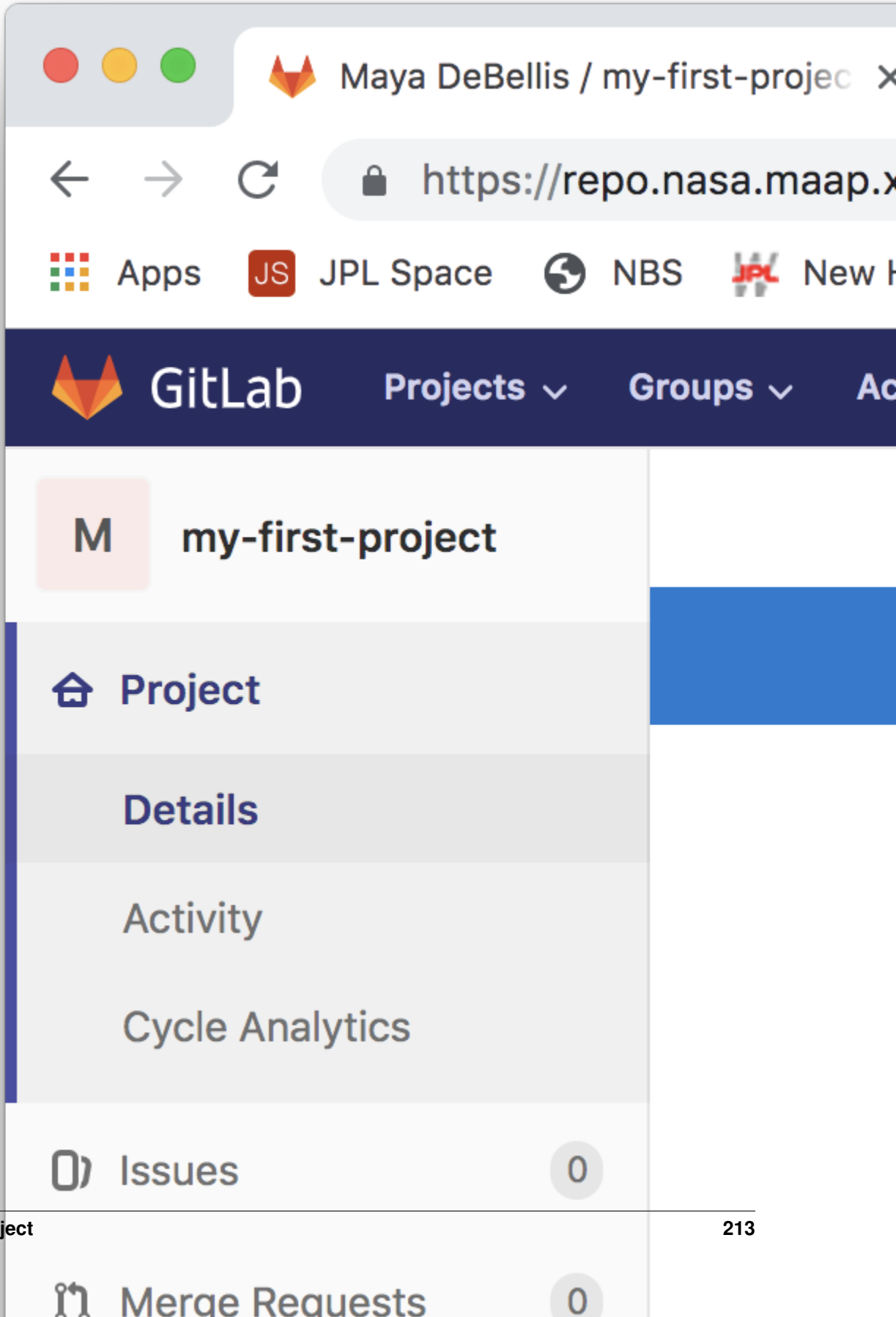
NOTE: Projects will default to private, meaning only you can see it, and if you want to allow another user to contribute, you will have to explicitly add them in the GitLab interface. Adding a user to a workspace that has a project in it does not automatically give them access to your GitLab project, you must also add them to the private project if you want to allow them to make changes.

Projects set to internal can be seen by any authenticated MAAP user. Projects set to public can be seen by anyone, regardless of whether they are a MAAP user or not.

Settings for your new project:

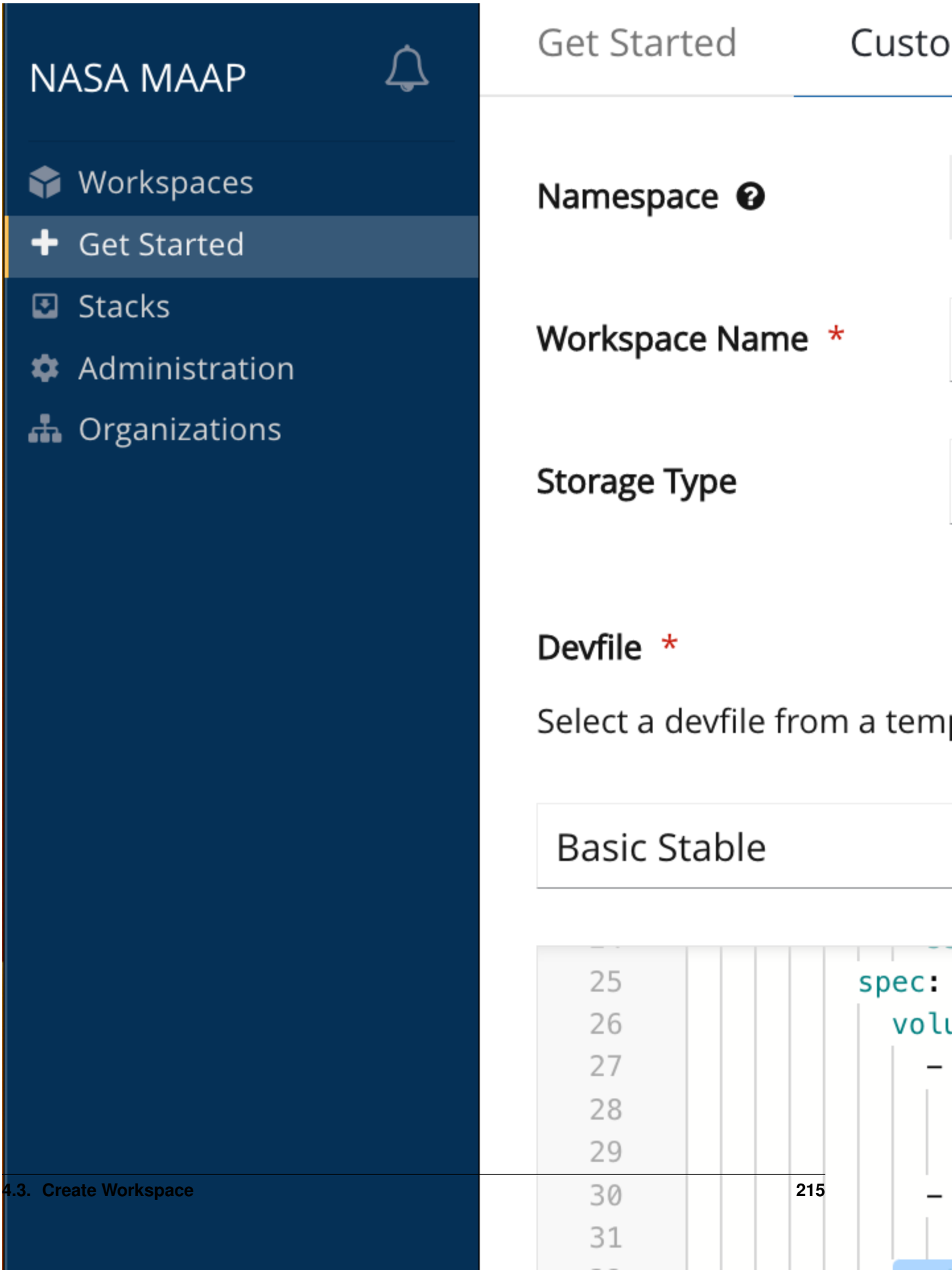

New Project · GitLab
https://repo.nasa.maap.xyz/projects/new
Apps
JPL Space
NBS
New Hires Corner...
ECC
MAAP
ECCO
GitLab
Projects
Groups
Activity
Milestones
Snippets
New project
A project is where you house your files (repository), plan your work (issues), and publish your documentation (wiki), among other things.
All features are enabled for blank projects, from templates, or when importing, but you can disable them afterward in the project settings.
Information about additional Pages templates and how to install them can be found in our Pages getting started guide.
Tip: You can also create a project from the command line. Show command
Blank project
Project name
my-first-project
Project URL
Want to house several deper
Project description (optiona
Description format
Visibility Level
Private
Project access must
Internal
The project can be a
Public
The project can be a
Initialize repository with
Allows you to immediately
Create project

Once you have created a project, copy the https url for the project that can be found under the clone button's dropdown. You will use this to add your project to your Che workspace.


Maya DeBellis / my-first-projec
https://repo.nasa.maap.x
Apps
JS
JPL Space
NBS
New
GitLab
Projects
Groups
Ac
M
my-first-project
Project
Details
Activity
Cycle Analytics
Issues
0
Merge Requests
0

4.3 Create Workspace

Navigate to the side panel tab `Workspaces` and click `Add Workspace`. On the Get Started tab, you can select a stack to get started quickly. If you would like a more customized approach, on the Custom Workspace tab you can name your workspace, select the storage type, and select a devfile template from available stacks or upload your own devfile. Here, you can edit the devfile under `containers/jupyter/resources/limits/memory` to alter the memory limit for your workspace.


NASA MAAP
Workspaces
Get Started
Stacks
Administration
Organizations
Get Started
Custo
Namespace
Workspace Name *
Storage Type
Devfile *
Select a devfile from a temp
Basic Stable
25
26
27
28
29
30
31
spec:
volu

If you want to edit the devfile or your workspace's memory limit after its creation, just click on the `Workspaces` side panel tab, then the workspace, then the Devfile tab at the top of the page.

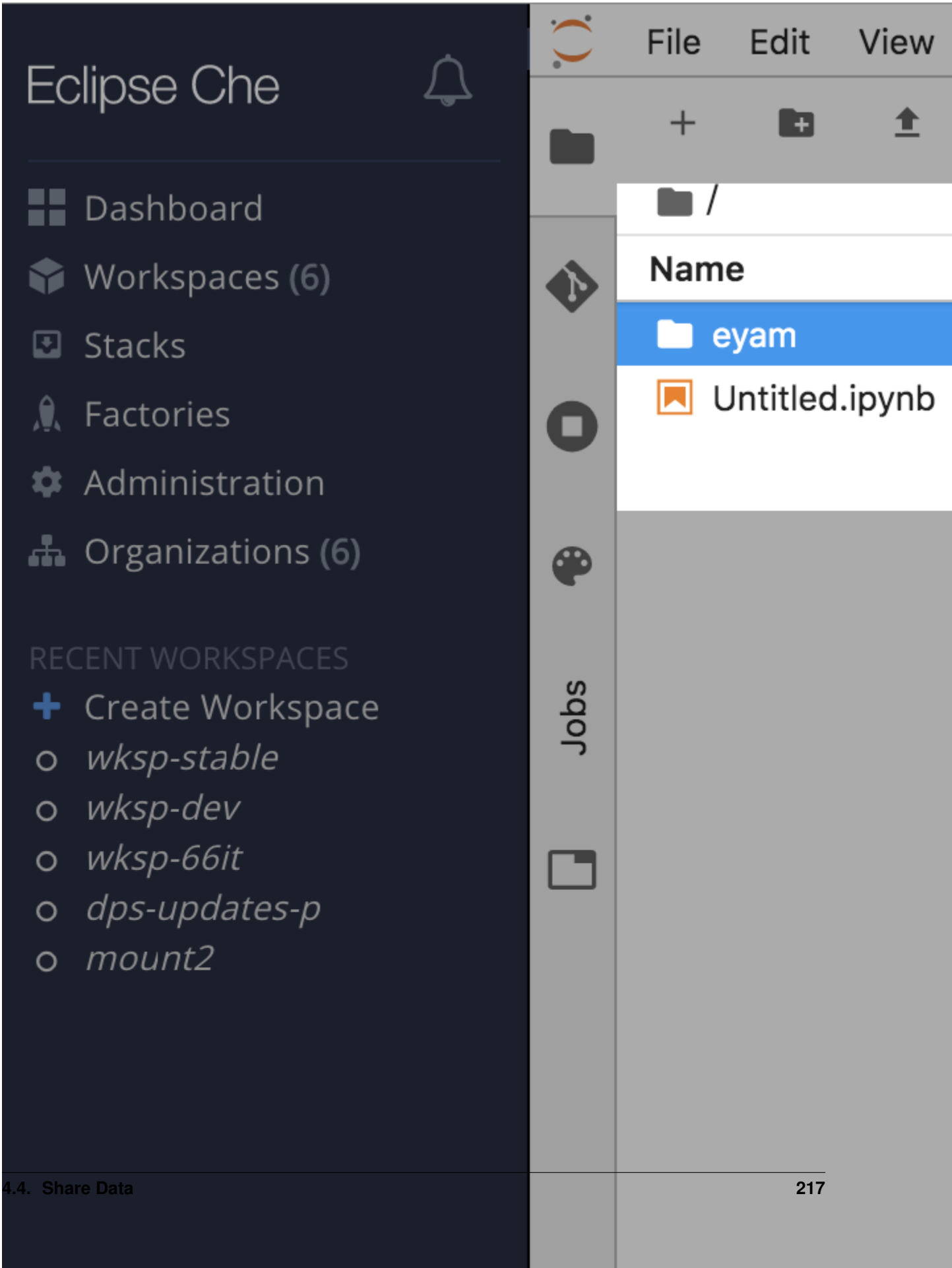
DISCLAIMER: MAAP might change the way that we do workspace memory limits in the future, including restricting the memory that a user can allocate for their workspace. For now, as a courtesy to other users only increase the memory limit if your kernel keeps running out of memory and crashing. If you're not sure why your kernel is crashing after increasing your workspace's memory limit, please contact the development team. We recommend using a memory limit of 16GB (which means altering the devfile to `memory: 15258Mi`).

4.4 Share Data

Users who have access to a workspace have access to all the files contained in that workspace.

All users have their own personal s3-bucket folder mounted in the FileBrowser home In `/projects`, each user has a personal s3-hosted folder with the same name as their CAS username. Files in this folder are automatically uploaded to s3 and are accessible from any workspace a user signs into.

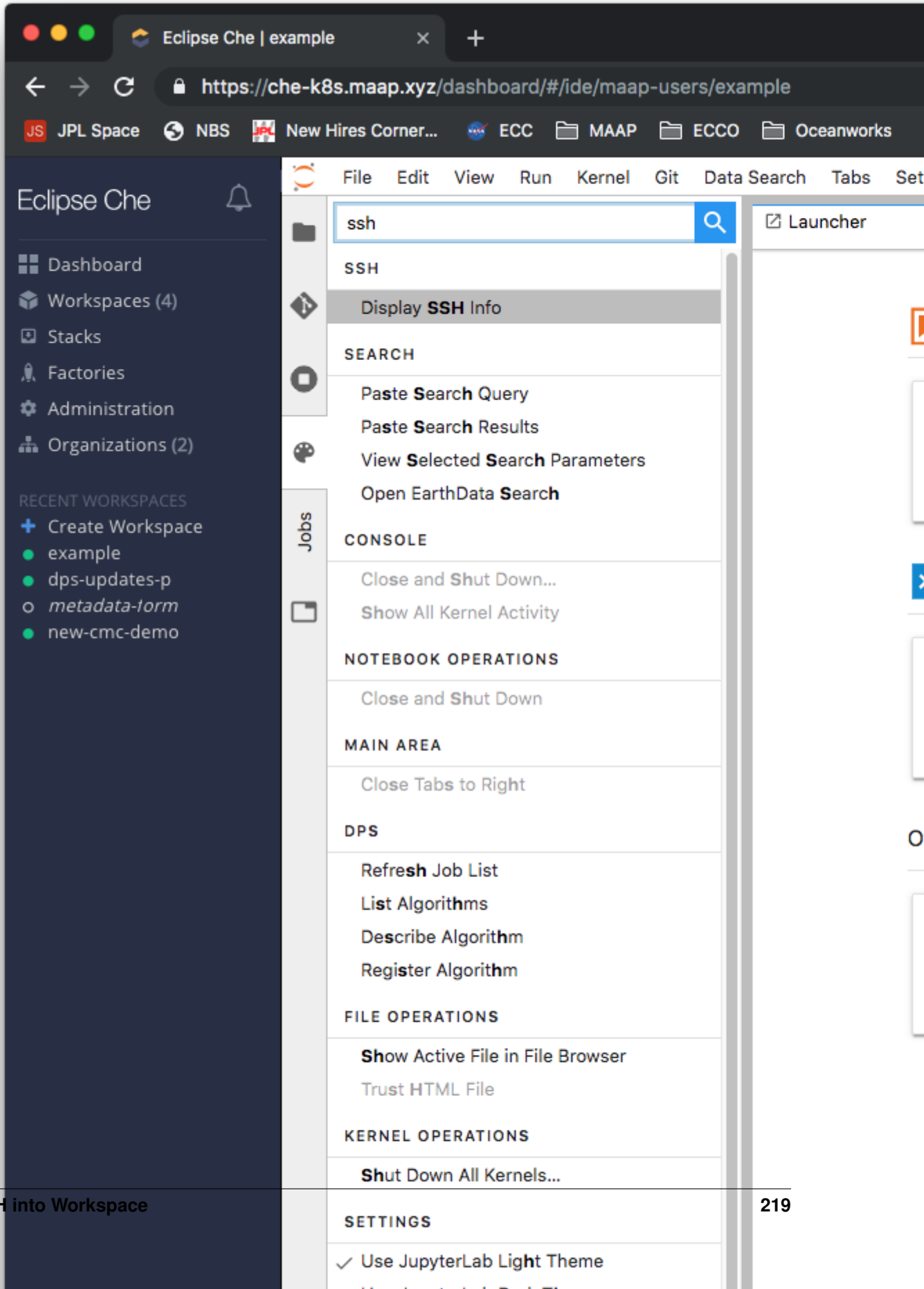
The intention of this mounted folder is that you can use this to share data with others, and also store files you want to access across different workspaces. It is not intended that you do all of your work in this directory. Because this directory is mounted to s3, you will notice that processes are slower when working in this directory.

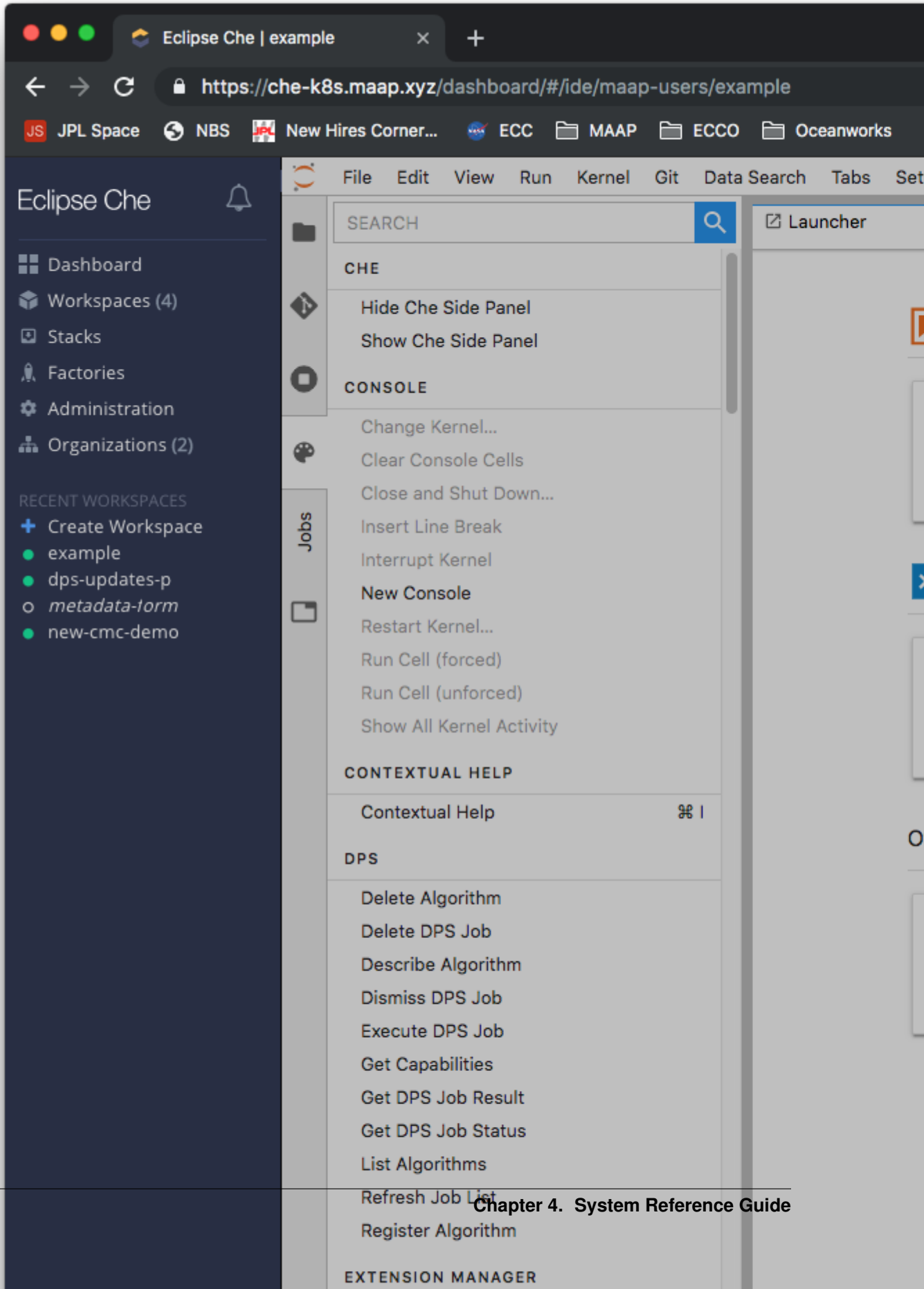

Eclipse Che
Dashboard
Workspaces (6)
Stacks
Factories
Administration
Organizations (6)
RECENT WORKSPACES
Create Workspace
wksp-stable
wksp-dev
wksp-66it
dps-updates-p
mount2
File
Edit
View
/
Name
eyam
Untitled.ipynb
Jobs

Users can create a shareable link for any files in their folder that is hosted on s3. To do this, go to `Command Palette -> User -> Get Presigned S3 URL` and enter the relative path to the file you want to share. The link will expire after 12 hours.

4.5 SSH into Workspace

As an alternative to using the jupyter interface, you can SSH directly into the container that your workspace set up. In order to get the IP and port information, navigate to the command palette of the jupyter interface. Find the command `Display SSH Info`, which will display the information you need (you can easily search for SSH). Your public SSH key that you added to your account will be added to any workspace you create. If you did not upload an SSH key to your profile, you will not be able to SSH in and must use the jupyter interface.


Eclipse Che | example
https://che-k8s.maap.xyz/dashboard/#/ide/maap-users/example
JPL Space
NBS
New Hires Corner...
ECC
MAAP
ECCO
Oceanworks
Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
example
dps-updates-p
metadata-torm
new-cmc-demo
File
Edit
View
Run
Kernel
Git
Data Search
Tabs
Launcher
ssh
SSH
Display SSH Info
SEARCH
Paste Search Query
Paste Search Results
View Selected Search Parameters
Open EarthData Search
CONSOLE
Close and Shut Down...
Show All Kernel Activity
NOTEBOOK OPERATIONS
Close and Shut Down
MAIN AREA
Close Tabs to Right
DPS
Refresh Job List
List Algorithms
Describe Algorithm
Register Algorithm
FILE OPERATIONS
Show Active File in File Browser
Trust HTML File
KERNEL OPERATIONS
Shut Down All Kernels...
SETTINGS
Use JupyterLab Light Theme
Jobs


Eclipse Che | example
https://che-k8s.maap.xyz/dashboard/#/ide/maap-users/example
JPL Space
NBS
New Hires Corner...
ECC
MAAP
ECCO
Oceanworks
Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
example
dps-updates-p
metadata-form
new-cmc-demo
File
Edit
View
Run
Kernel
Git
Data Search
Tabs
SEARCH
Launcher
CHE
Hide Che Side Panel
Show Che Side Panel
CONSOLE
Change Kernel...
Clear Console Cells
Close and Shut Down...
Insert Line Break
Interrupt Kernel
New Console
Restart Kernel...
Run Cell (forced)
Run Cell (unforced)
Show All Kernel Activity
CONTEXTUAL HELP
Contextual Help
DPS
Delete Algorithm
Delete DPS Job
Describe Algorithm
Dismiss DPS Job
Execute DPS Job
Get Capabilities
Get DPS Job Result
Get DPS Job Status
List Algorithms
Refresh Job List
Register Algorithm
EXTENSION MANAGER
Jobs

4.5.1 Accessing MAAP workspaces over SSH

If you would like to have your MAAP workspace mounted (via SSH) on your local computer, follow these steps. It is completely optional and if you do not know why you would want this, feel free to skip this part of the Getting Started Guide.

Basically, this works the same as on any unix-based system: You need to add your Public SSH key from your computer/laptop to the `~/.ssh/authorized_keys` file in MAAP, as described below.

Generate a Public SSH Key

First, you need an SSH key for your personal machine. Here are example instructions for creating an SSH key. Note that there are different instructions for Windows and unix-based systems (MacOS, Linux). The outcome of this will be a Public SSH key on your local computer.

Add the Public SSH Key to MAAP

By default, you cannot see or edit hidden files in the MAAP Jupyter account. Therefore, you should do the following workaround:

1. On your local computer, copy the contents of the public SSH key you just created. It will usually have a `.pub` extension (the default file is `~/.ssh/id_rsa.pub`) and its contents will look something like this (one long line):

   ```
   ssh-rsa AAAAB3Nza[...]TeDx1 ashiklom@gs610-ashiklom.ndc.nasa.gov
   ```

2. Use the Jupyter interface (right click in the file browser) to create a new folder in your home directory called `ssh`.
3. Inside that folder, use the Jupyter interface (right click in the file browser) to create a new file called `authorized_keys` (NOTE: no file extension!).
4. Double click the file you just created to open it in the editor.
5. Copy the contents of your SSH key from your computer into this file. Save it and close it.
6. Return to your home directory, right click the `ssh` folder, click "Rename", and rename it to `.ssh` (period in front). You will get a 404 error, but the operation will have succeeded, and folder will 'disappear'.
   - To check that this worked, in a terminal window, run `ls -a` in your home directory. You should see the `.ssh` folder in there.
7. Repeat these steps to add more SSH keys from different machines.
   - Note that, to see the `.ssh` folder from JupyterLab, you'll need to temporarily rename it to `ssh`. The only way to do this is via the terminal — `mv .ssh ssh`. Don't forget to rename it back to `.ssh` when you're done!

To access a MAAP workspace over SSH, you will need a workspace-specific IP address and port. You can find this by going to the JupyterLab commands menu — the 4th button in the left panel of the Jupyter interface; shows a looking glass over a list; alternatively, press Control + Shift + C (Command + Shift + C on Mac) — and search for "SSH". You'll see "Display SSH Info". Click this button. A dialog box with the correct SSH command will appear.

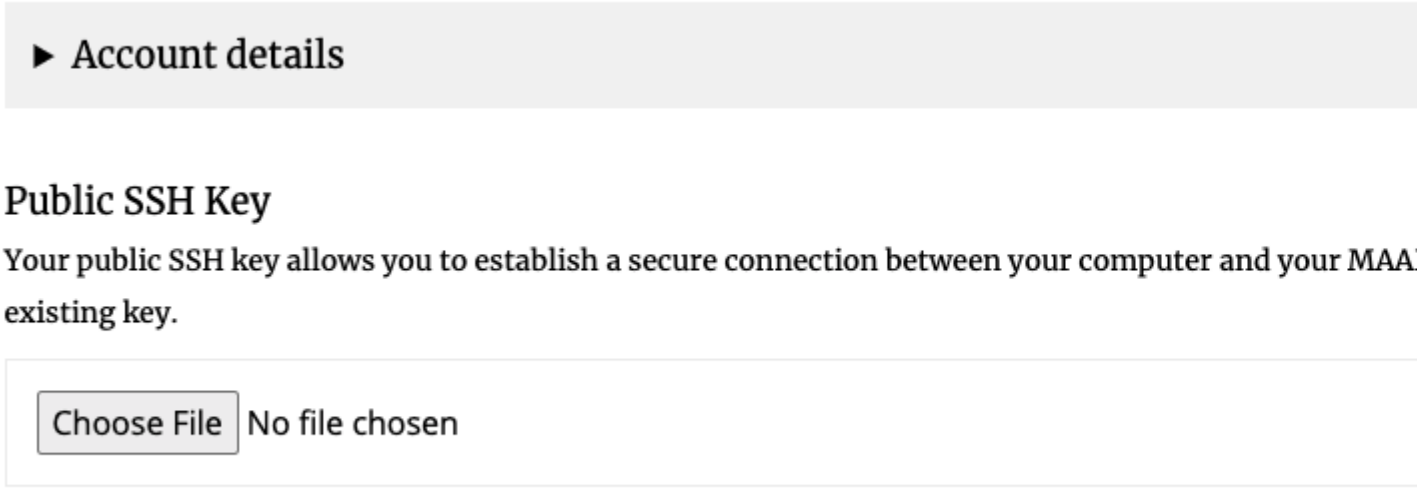
Uploading your public SSH key

Is this an alternative way of doing the above? Or does this even work?

In order to access your ADE workspaces using SSH, you'll need to upload your public SSH key to your MAAP profile using the MAAP portal at https://ops.maap-project.org.

Click on your profile name (or the "Login" button) in the top right corner shown here:

On your profile page, click the "Choose File" button to upload your key.



After uploading your key, the SSH connection will be enabled **after your next** login into the ADE.

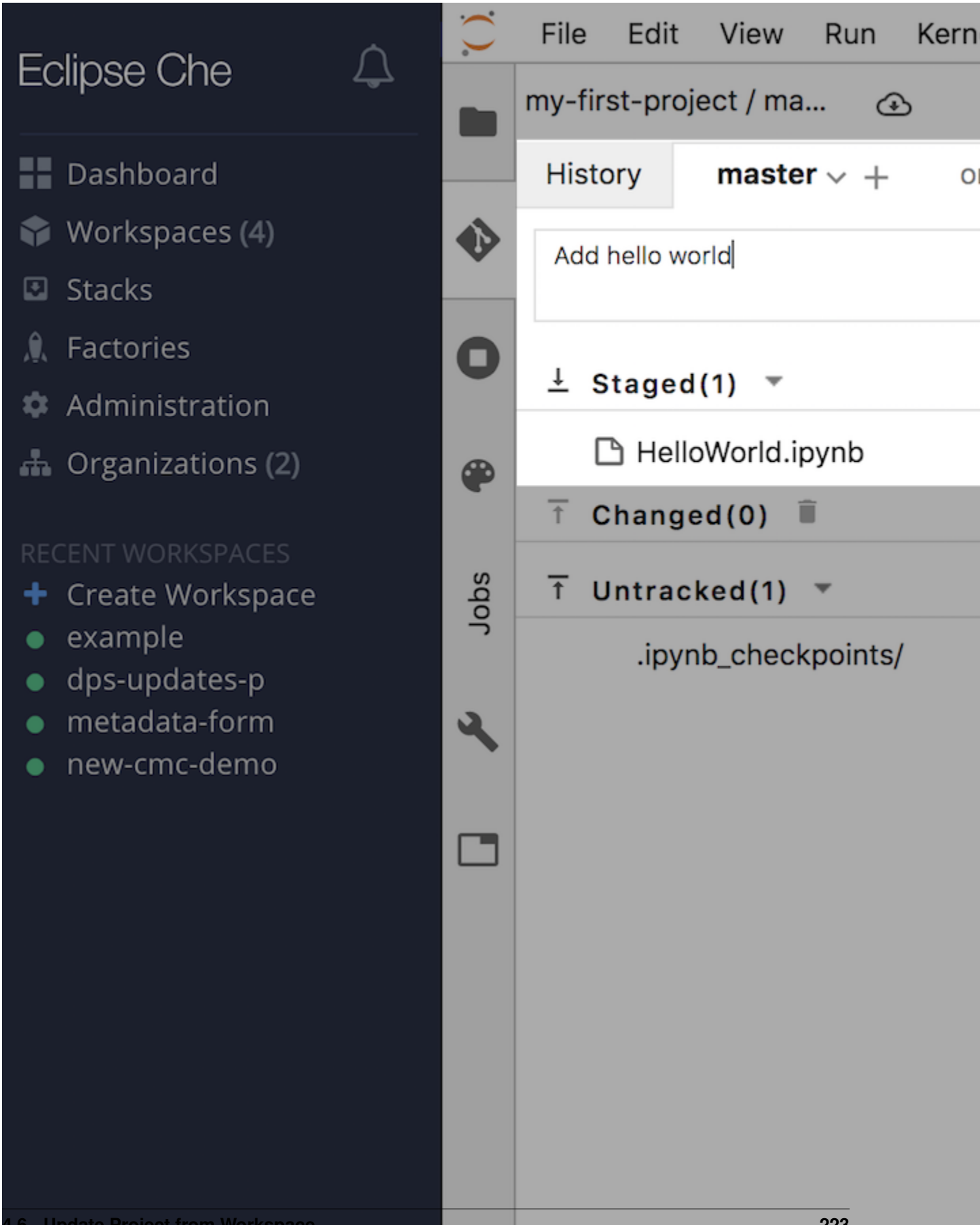
4.6 Update Project from Workspace

All projects imported from GitLab can be found in `/projects`.

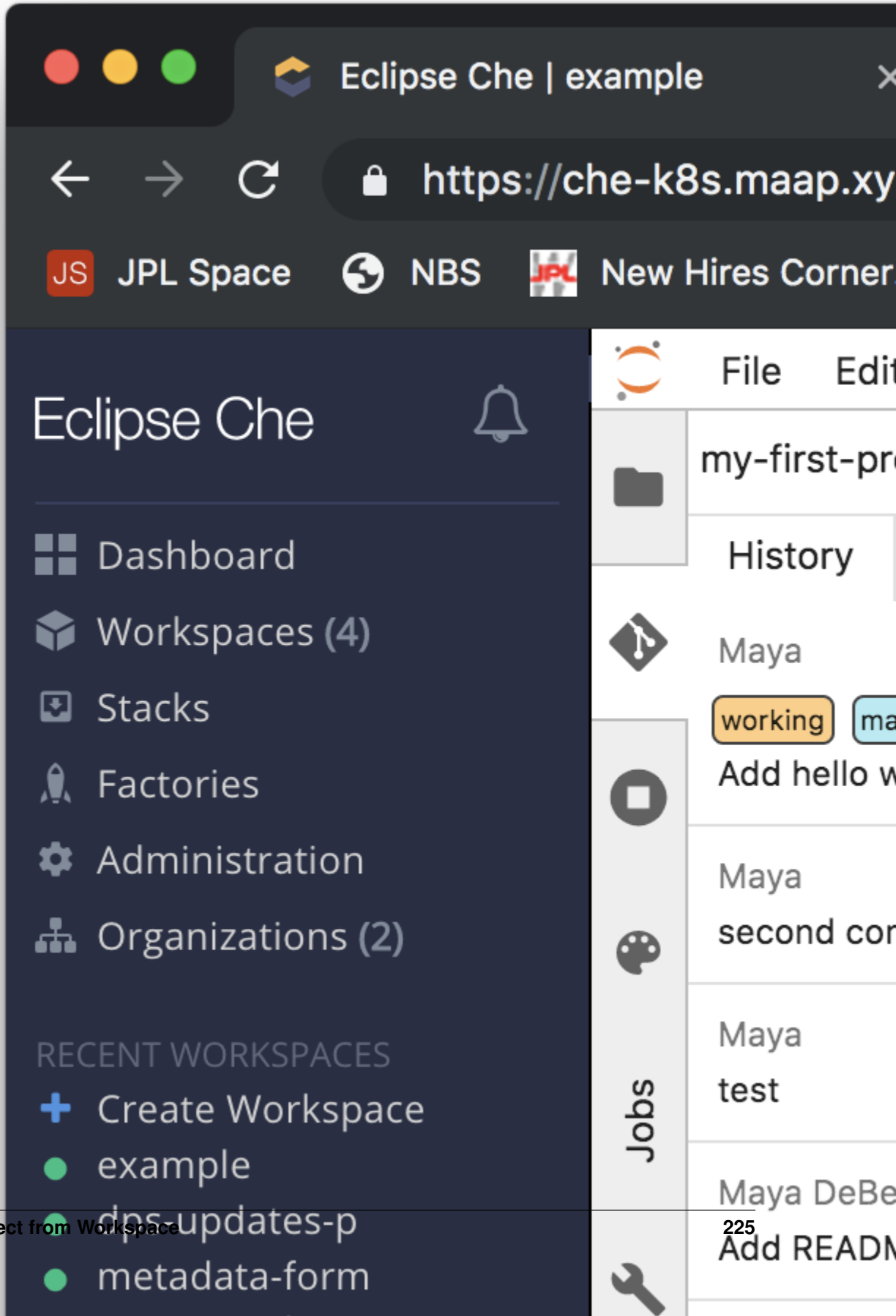
Using the left side panel in the jupyter interface, you can push changes to your gitlab project.

If you are more comfortable using the command line to interact with git, you do not need to use the side panel. It will work the same way in the terminal, once you navigate to the project's filepath.

When you are ready to update your project with your changes, navigate to the git panel. Add the files you want to change to the list of staged changes. Then write a commit message, and click the check. Now you need to push your changes by selecting the push changes button on the toolbar.


Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
example
dps-updates-p
metadata-form
new-cmc-demo
File
Edit
View
Run
Kern
my-first-project / ma...
History
master
Add hello world
Staged(1)
HelloWorld.ipynb
Changed(0)
Untracked(1)
.ipynb_checkpoints/
Jobs

If you want to check your commit history, look at branches, and confirm your updates have been pushed, you can see this on the history tab.

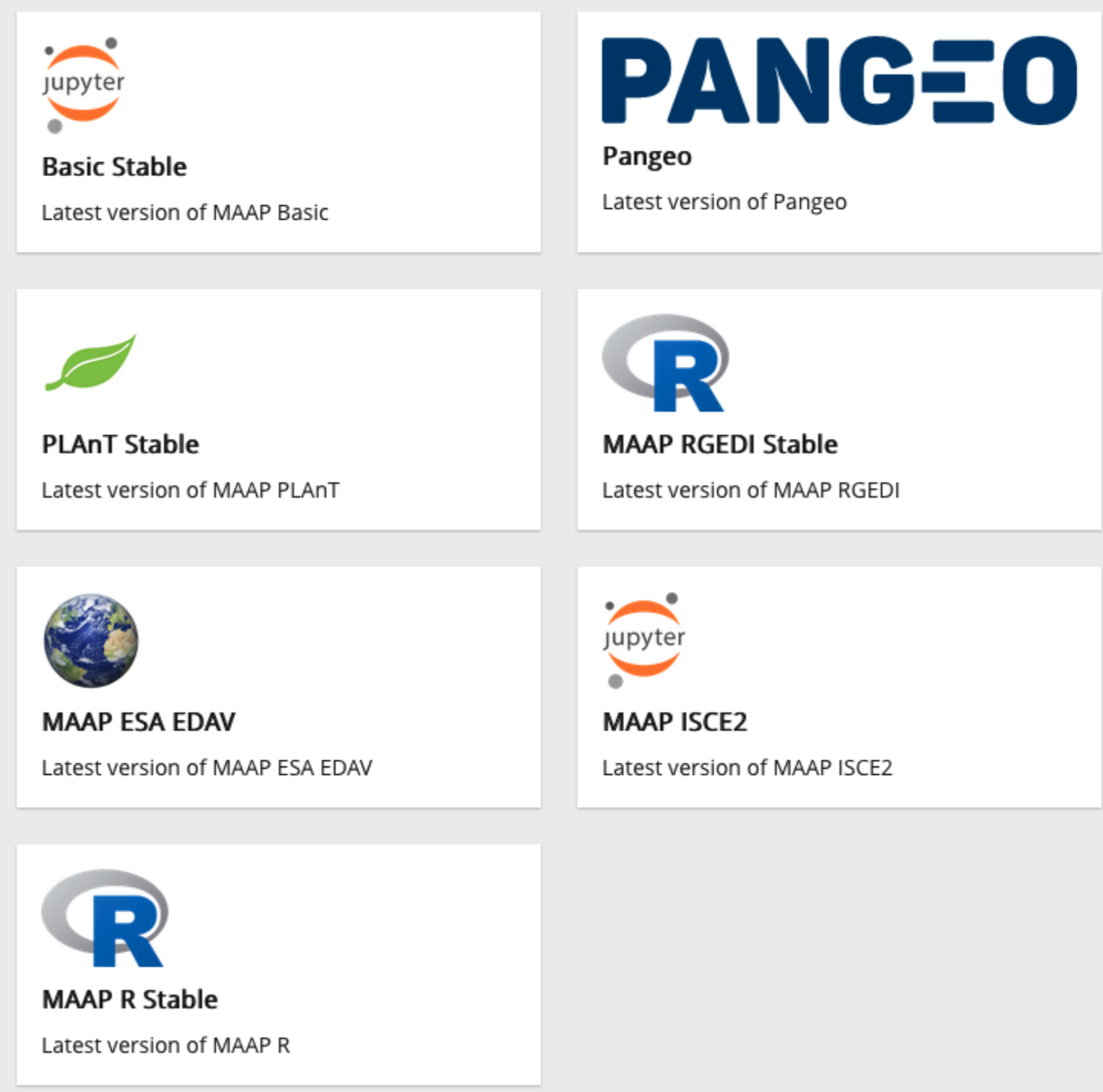

Eclipse Che | example
https://che-k8s.maap.xy
JS JPL Space
NBS
New Hires Corner
Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
example
dps-updates-p
metadata-form
File
Edit
my-first-pr
History
Maya
working
Add hello w
Maya
second cor
Maya
test
Jobs
Maya DeBe
Add READM

4.7 Environments

This document guides MAAP users in the process of selecting, extending existing environments (the set of libraries availables for analysis) or creating custom environments.

4.7.1 Workspaces

The MAAP ADE offers various workspace options, each workspace coming with its own environment that has pre-installed essential libraries for computing and geospatial analysis. At the time of writing this guide, here are the options :



For example, the `MAAP RGEDI Stable` and `MAAP R Stable` workspace options come with various pre-installed R packages.

For more information : Each of these options rely on Docker images that were build off from Dockerfiles that are publicly available in the MAAP workspace repository. If you want to learn more about what libraries each image contains, check out this repository.

4.7.2 Extending environments

Users may need libraries for their specific analysis purposes that are not present in the environments of the different workspace options offered. In this case, ideally, the steps should be the following :

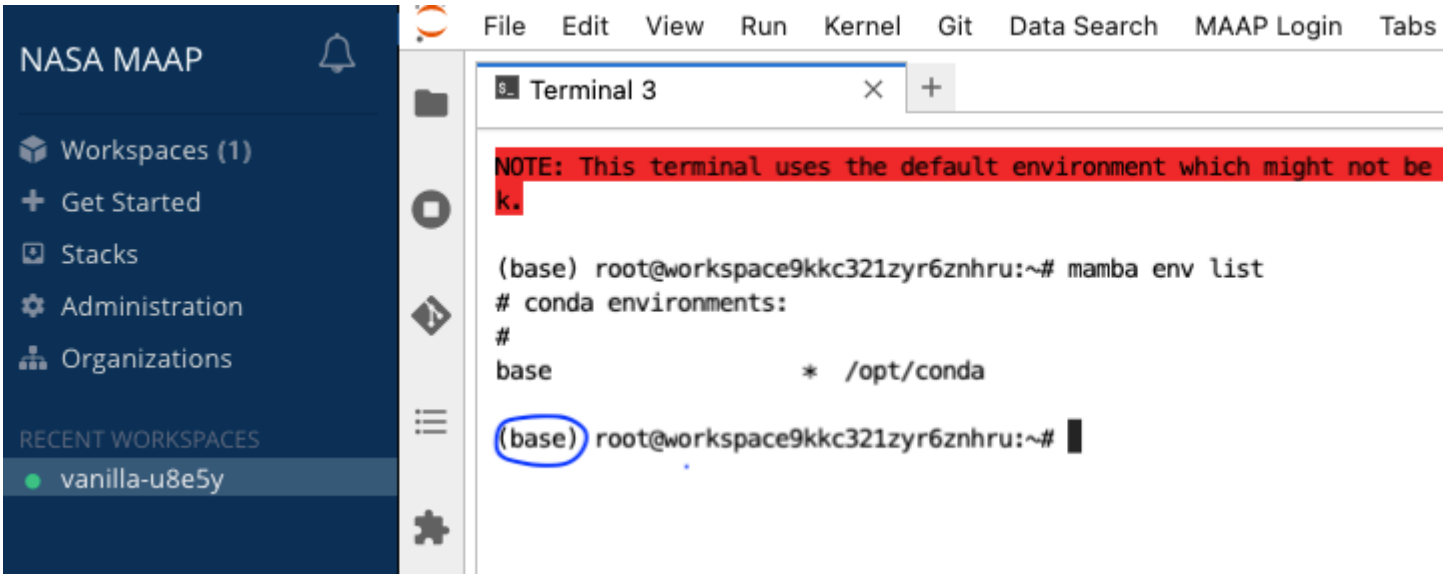
1. The user explores her/his environment need by extending the environment of an existing workspace or creating her/his custom environment in an existing workspace (see next sections).
2. Once that is done, the user submits a ticket/coordinates with the platform team to create a new workspace option with the requested, finalized environment.

The above approach is ideal because modifications to the pre-defined workspace environment do not survive a workspace restart (see next sections), and because sharing new experimented environments is valuable.

The next sections explain how to extend environments or create custom environments, and for this, introduces information regarding which environment management solution we are using.

Package manager

We use `mamba` (a fast `conda` drop-in replacement) as a package manager to install, update or remove packages (libraries). `mamba` works with 'environments' that are directories in your local file system containing a set of packages. When you work 'in a given environment', it means that your programs will look for dependencies in that environment's `mamba` directory. All workspaces launch with an environment called `base`, which is a `mamba` environment that has all the pre-installed libraries. If you open a terminal launcher after creating a `Basic Stable` workspace :



You can notice that a `base` `mamba` environment is activated, and its libraries are located in `/opt/conda`.

Extending the `base` environment in a given workspace session.

Note : any modification to the ''base'' environment does not survive a workspace restart. In other words, modifications to ''/opt/conda'' disappear after a workspace restart.

Extending an existing `mamba` environment means adding packages on top of what it contains, which works provided there are no dependency conflicts. You can install libraries using the `mamba install` command to install additional packages in your current environment (run `mamba --help` to learn more about how to use `mamba` commands). For example :

```
mamba install xarray
```

However, it is recommended to use configuration files for reproducibility and shareability. With this approach, assuming your configuration file is named `config.yml`, the command to use is :

```
mamba env update -f config.yml
```

For more details on configuration files, see the *Custom environments section* and for an example of this command, refer to the *subsection about updating an environment with a configuration file*.

4.7.3 Custom environments

For the rest of this README, in each section we provide a link to download an example YAML configuration file.

You can use the `mamba` CLI to create a new, custom environment. The parameters (the list of libraries, the location where to search for them, etc...) can be passed either from a configuration YAML file or directly on the console. We recommend using the first option (a YAML file is easier to share and modify).

Basic custom environment

*Example config file for this section*here.

This configuration installs specific versions `python`, `pandas` and `geopandas` from `conda-forge`. If versions aren't specified, the latest is installed. We recommend to always specify the version for reproducibility. The basic command to create this environment would be :

```
mamba env create -f env-example.yml
```

However, this stores this environment files in `/opt/conda`, which is a directory that is recreated when the workspace restarts, and so custom environments are lost. Therefore, you want to specify a storage location in your user directory with the `--prefix` parameter

```
mamba env create -f env-example.yml --prefix /projects/env
```

and to activate it :

```
mamba activate env-example
```

Updating an existing environment with a configuration file

*Example config file for this section*here.

You can *update* an existing environment with a configuration file as well. For example, let's assume you have a `mamba` environment with a set of packages already installed in it (for example the `base` environment), but it doesn't have `xarray` and `geopandas`. Using the linked example config :

```
mamba env update -f env-extend.yml
```

This command will update `base` by adding `xarray` and `geopandas`, provided it does not cause conflicts with the existing libraries.

Using `pip` for python packages

*Example config file for this section*here.

Some python packages might not be availabe in the channel you are using, or in any `mamba` channel. If that package however is in `PyPI` (the official python package repository), one can use `pip` within a `mamba` environment to download packages. The recommended way is to specify this in the configuration file. In the linked example, we add `stackstac` as a dependency to install from `PyPI` because it is not available in the `conda-forge` channel.

Using custom environments in jupyter notebooks

*Example config file for this section*here.

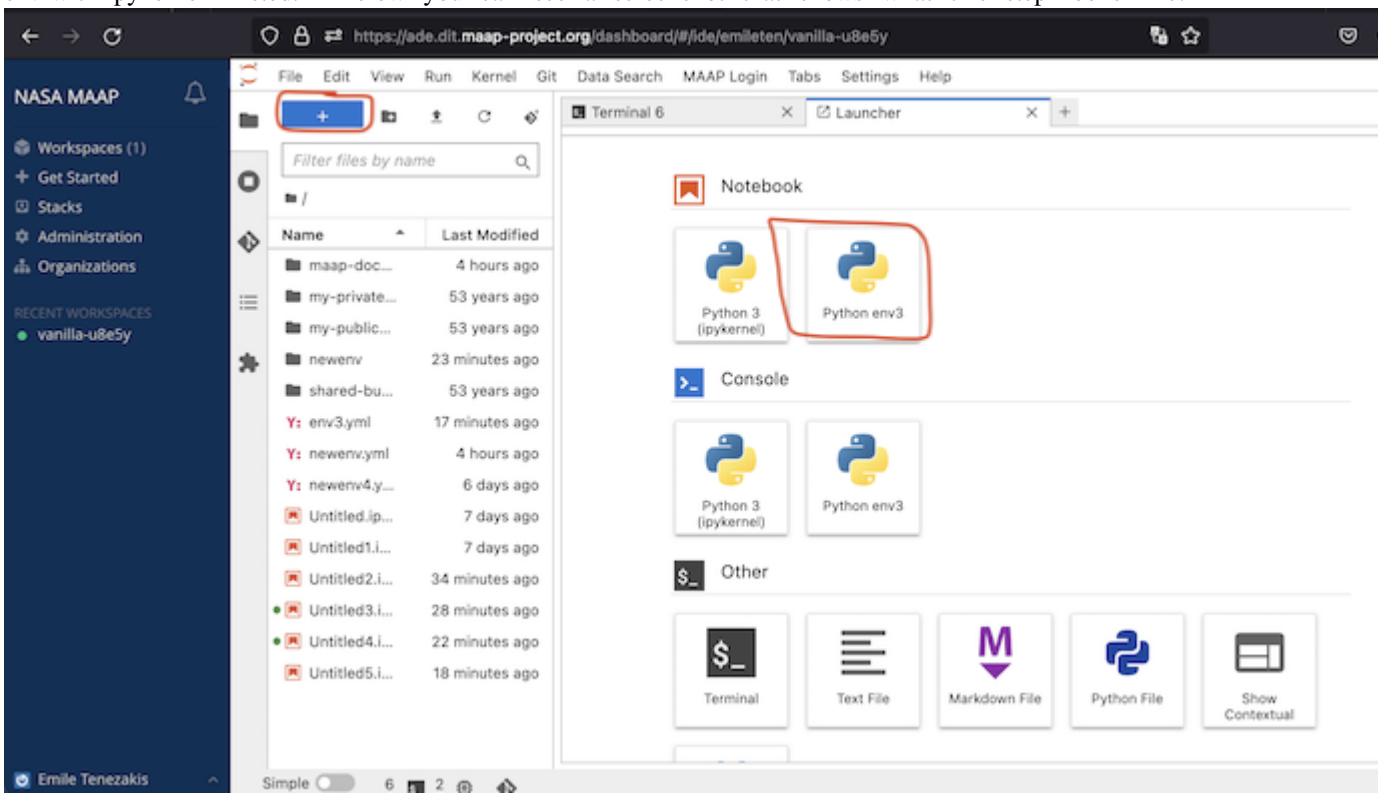
The following instruction steps are for python kernels.

- Make sure ipykernel is listed as a dependency in your configuration file.
- Create your environment using the linked configuration file.
- Install the environment as a kernel by running the following command (parameter values follow the example mentioned):

```
python -m ipykernel install --user --name env-with-ipykernel --display-name
↪"Python env-with-ipykernel".
```

  The above command installs the environment as a kernel in Jupyter, making it accessible in the notebook with a display name of "Python env-with-ipykernel".

- Wait around 30 seconds and launch a new notebook. Among the kernel options, you should see "Python env-with-ipykernel" listed. Below you can see a screenshot that shows what this step looks like:



Suggested packages for custom environment

*Example config file for this section*here

MAAP users typically use the python `maap-py`. It's pre-installed in all workspaces, in the `base` mamba environment, but any custom environment should specify it, otherwise it is not going to be accessible from that environment. However, `maap-py` is not packaged in a public package repository, like `PyPI` or `conda-forge`. It is possible to install it directly from its github repository with `pip` though. See the configuration example linked. You can note that in the example, `maap-py` is 'versioned' using a commit hash (at the end of the github URL).

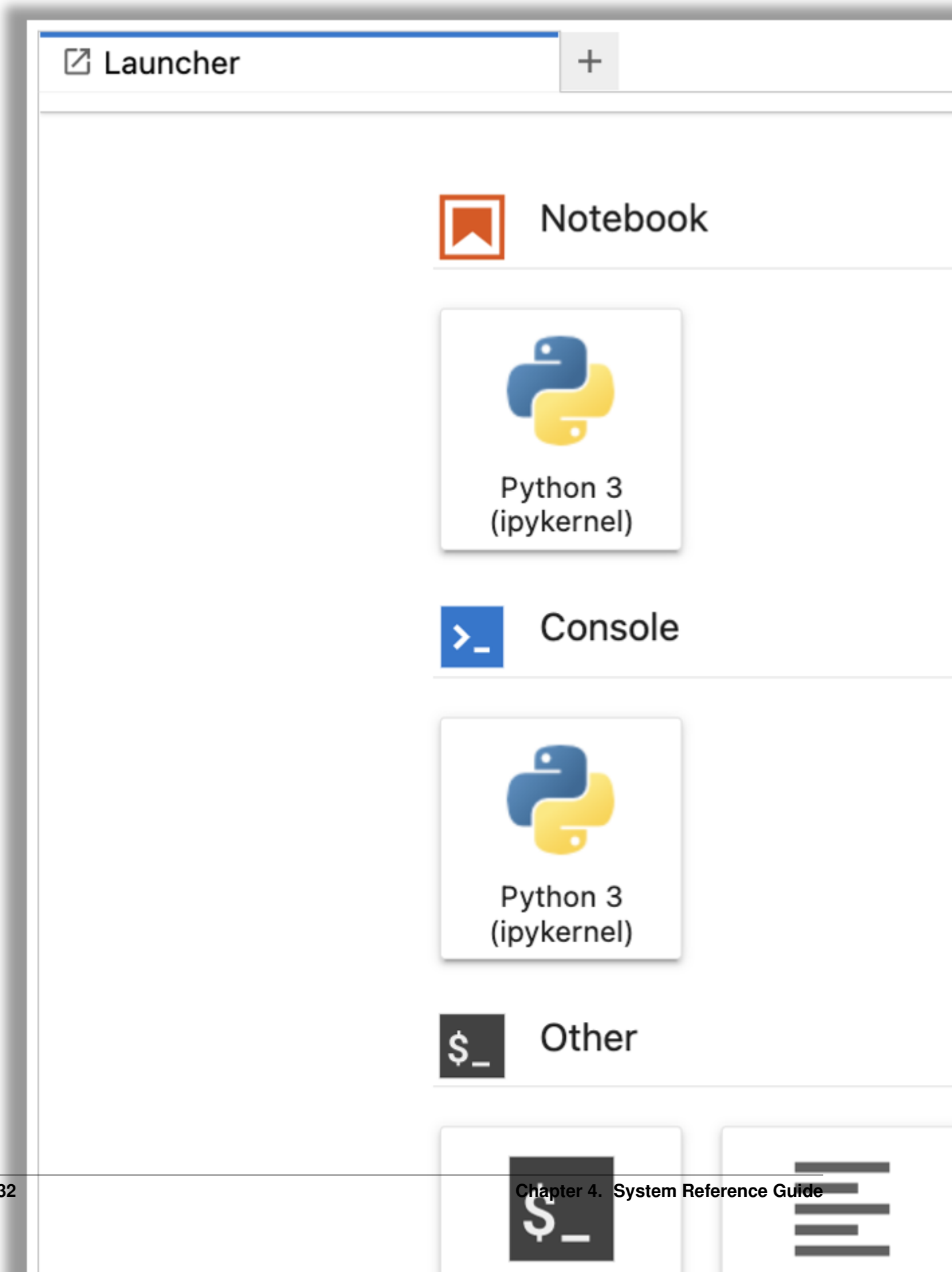
4.8 ADE Custom Extensions

4.8.1 Jobs UI

The Jobs UI allows users to submit and view DPS jobs from their Jupyter workspace. Users can monitor job status, access generated products, view errors, and view other job metadata.

Access

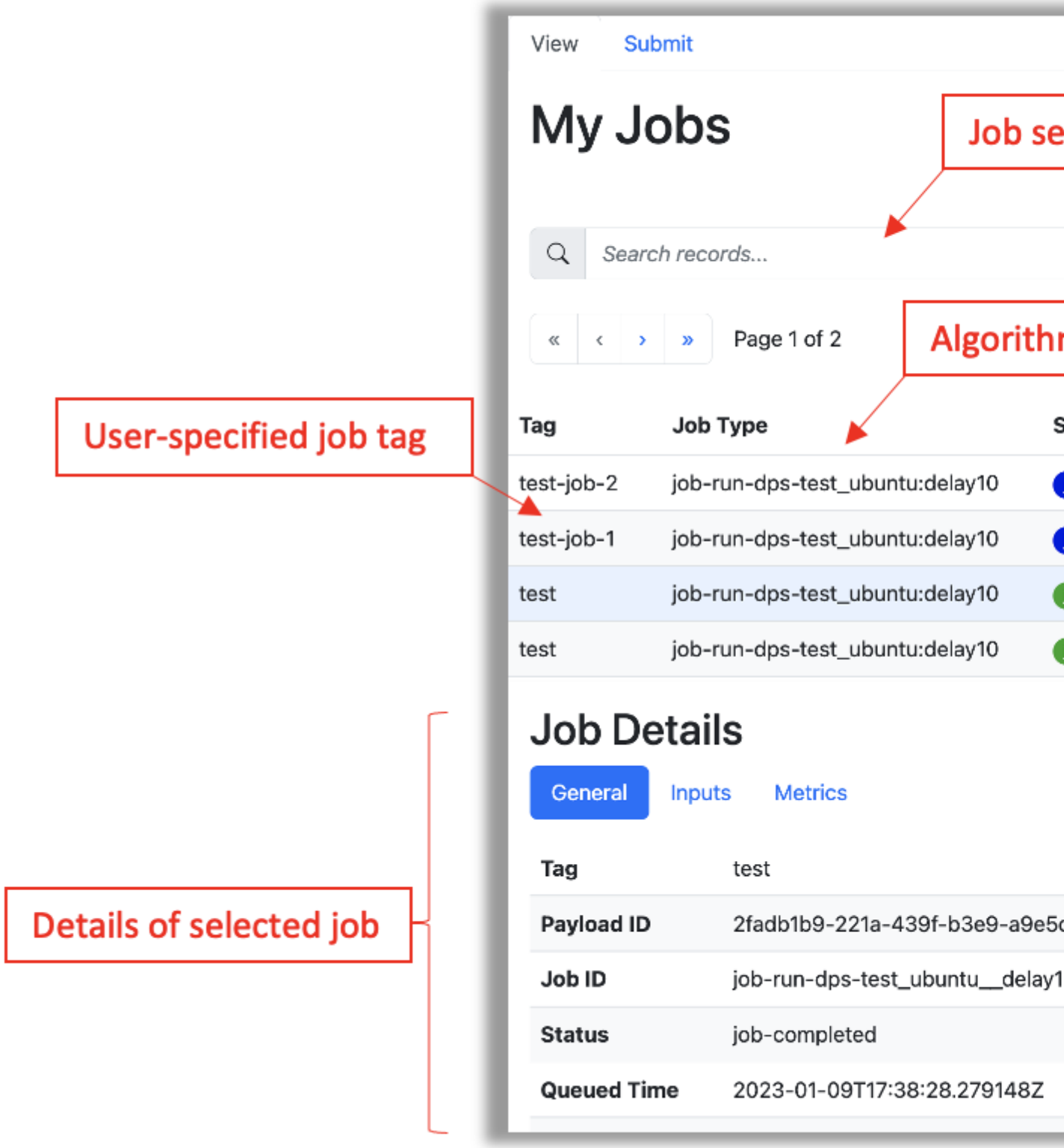
1. From your workspace click on the **View & Submit Jobs** card on the launcher tab.


Launcher
Notebook
Python 3
(ipykernel)
Console
Python 3
(ipykernel)
Other

View Jobs

The **View** pane lists all the jobs submitted by the user. The top table shows only a few key fields. Users can click on any row to view detailed data for the selected job such as inputs, outputs generated, and errors produced if the job failed to complete successfully.

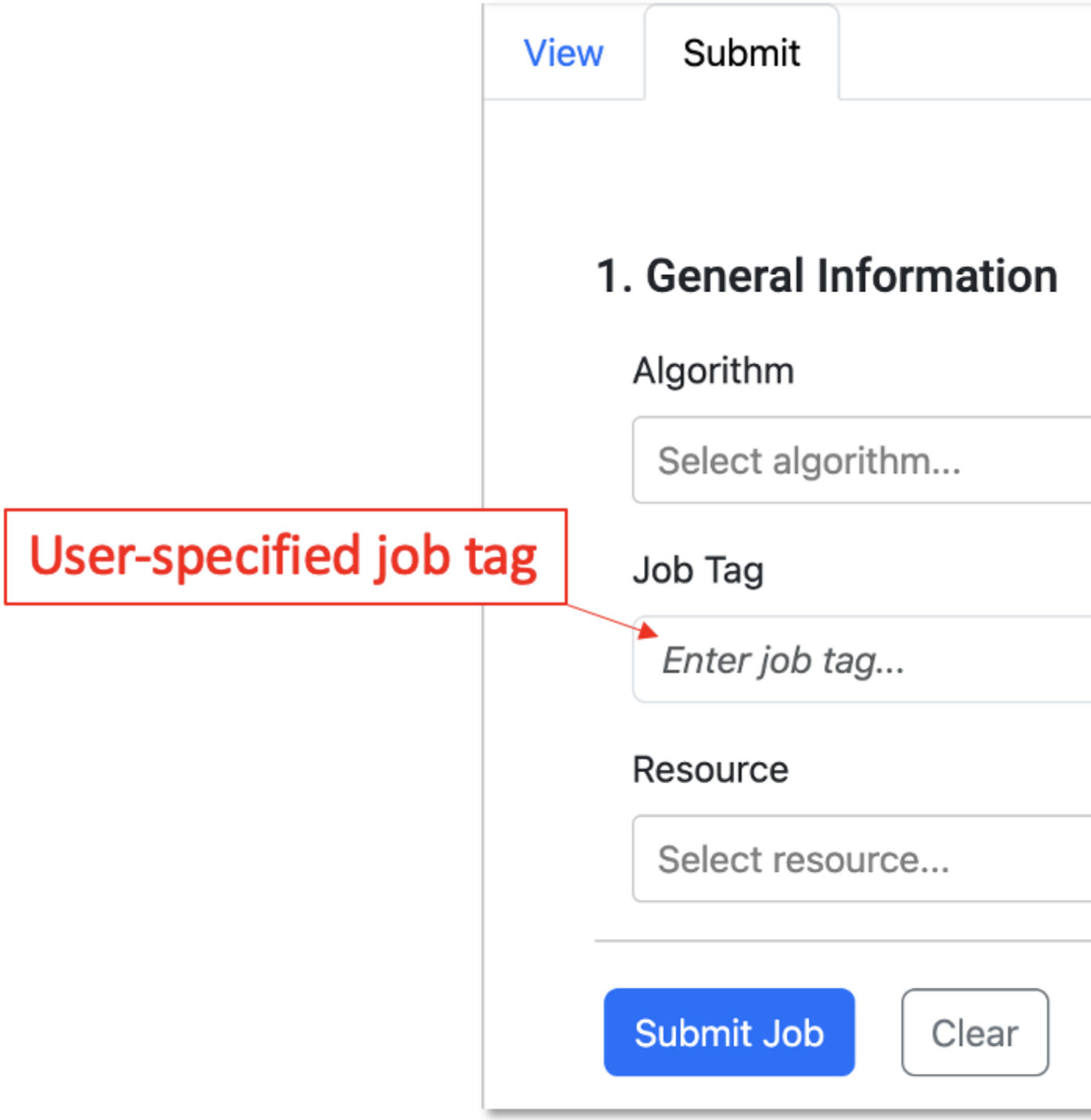
Users can sort jobs by queued, start, and end time in ascending/descending order. Users may use the search bar to filter the job list down to jobs containing the user-provided string in any of the fields shown.


View
Submit
My Jobs
Job se
Search records...
« ‹ › »
Page 1 of 2
Algorithm
User-specified job tag
Tag
Job Type
test-job-2
job-run-dps-test_ubuntu:delay10
test-job-1
job-run-dps-test_ubuntu:delay10
test
job-run-dps-test_ubuntu:delay10
test
job-run-dps-test_ubuntu:delay10
Job Details
General
Inputs
Metrics
Details of selected job
Tag
test
Payload ID
2fadb1b9-221a-439f-b3e9-a9e5
Job ID
job-run-dps-test_ubuntu__delay1
Status
job-completed
Queued Time
2023-01-09T17:38:28.279148Z

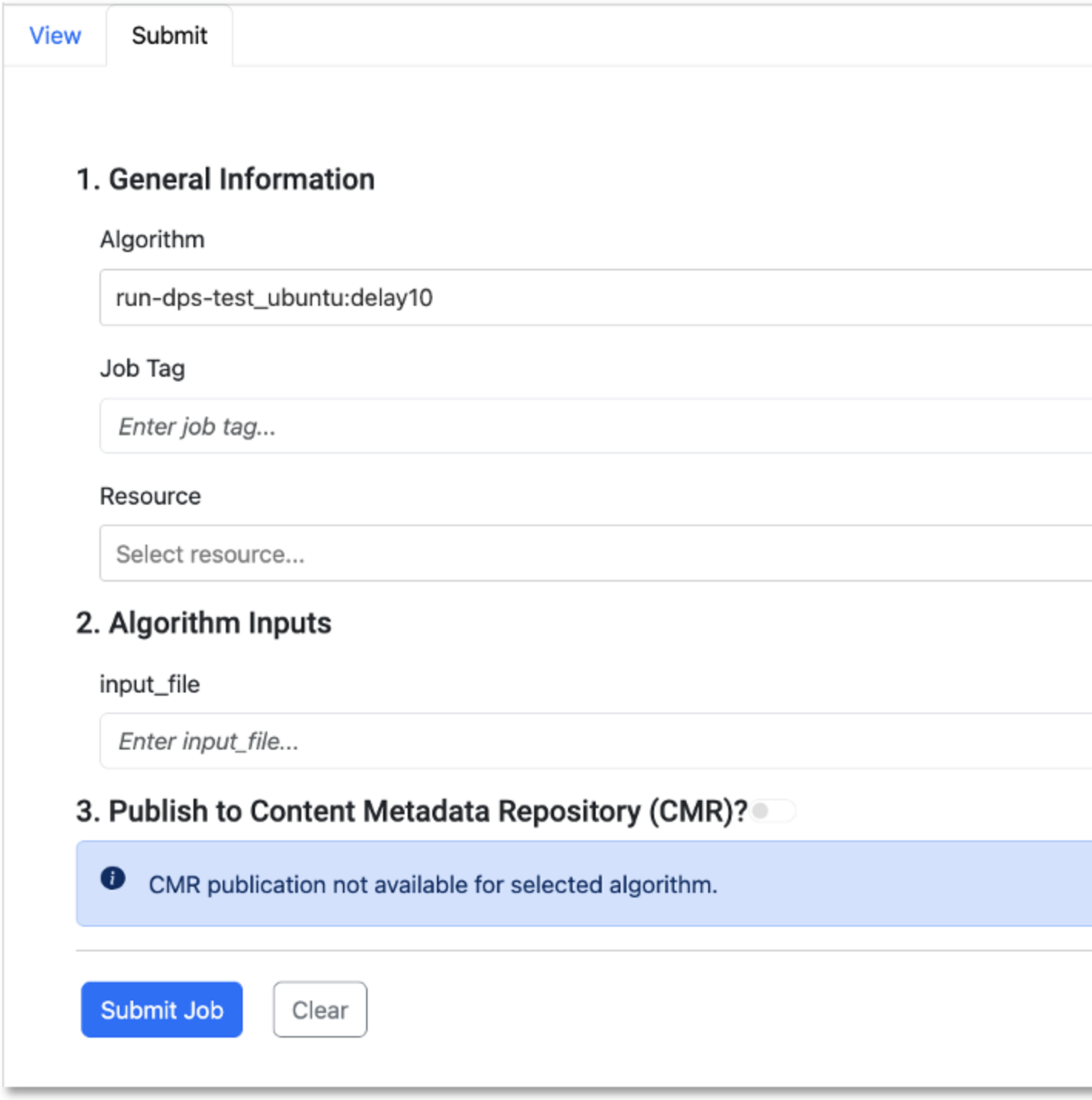
Submit Jobs

Users can submit individual jobs from the **Submit** pane. The following are the minimum requirements for submitting a job:

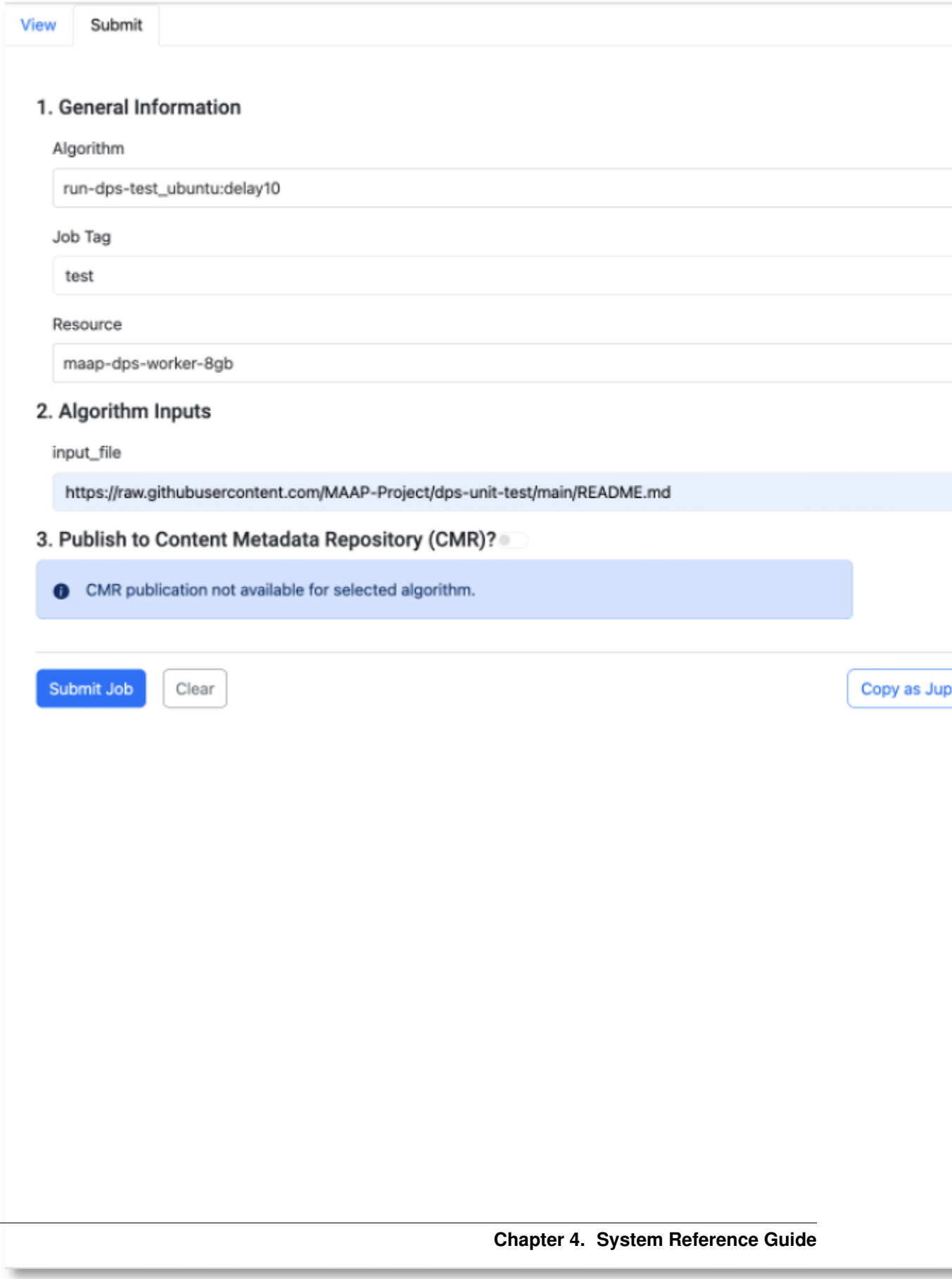
1. Select an algorithm from the dropdown menu. Note: only registered algorithms will be shown.
2. Provide a tag that may then be used to easily search for and identify the submitted job.
3. Select the resource to use for algorithm execution.



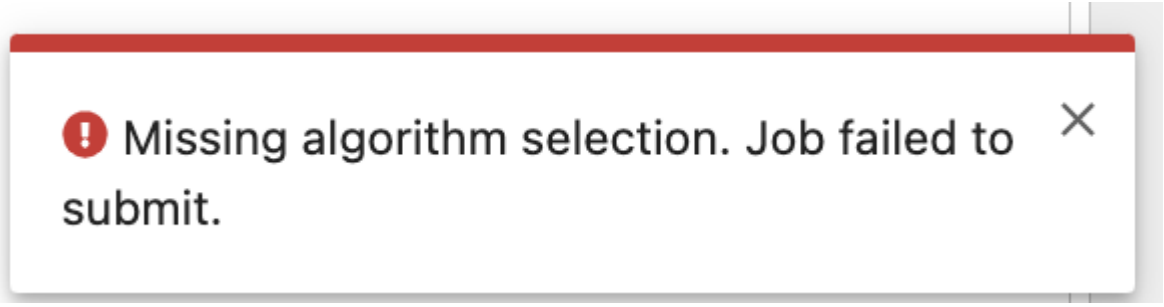
Algorithms may contain additional inputs that users may have to provide.



Once all inputs have been provided, the user may click **Submit Job** to submit the job. If the job was submitted successfully, a toast will appear in the bottom right corner containing the unique job id.

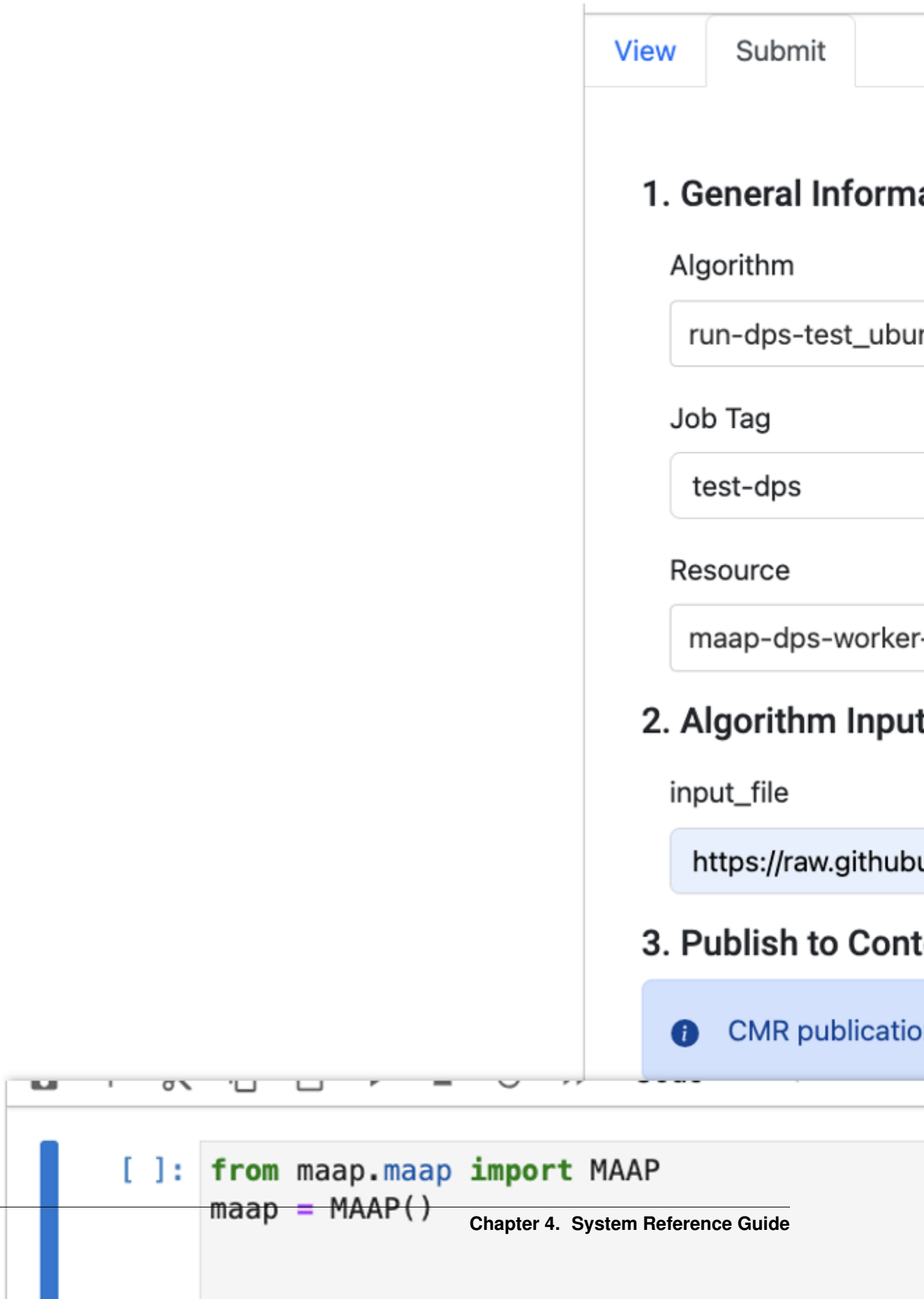

View
Submit
1. General Information
Algorithm
run-dps-test_ubuntu:delay10
Job Tag
test
Resource
maap-dps-worker-8gb
2. Algorithm Inputs
input_file
https://raw.githubusercontent.com/MAAP-Project/dps-unit-test/main/README.md
3. Publish to Content Metadata Repository (CMR)?
CMR publication not available for selected algorithm.
Submit Job
Clear
Copy as Jup

If the job failed to submit, a toast will appear indicating the job failed to submit.



Generate Job Submission Command

Users may fill out the job submission form and - instead of submitting the job - click the **Copy Jupyter Notebook Code** button to copy the **maap-py** job submission command to their clipboard to then paste it into a Jupyter notebook.

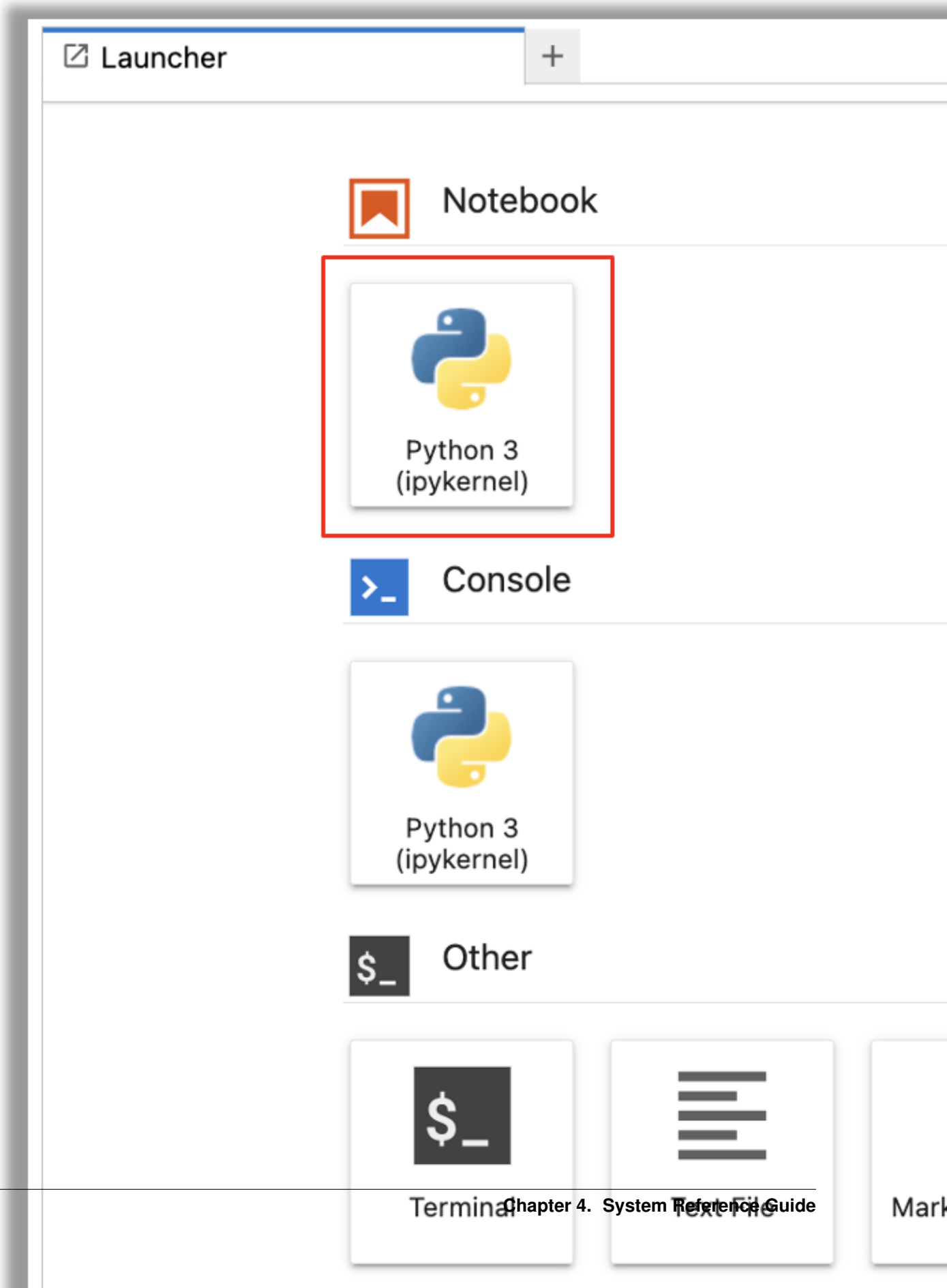

View
Submit
1. General Informa
Algorithm
run-dps-test_ubur
Job Tag
test-dps
Resource
maap-dps-worker-
2. Algorithm Input
input_file
https://raw.githubu
3. Publish to Cont
CMR publicatio
[ ]:
from maap.maap import MAAP
maap = MAAP()

4.8.2 MAAP Libraries

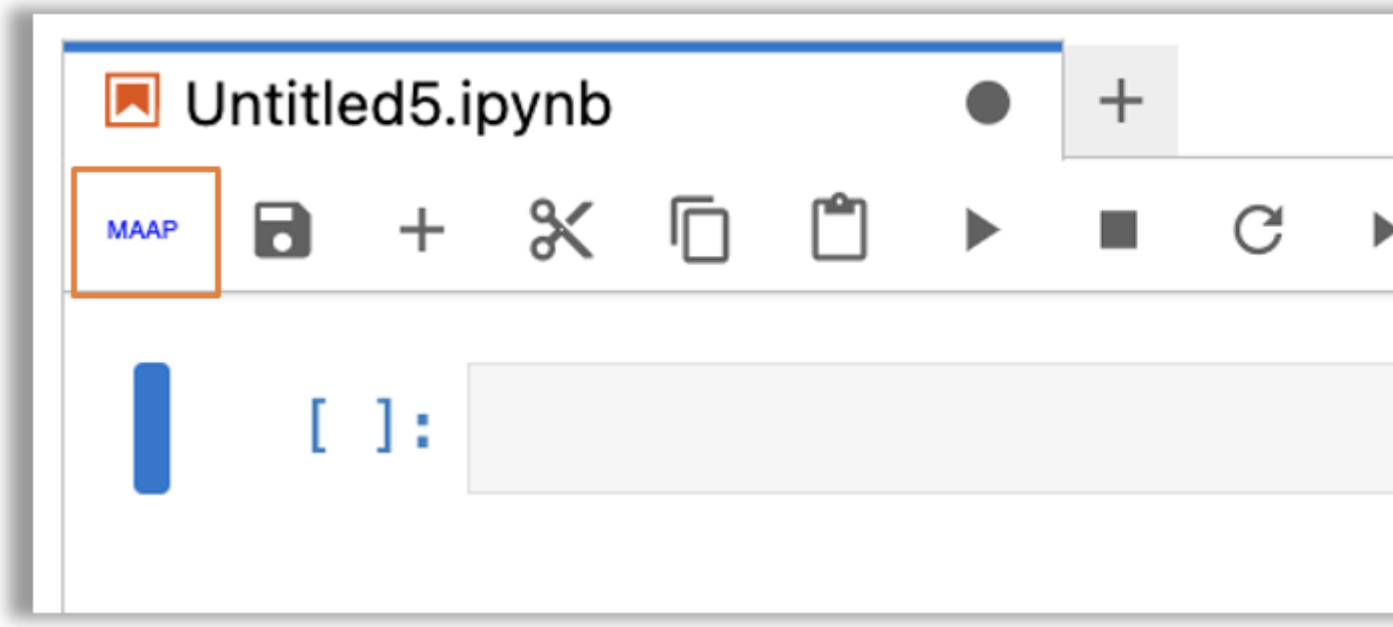
This Jupyter extension generates boilerplate code that imports and initializes MAAP libraries for usage in Jupyter notebooks. It currently supports the maap-py client library and ipycmc.

Access

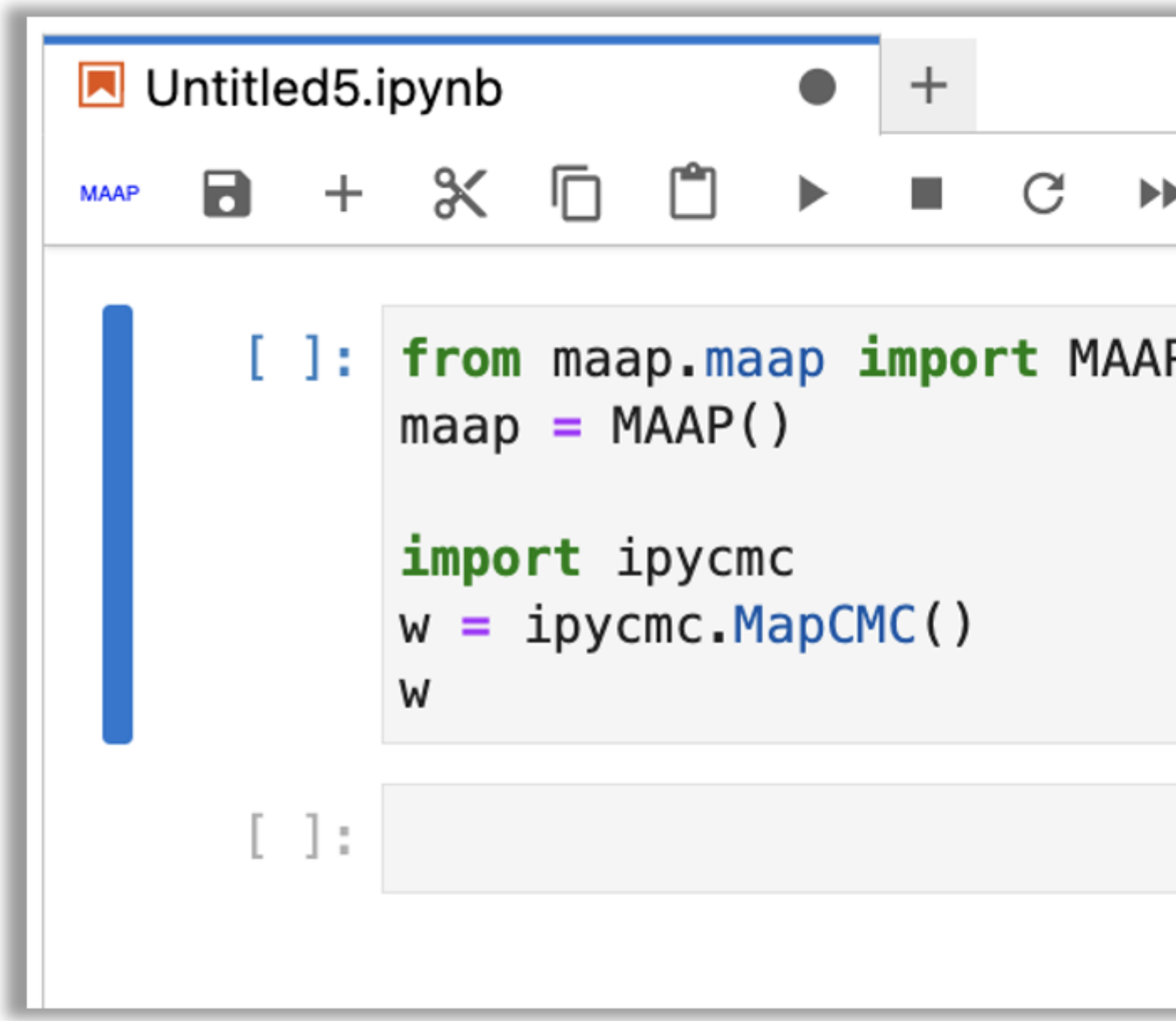
1. From your workspace, create a Jupyter notebook from the launcher pane.


Launcher
Notebook
Python 3
(ipykernel)
Console
Python 3
(ipykernel)
Other
Terminal
Text File
Mark

2. Click on the **MAAP** icon.



3. The Jupyter notebook will be populated with the following code block.

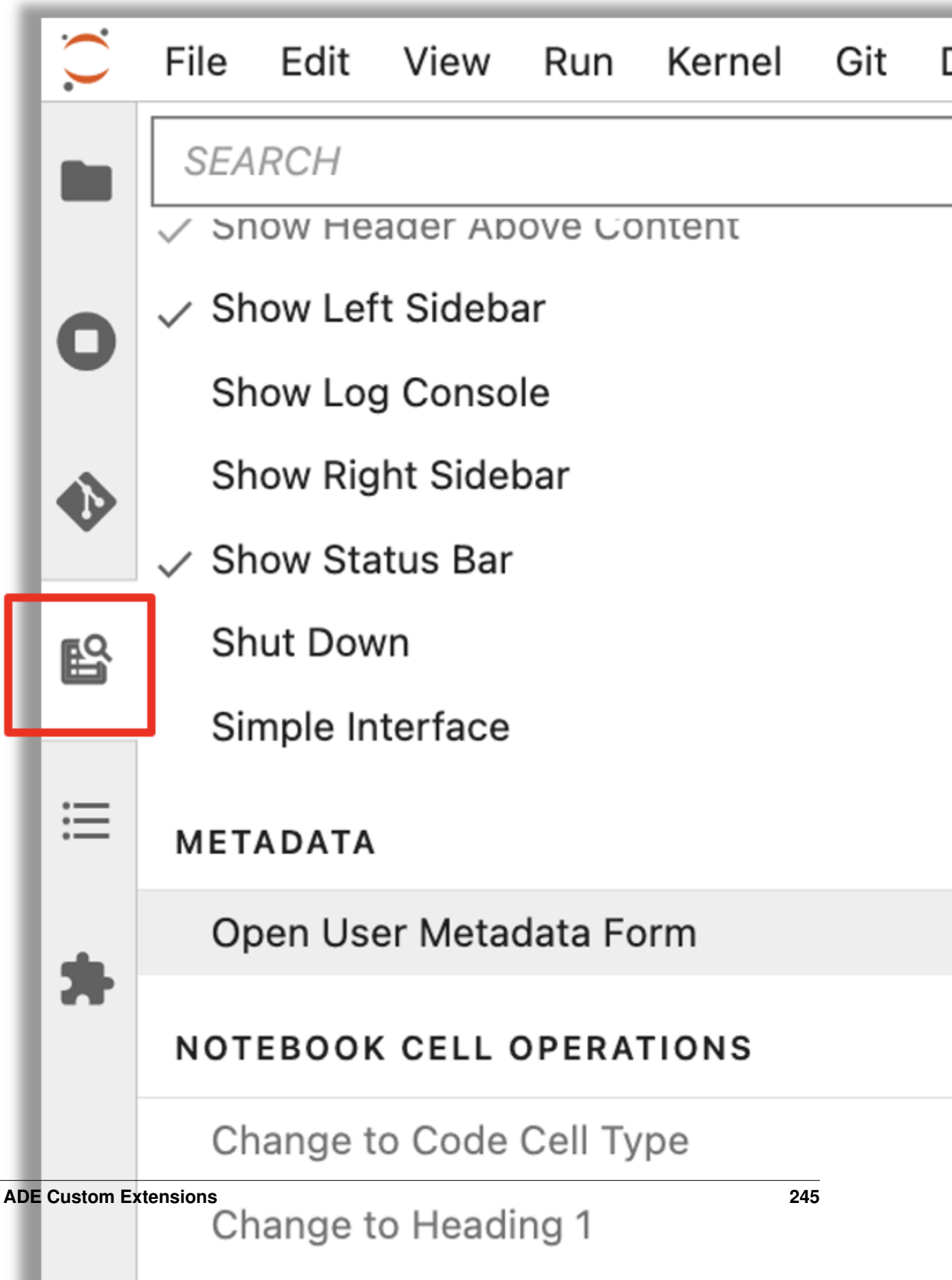


4.8.3 User Metadata Form

This extension allows users to easily access the **MAAP User Shared Data Questionnaire**.

Access

1. From your workspace, click on the command icon on the left ribbon menu and scroll down to the Metadata section.


File
Edit
View
Run
Kernel
Git
SEARCH
Show Header Above Content
Show Left Sidebar
Show Log Console
Show Right Sidebar
Show Status Bar
Shut Down
Simple Interface
METADATA
Open User Metadata Form
NOTEBOOK CELL OPERATIONS
Change to Code Cell Type
Change to Heading 1

2. Click on **Open User Metadata Form** to be redirected to the shared-data questionnaire in a new window.

4.8.4 Jupyter Server Extension

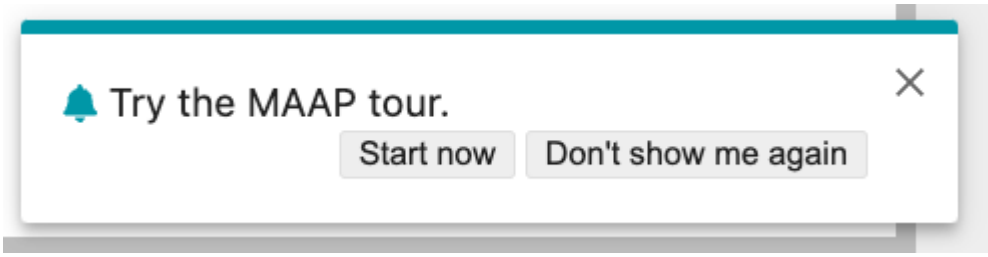
This is the backend extension for the MAAP common collection of custom Jupyter extensions. This backend is accessible to users and other Jupyter extensions through a RESTful interface. The supported endpoints interact with the MAAP API and user workspace. For a list of the endpoints, refer to the repo documentation.

4.8.5 Maap Help Jupyter Extension

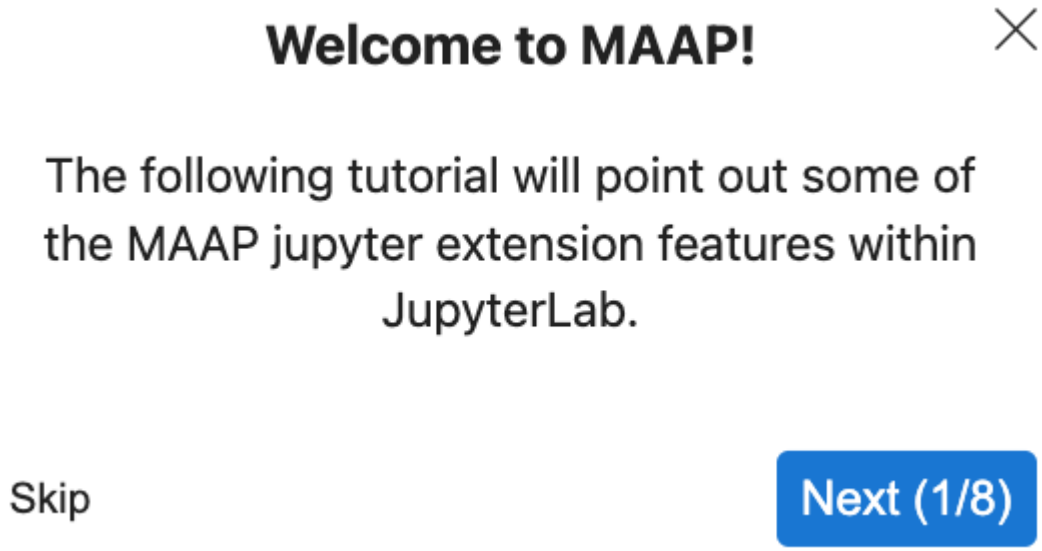
This Jupyter extension provides an interactive tutorial on a new user's first launch of the ADE and adds MAAP specific information to the help tab. The interactive tutorial can be accessed again from the help menu.

Interactive Tour

1. Launch the ADE for your first time as a new user, and see the MAAP tour notification



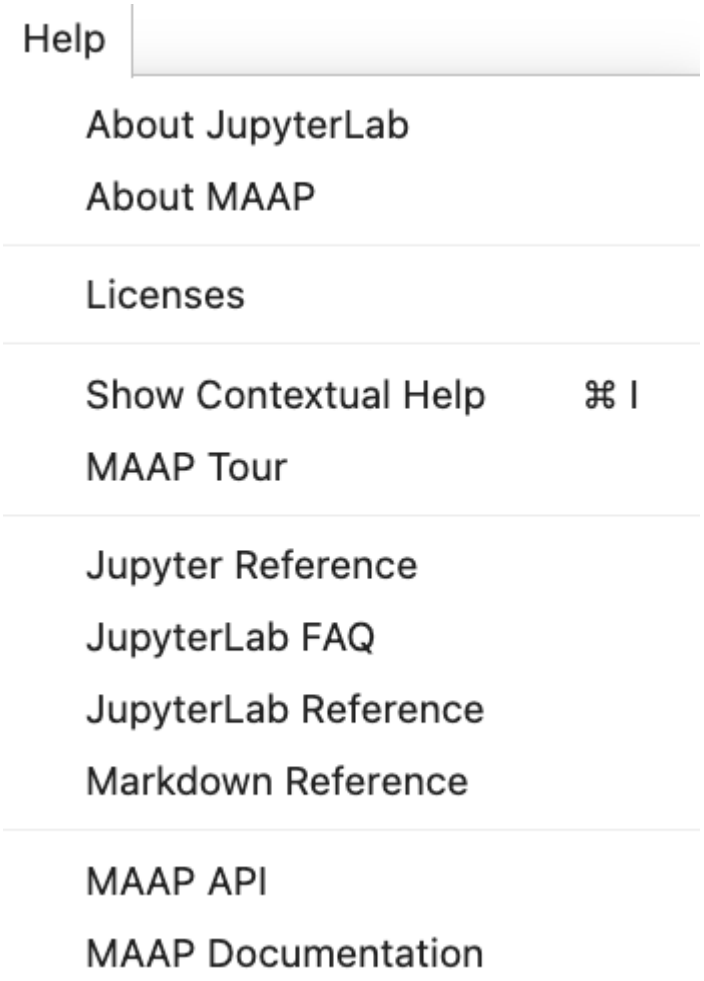
2. Step through the tour with the "Next" button.



3. The MAAP tour will no longer appear on launch if you a) finish the tour, b) press skip during the tour, or c) press "Don't show me again." You can revisit the MAAP tour by selecting it from the help menu.

MAAP Help Menu

Options show MAAP specific help information along with default JupyterLab help information.



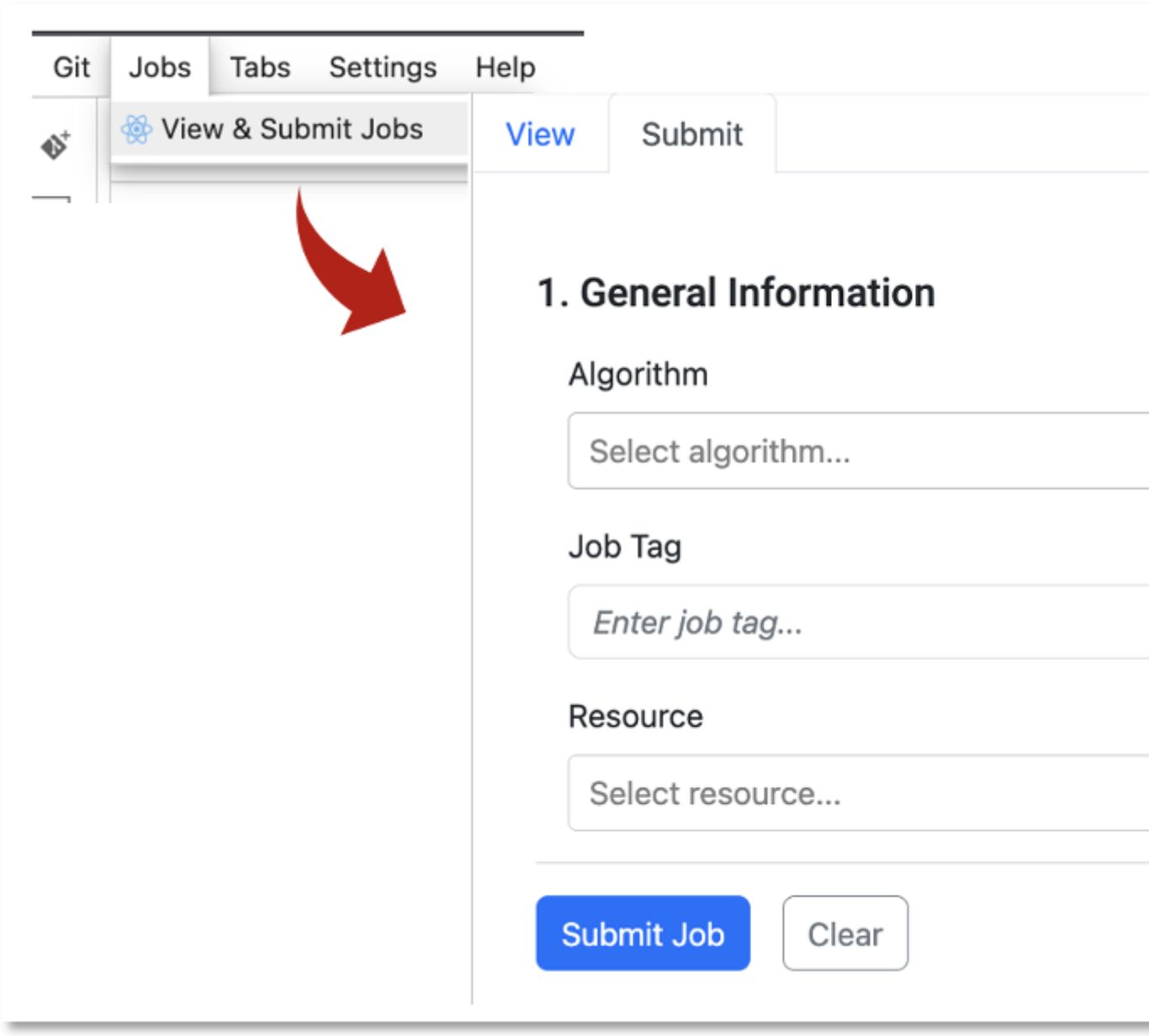
4.9 FAQ

4.9.1 How do I submit a job?

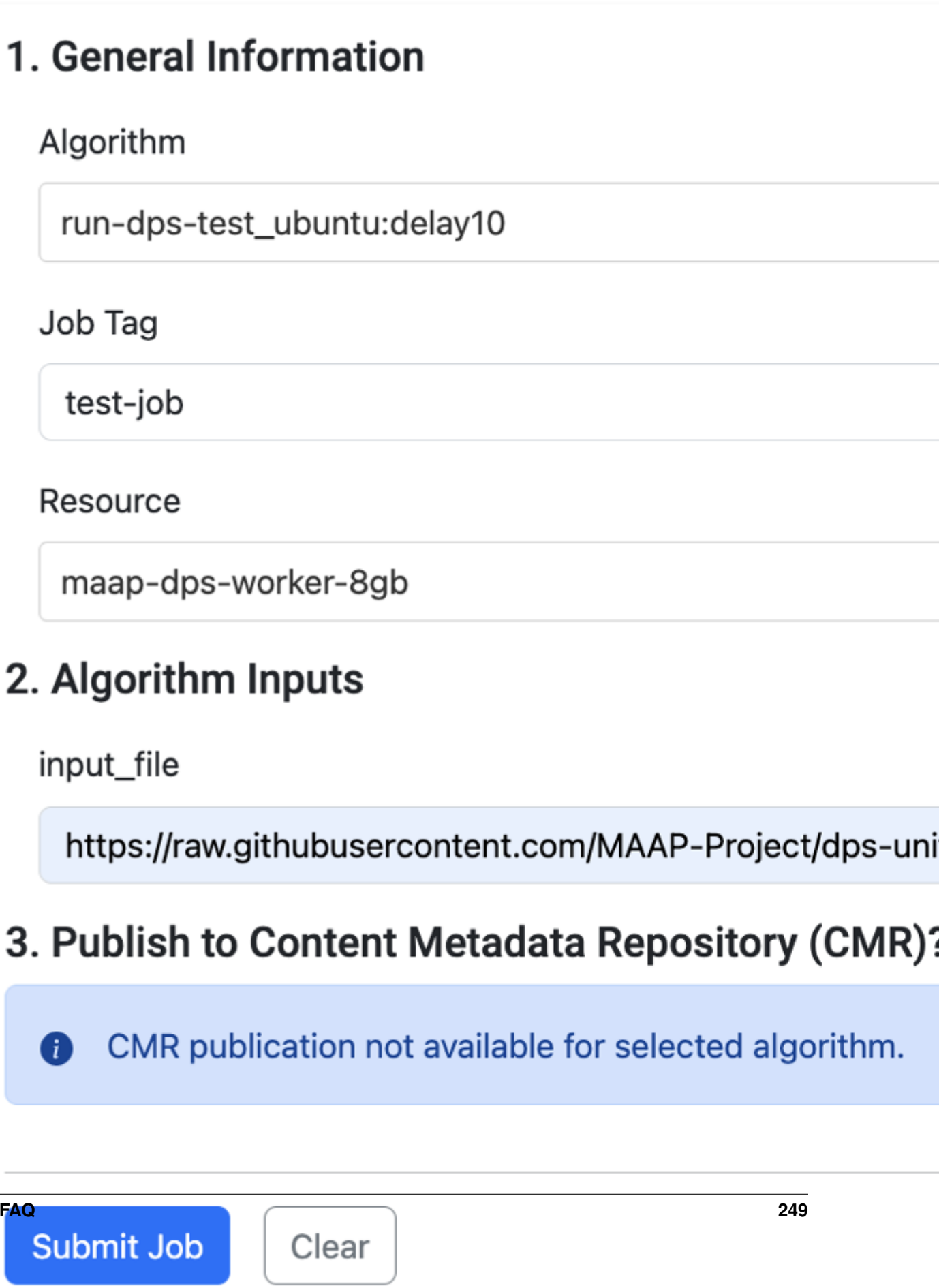
Users may submit jobs using the Jobs UI or the maap-py client library.

Jobs UI

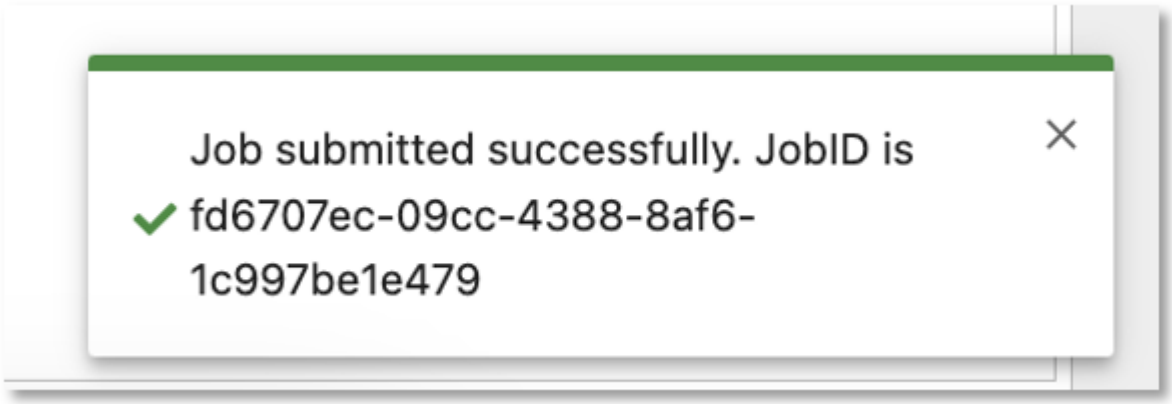
1. Open the Jobs UI from the `Jobs` menu then navigate to the Jobs UI `Submit` tab.



2. Select the algorithm you would like to run then enter the algorithm inputs.

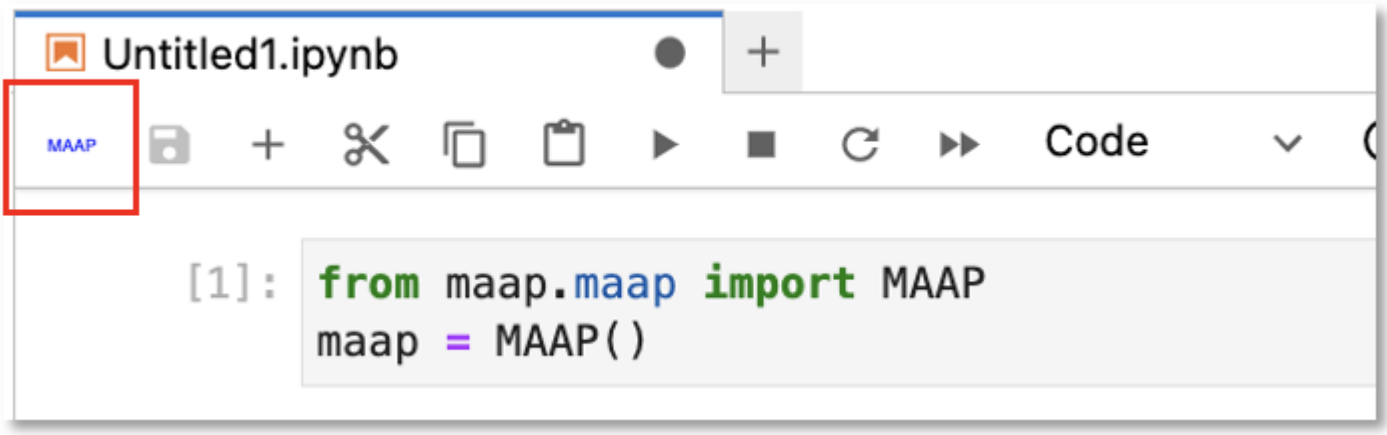

1. General Information
Algorithm
run-dps-test_ubuntu:delay10
Job Tag
test-job
Resource
maap-dps-worker-8gb
2. Algorithm Inputs
input_file
https://raw.githubusercontent.com/MAAP-Project/dps-uni
3. Publish to Content Metadata Repository (CMR)?
CMR publication not available for selected algorithm.
Submit Job
Clear

3. Click the `Submit Job` button. A toast will appear in the bottom right of the screen indicating whether job submission was successfull or not.



MAAP-py Library

1. Open a Jupyter Notebook then click the `MAAP` button from the notebook toolbar. This will paste the code snippet below into your notebook.



2. Provide the MAAP host. For DIT, this would be `api.dit.maap-project.org`

```
from maap.maap import MAAP
maap = MAAP(maap_host='api.dit.maap-project.org')
```

3. Use the `submitJob` method and provide your algorithm inputs. The example below will run the `run-dps-test_ubuntu` algorithm.

```
maap.submitJob(identifier="test-job",
               algo_id="run-dps-test_ubuntu",
               version="delay10",
               username="anonymous",
               queue="maap-dps-worker-8gb",
               input_file="https://raw.githubusercontent.com/MAAP-Project/dps-unit-
↪test/main/README.md")
```

4. Run the notebook to submit the job. The cell output for a job that was submitted successfully will look similar to this:

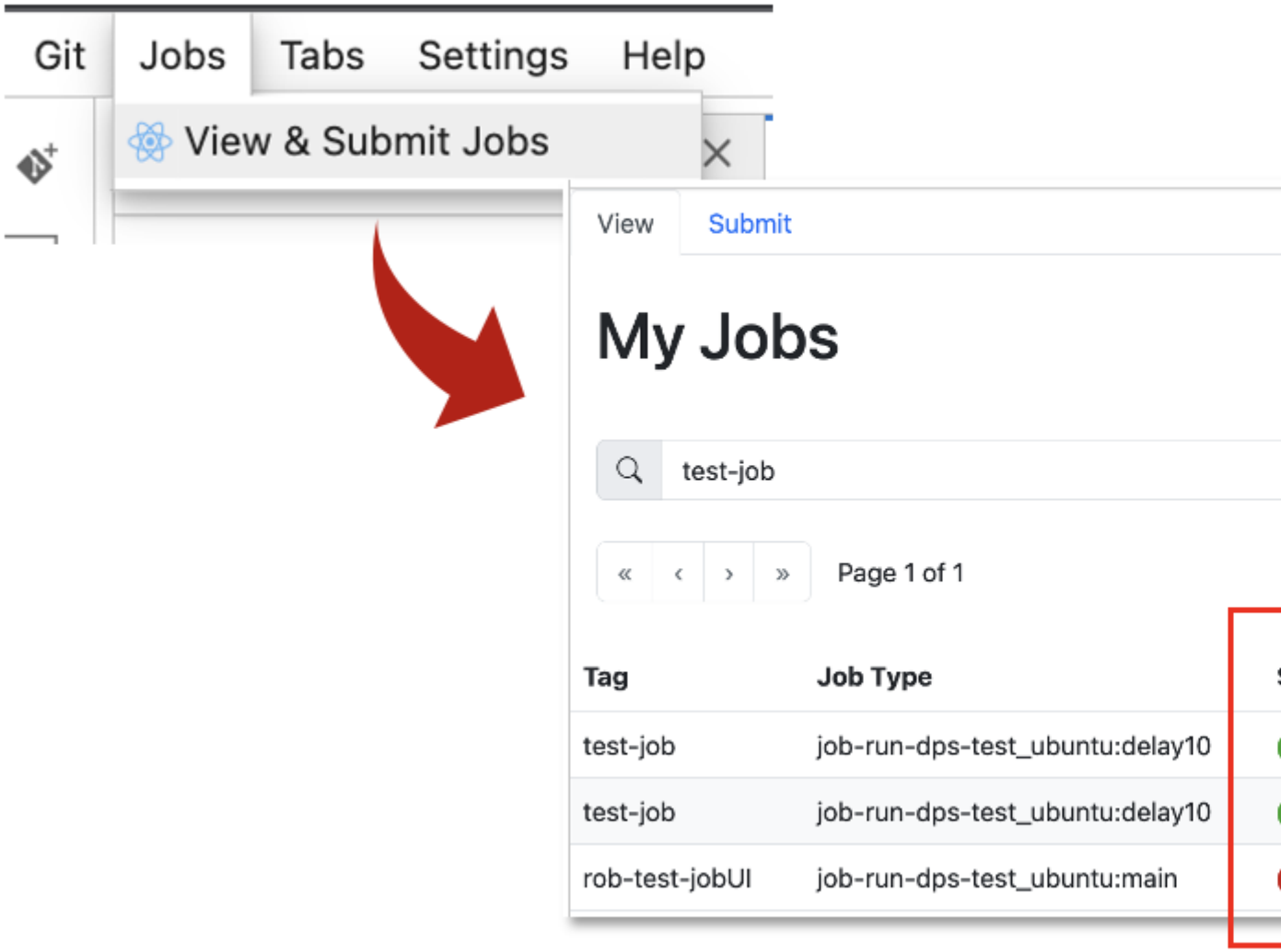
```
{'status': 'success',
 'http_status_code': 200,
 'job_id': '86fbac52-24b0-4963-8b67-59d0fc09946a'}
```

4.9.2 How do I check the status of my job?

Users may check the status of their job using the Jobs UI or the maap-py client library.

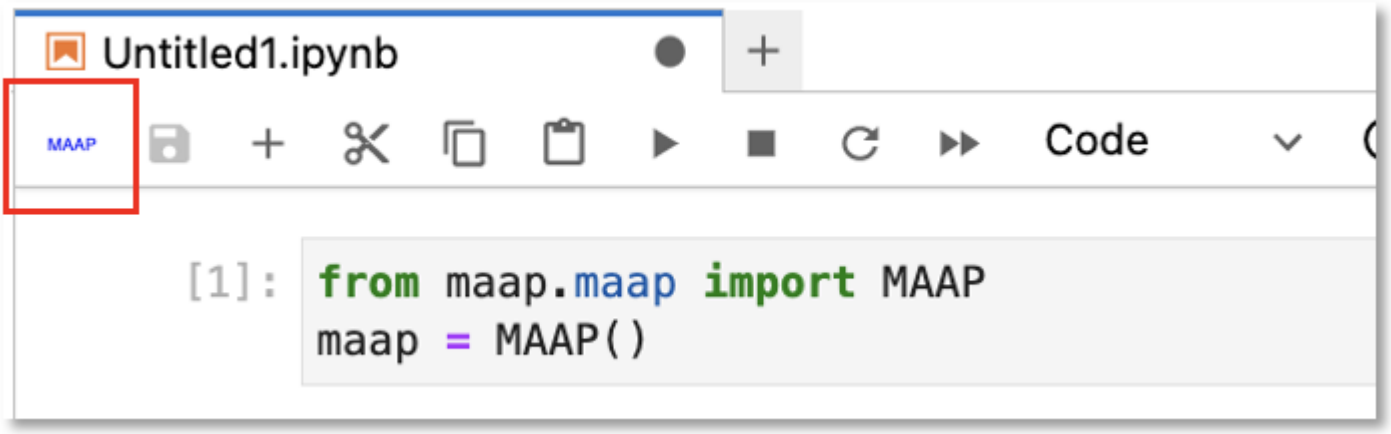
Jobs UI

1. Open the Jobs UI from the `Jobs` menu then navigate to the Jobs UI `View` tab. The `Status` column indicates the status for a given job.



MAAP-py Library

1. Open a Jupyter Notebook then click the `MAAP` button from the notebook toolbar. This will paste the code snippet below into your notebook.



2. Provide the MAAP host. For DIT, this would be `api.dit.maap-project.org`

```
from maap.maap import MAAP
maap = MAAP(maap_host='api.dit.maap-project.org')
```

3. Use the `getJobStatus` method and provide the job ID that was created upon job submission.

```
r = maap.getJobStatus("86fbac52-24b0-4963-8b67-59d0fc09946a")
r.text
```

4. Run the notebook to get the job status. The output should resemble the xml snippet below. In this example, the job status is `Succeeded`.

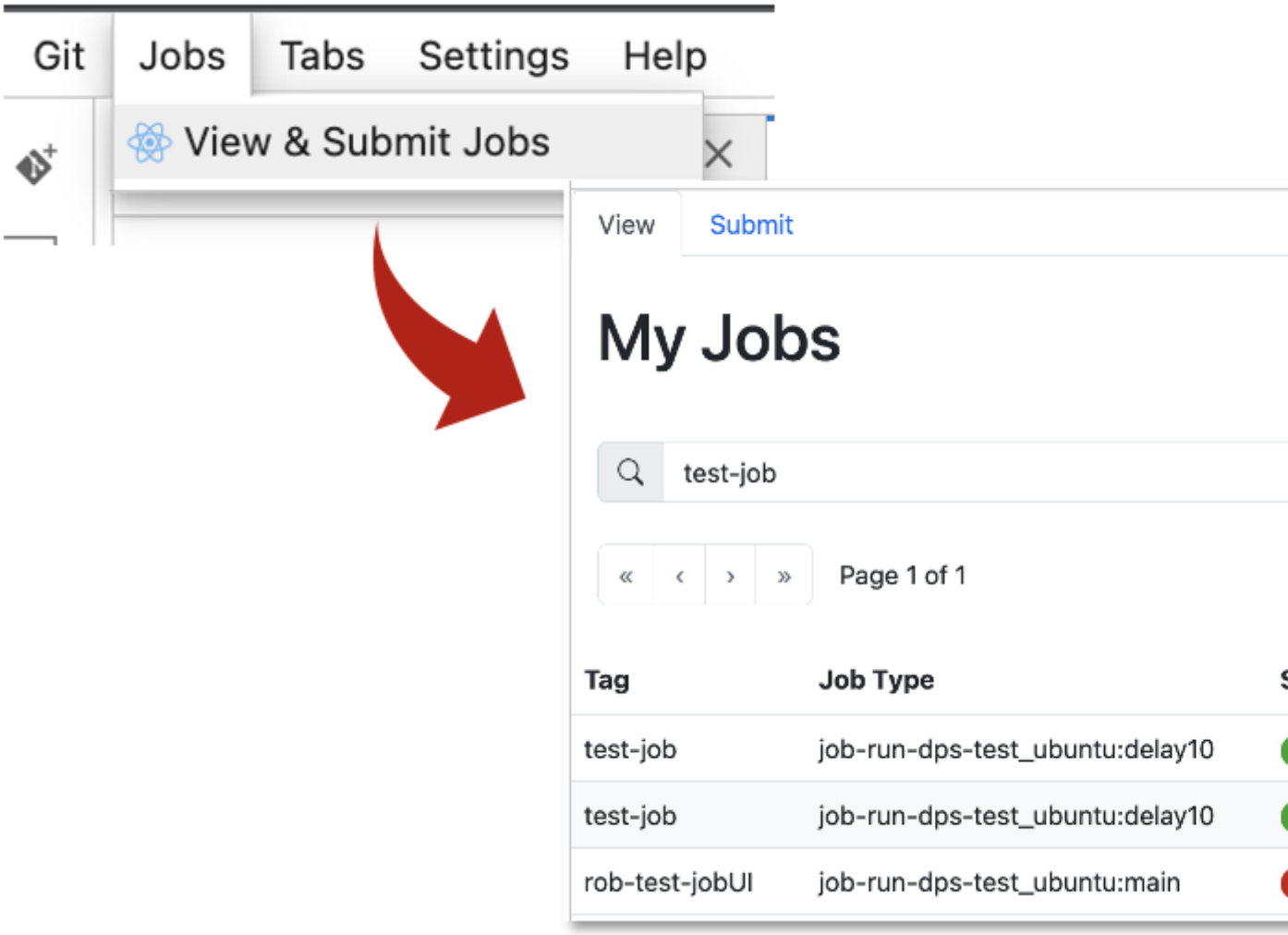
```
'<wps:StatusInfo xmlns:ows="http://www.opengis.net/ows/2.0" xmlns:schemaLocation=
↪"http://schemas.opengis.net/wps/2.0/wps.xsd" xmlns:wps="http://www.opengis.net/wps/
↪2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"><wps:JobID>86fbac52-24b0-
↪4963-8b67-59d0fc09946a</wps:JobID><wps:Status>Succeeded</wps:Status></wps:
↪StatusInfo>'
```

4.9.3 How do I get the outputs of my job?

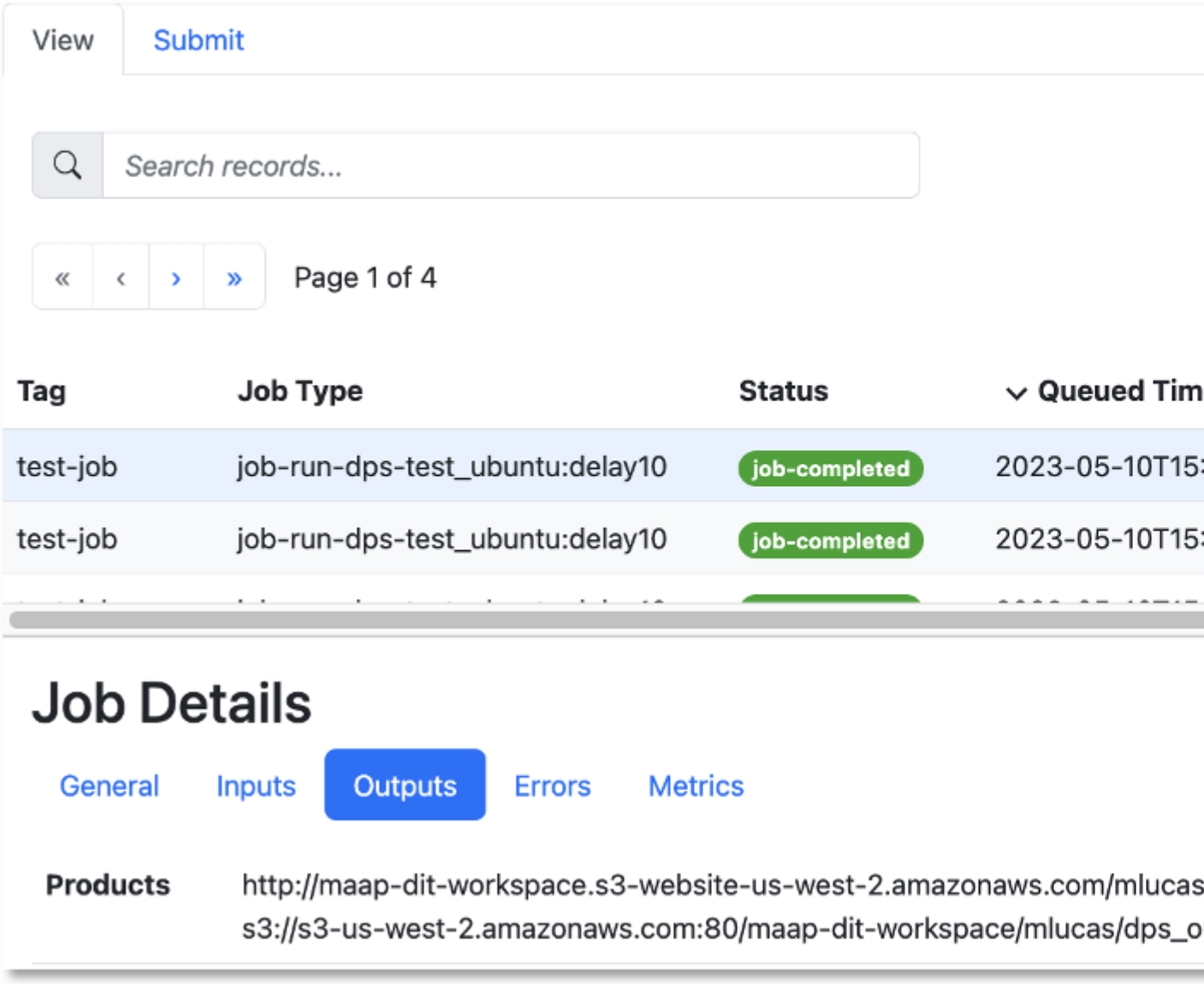
Users may get the outputs of their job using the Jobs UI or the maap-py client library.

Jobs UI

1. Open the Jobs UI from the `Jobs` menu then navigate to the Jobs UI `View` tab.



2. Select your job then under the `Job Details` table, click the `Outputs` tab. The `Products` field provides the path to the product directory within the workspace.

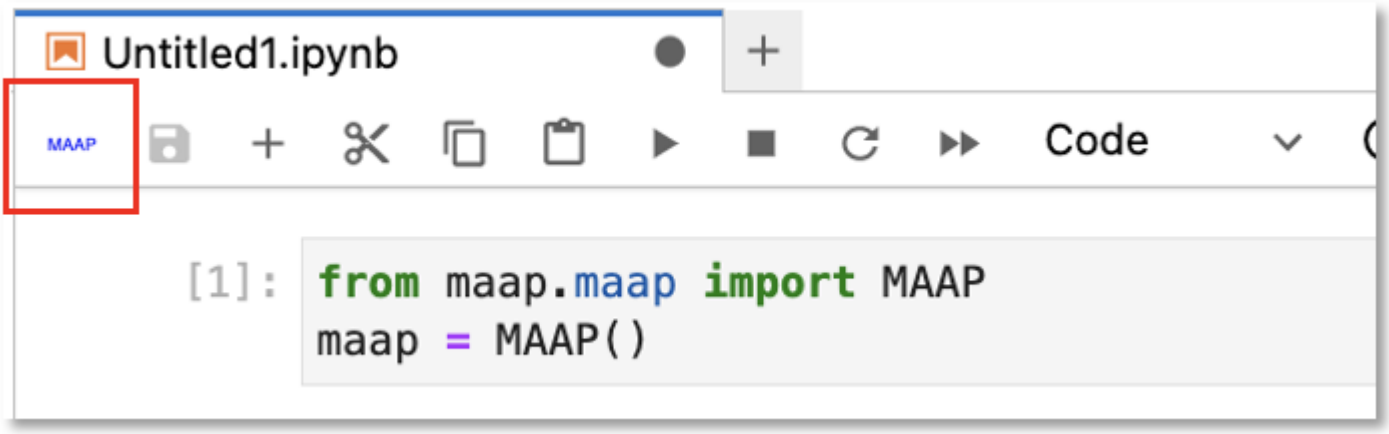


Navigating to the product directory for the selected job above shows the following:

```
(base) root@workspacelbx1uqgeyqdei83w:~/my-private-bucket/dps_output/run-dps-test_
↪ubuntu/delay10/2023/05/10/15/13/27/250000# pwd
/projects/my-private-bucket/dps_output/run-dps-test_ubuntu/delay10/2023/05/10/15/13/
↪27/250000

(base) root@workspacelbx1uqgeyqdei83w:~/my-private-bucket/dps_output/run-dps-test_
↪ubuntu/delay10/2023/05/10/15/13/27/250000# ls
_stderr.txt  _stdout.txt  output-2023-05-10T15:13:27.250000.context.json  output-2023-
↪05-10T15:13:27.250000.dataset.json  output-2023-05-10T15:13:27.250000.met.json ␣
↪write-output.txt
```

MAAP-py Library

1. Open a Jupyter Notebook then click the `MAAP` button from the notebook toolbar. This will paste the code snippet below into your notebook.



2. Provide the MAAP host. For DIT, this would be `api.dit.maap-project.org`

```
from maap.maap import MAAP
maap = MAAP(maap_host='api.dit.maap-project.org')
```

3. Use the `getJobResult` method and provide the job ID that was created upon job submission.

```
r = maap.getJobResult("86fbac52-24b0-4963-8b67-59d0fc09946a")
r.text
```

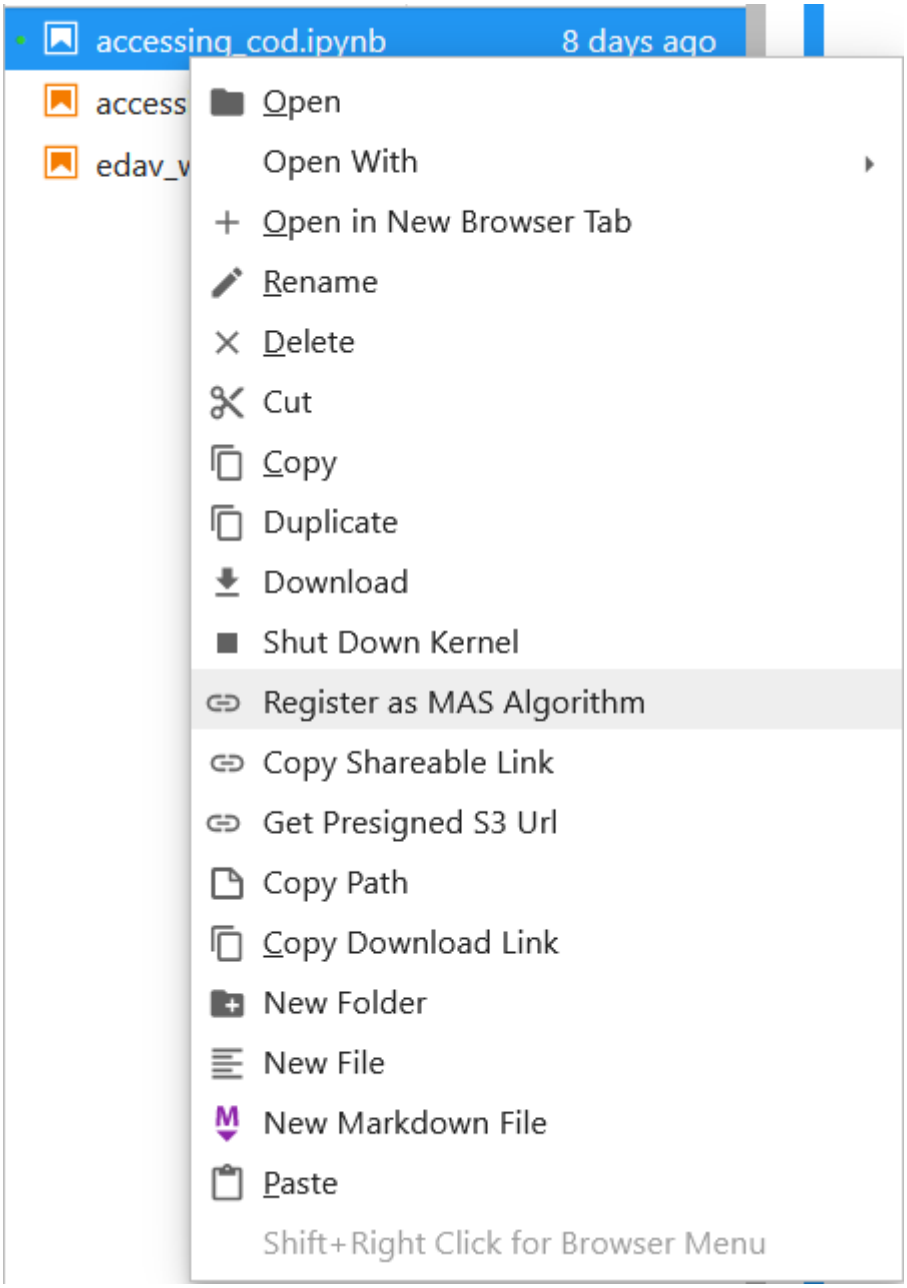
4. Run the notebook to get the job result. The output should resemble the xml snippet below.

```
<wps:Result xmlns:ows="http://www.opengis.net/ows/2.0" xmlns:schemaLocation="http://
↪schemas.opengis.net/wps/2.0/wps.xsd" xmlns:wps="http://www.opengis.net/wps/2.0"␣
↪xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"><wps:JobID>86fbac52-24b0-4963-
↪8b67-59d0fc09946a</wps:JobID><wps:Output id="output-2023-05-10T15:39:51.905070">
↪<wps:Data>http://maap-dit-workspace.s3-website-us-west-2.amazonaws.com/anonymous/
↪dps_output/run-dps-test_ubuntu/delay10/2023/05/10/15/39/51/905070</wps:Data><wps:
↪Data>s3://s3-us-west-2.amazonaws.com:80/maap-dit-workspace/anonymous/dps_output/run-
↪dps-test_ubuntu/delay10/2023/05/10/15/39/51/905070</wps:Data><wps:Data>https://s3.
↪console.aws.amazon.com/s3/buckets/maap-dit-workspace/anonymous/dps_output/run-dps-
↪test_ubuntu/delay10/2023/05/10/15/39/51/905070/?region=us-east-1&amp;tab=overview</
↪wps:Data></wps:Output></wps:Result>
```

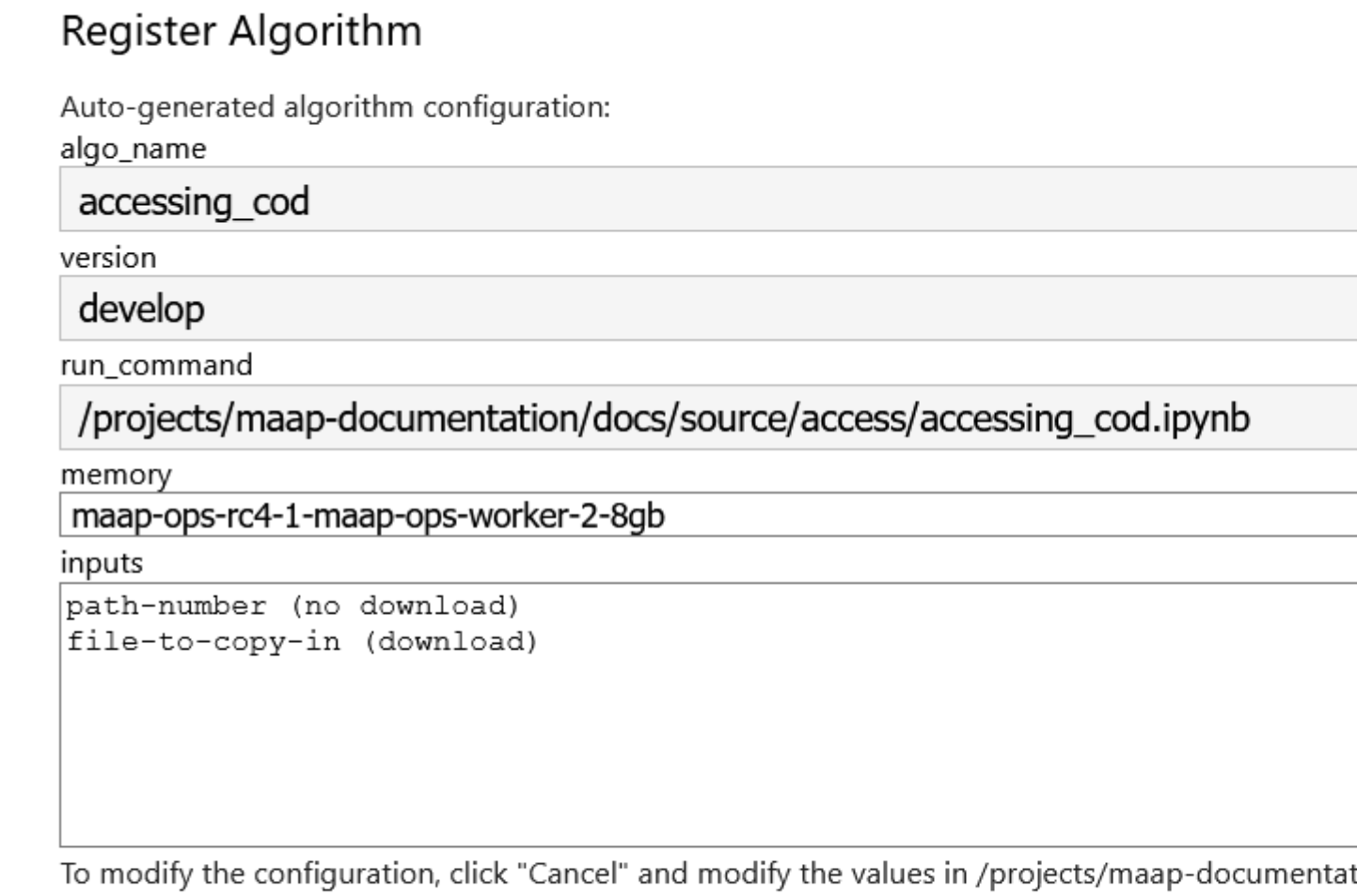
4.9.4 How Do I Register An Algorithm To MAS?

Before starting, make sure your algorithm can run in your environment and all changes have been committed/pushed to GitLab or your external repository.

When you are ready, in the file browser, right click the file which you wish to register as an algorithm and select *Register as MAS Algorithm*.



If your changes have not been committed, there will be a popup asking you to commit your changes first. The fields will be summarized for you in a form.



These are saved in the automatically generated file *algorithm_config.yaml*, in the same directory as the script to be run. Don't register the algorithm if any fields are missing or incorrect.

Click cancel and open the file *algorithm_config.yaml* to edit those fields.

```yaml
# DO NOT DELETE
# THIS CONFIG IS AUTO-GENERATED BY ADE UI
algo_name: accessing_cod
version: develop
environment: ubuntu
repository_url: https://github.com/MAAP-Project/maap-documentation.git
docker_url: mas.maap-project.org:5000/root/ade_base_images/vanilla:${versio
# queue chosen when registering
memory: 15GB

# fill out these fields
# explain what this algorithm does
description:
# path to the wrapper script for running the algorithm
run_command: /projects/maap-documentation/docs/source/access/accessing_cod.
# set a storage value in GB or MB or KB, e.g. "100GB", "20MB", "10KB"
disk_space: 10GB
inputs:
# remove below this line if no inputs
# rename and set algorithm input names accordingly
  - name: path-number
    download: False
  - name: file-to-copy-in
    download: True


```

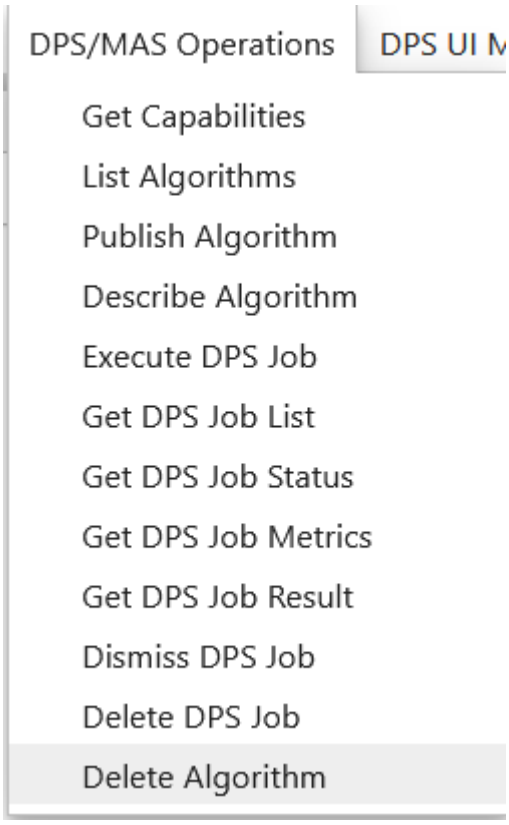
When everything is correct, save your config changes and right-click the file again to register your algorithm.

Notes:

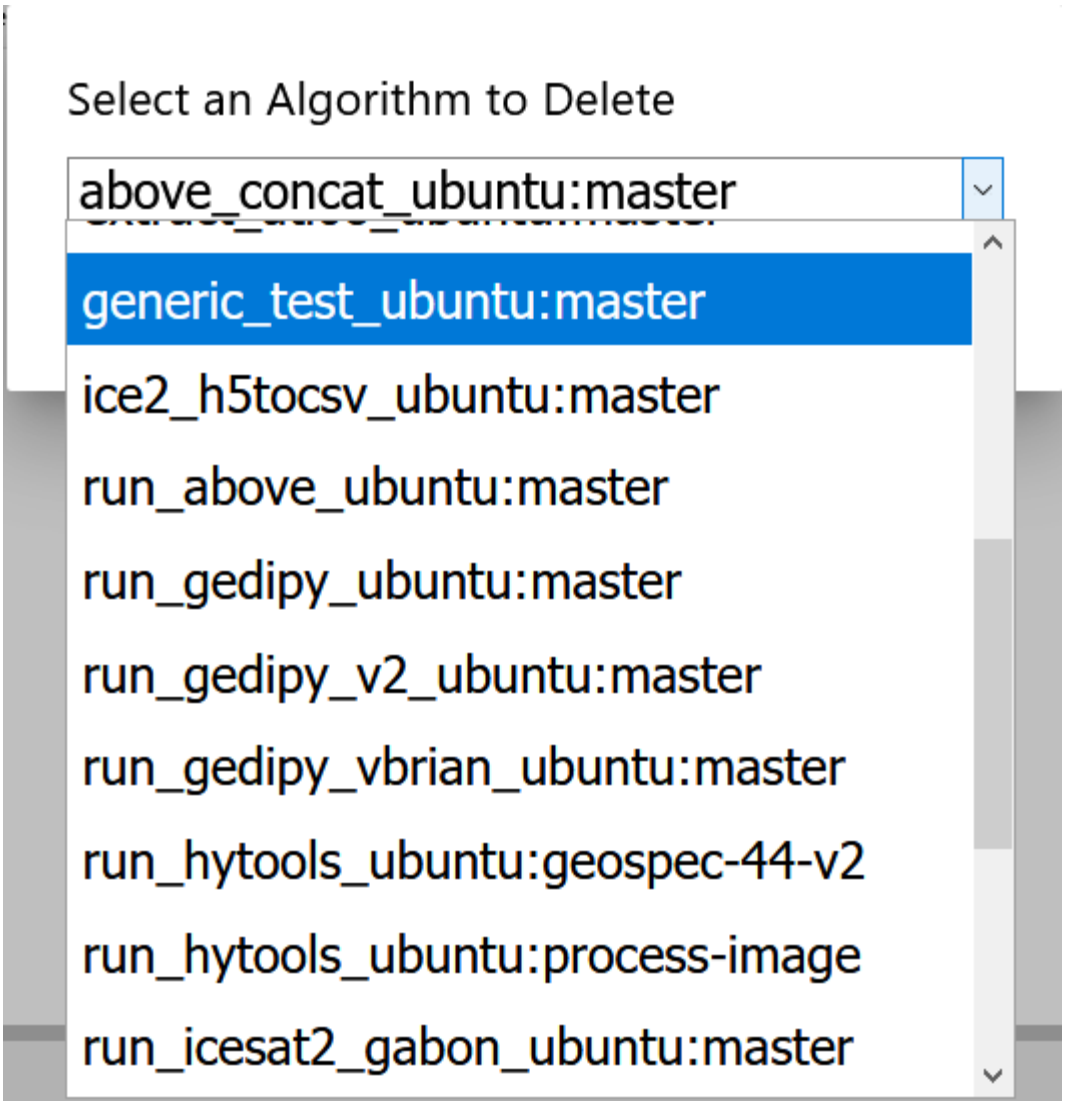
There will only be one *algorithm_config.yaml* file for any directory. If you have multiple scripts to be registered as algorithms, consider separating them into different folders.The outputs should be written to a folder named outputs.

4.9.5 How Do I Delete My Algorithm From MAS?

Open the *DPS/MAS Operations* menu and select *Delete Algorithm*.



A dropdown list of the currently available algorithms will show up.



Select the algorithm you wish to delete and press OK.

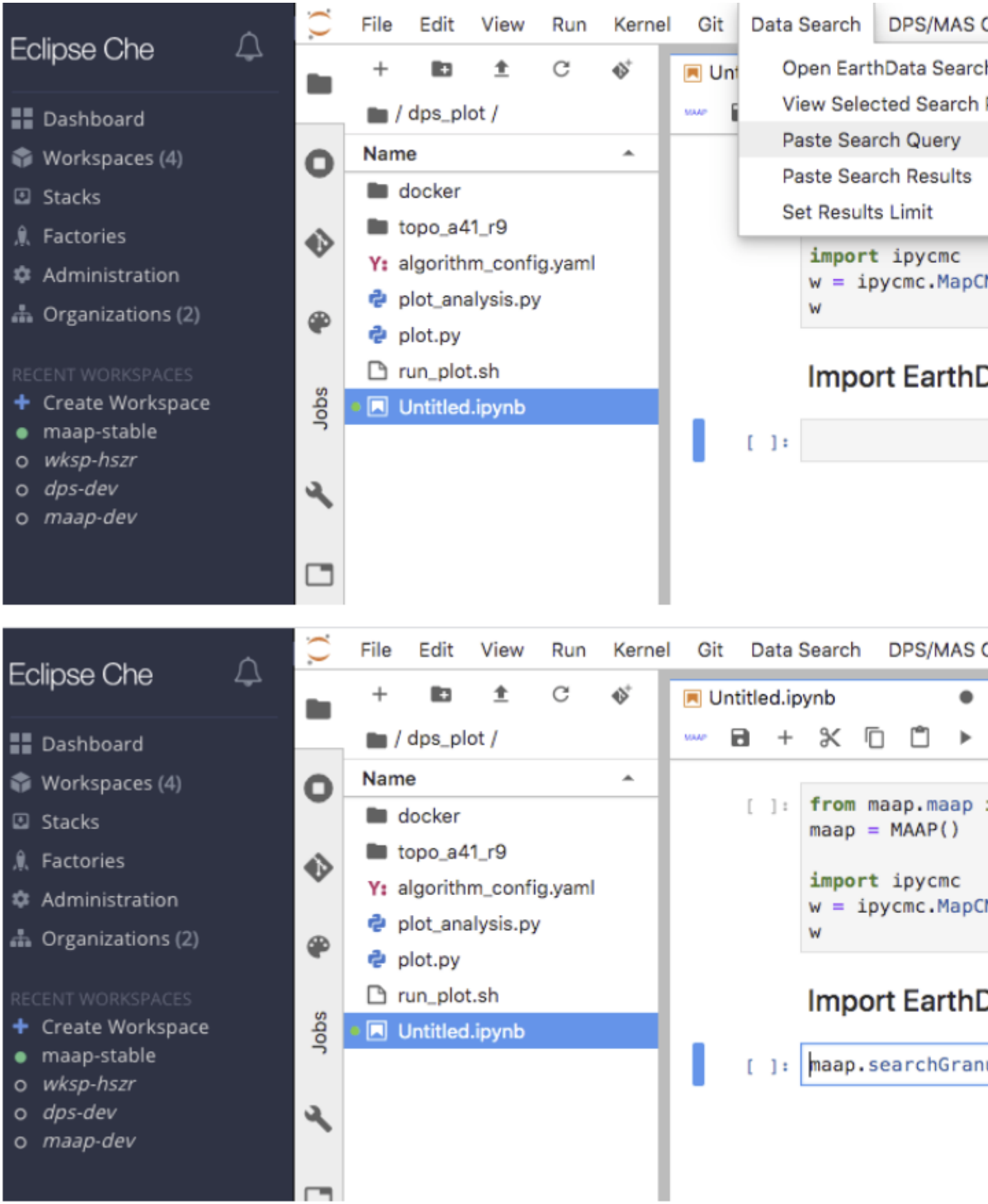
4.9.6 Which Files In My Workspace Are Persistent?

When you first start your workspace, there should be at least three folders already in the file browser: che, maap-users, and <username>.

- che contains config files for your ADE. You don't need to worry about it.
- maap-users is an s3-backed folder shared among all members of the maap-users organization (every registered MAAP user). If you are in other MAAP organizations, each organization will have its own shared s3-backed folder.
- Finally, <username> is also s3-backed, but only shown to you. However, when you share your workspace with another user, both users' personal folders will be mounted and accessible from the shared workspace. Anything placed within a s3-backed folder will be added to s3 and be persistent. Anything outside those folders will be ephemeral. Please clone your git repositories (e.g., dps_plot) outside the s3-backed folders, as they should already be version controlled elsewhere.

4.9.7 How Do I Copy My EARTHDATA Search Into My Jupyter Notebook?

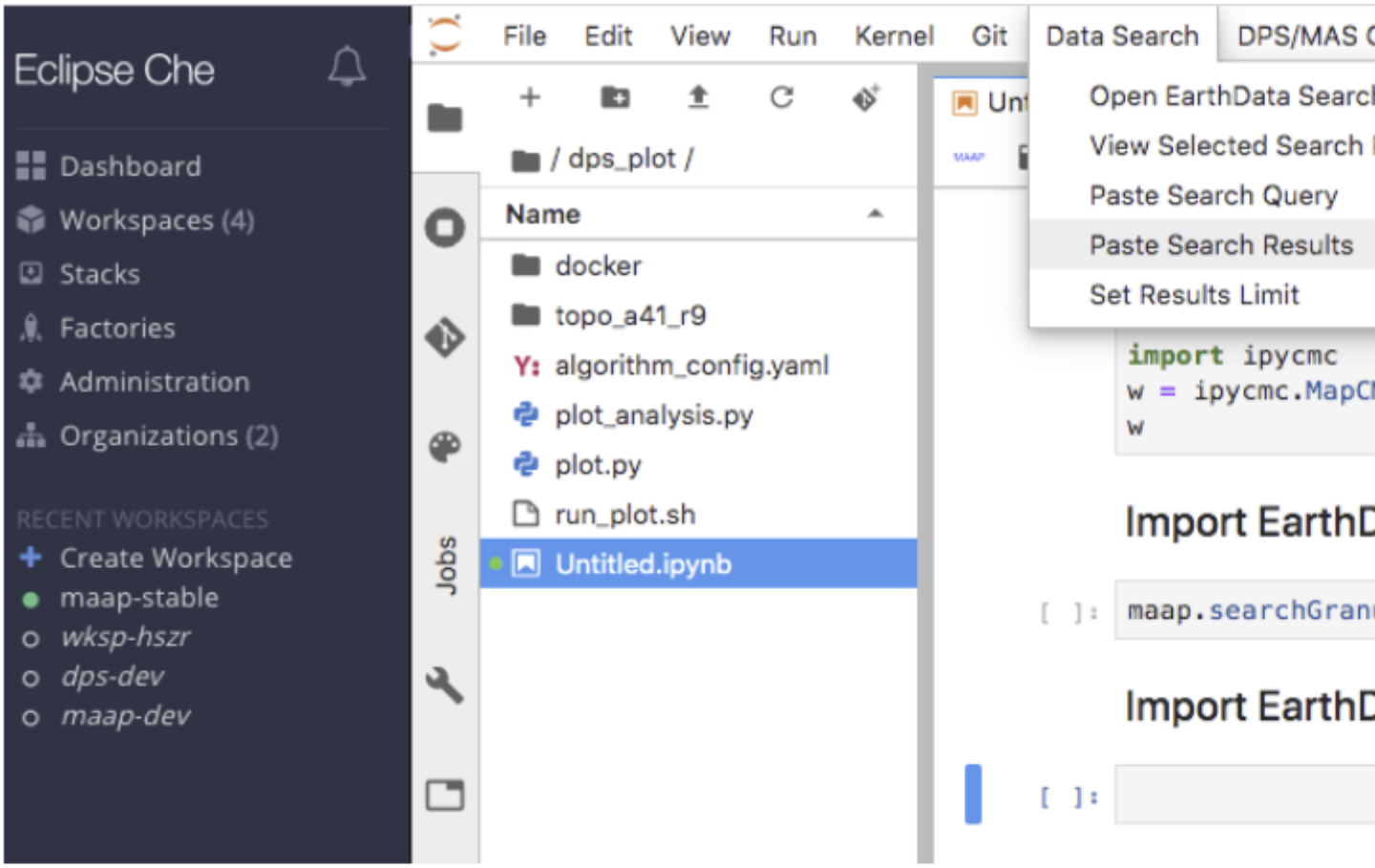
After setting your search parameters, switch tabs back to your Jupyter notebook. At the top, open the Data Search menu, and select *Paste Search Query*.

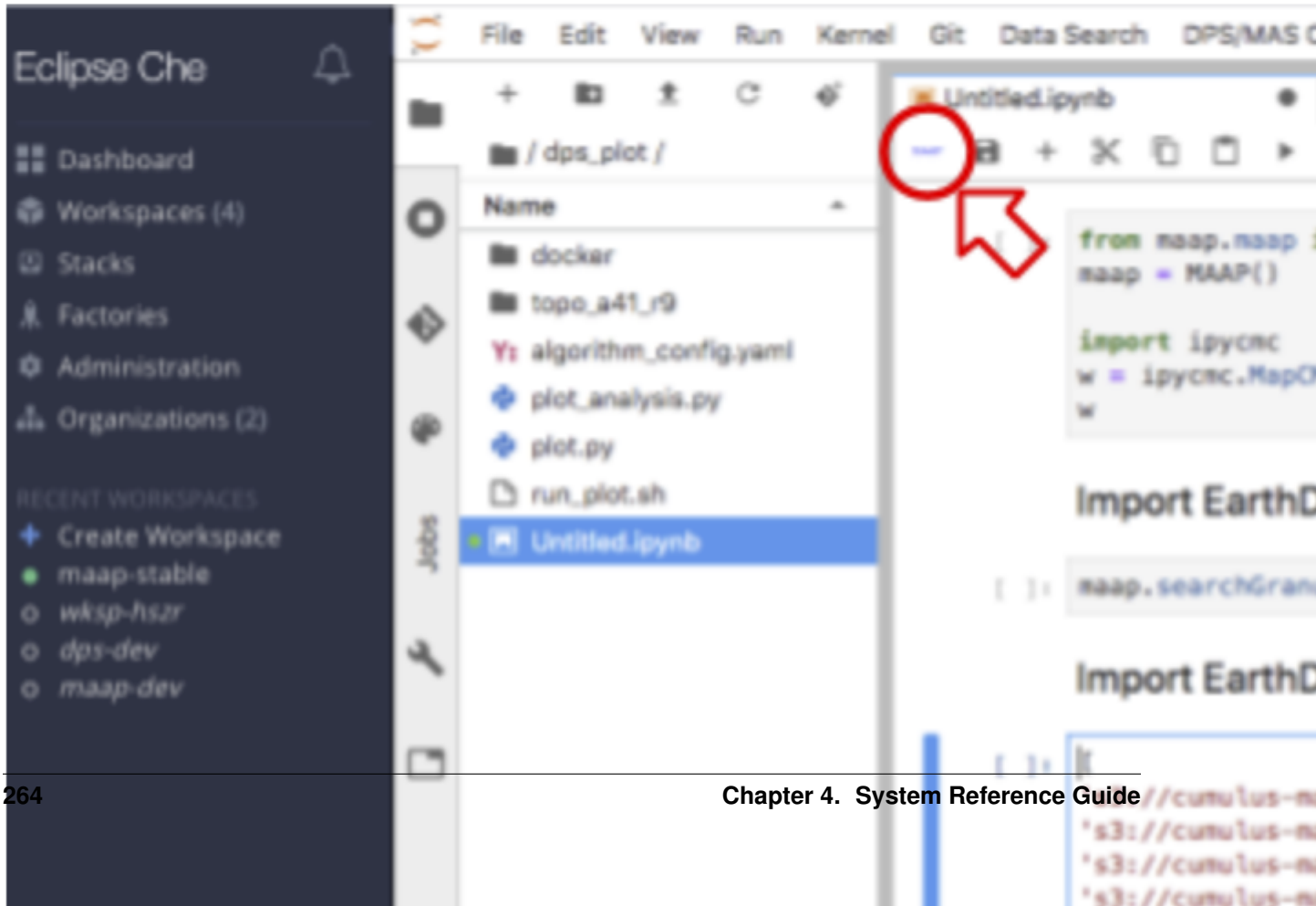

Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
maap-stable
wksp-hszr
dps-dev
maap-dev
File Edit View Run Kernel Git Data Search DPS/MAS
/ dps_plot /
Name
docker
topo_a41_r9
algorithm_config.yaml
plot_analysis.py
plot.py
run_plot.sh
Untitled.ipynb
Jobs
Open EarthData Search
View Selected Search
Paste Search Query
Paste Search Results
Set Results Limit
import ipycmc
w = ipycmc.MapC
w
Import EarthD
[ ]:
Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
maap-stable
wksp-hszr
dps-dev
maap-dev
File Edit View Run Kernel Git Data Search DPS/MAS
Untitled.ipynb
/ dps_plot /
Name
docker
topo_a41_r9
algorithm_config.yaml
plot_analysis.py
plot.py
run_plot.sh
Untitled.ipynb
Jobs
[ ]:
from maap.maap
maap = MAAP()
import ipycmc
w = ipycmc.MapC
w
Import EarthD
[ ]:
maap.searchGran

Caveat: This call uses the MAAP Python library. Make sure you import it before running the inserted code. You can do this by clicking on the blue "MAAP" text just below your notebook name.

4.9.8 How Do I Import Granules Over From My EARTHDATA Search Into My Jupyter Notebook?

After setting your search parameters, switch tabs back to your Jupyter notebook. At the top, open the Data Search menu, and select *Paste Search Results*.


Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
maap-stable
wksp-hszr
dps-dev
maap-dev
File Edit View Run Kernel Git
Data Search
Open EarthData Search
View Selected Search
Paste Search Query
Paste Search Results
Set Results Limit
/ dps_plot /
Name
docker
topo_a41_r9
algorithm_config.yaml
plot_analysis.py
plot.py
run_plot.sh
Untitled.ipynb
Jobs
import ipycmc
Import EarthD
maap.searchGran

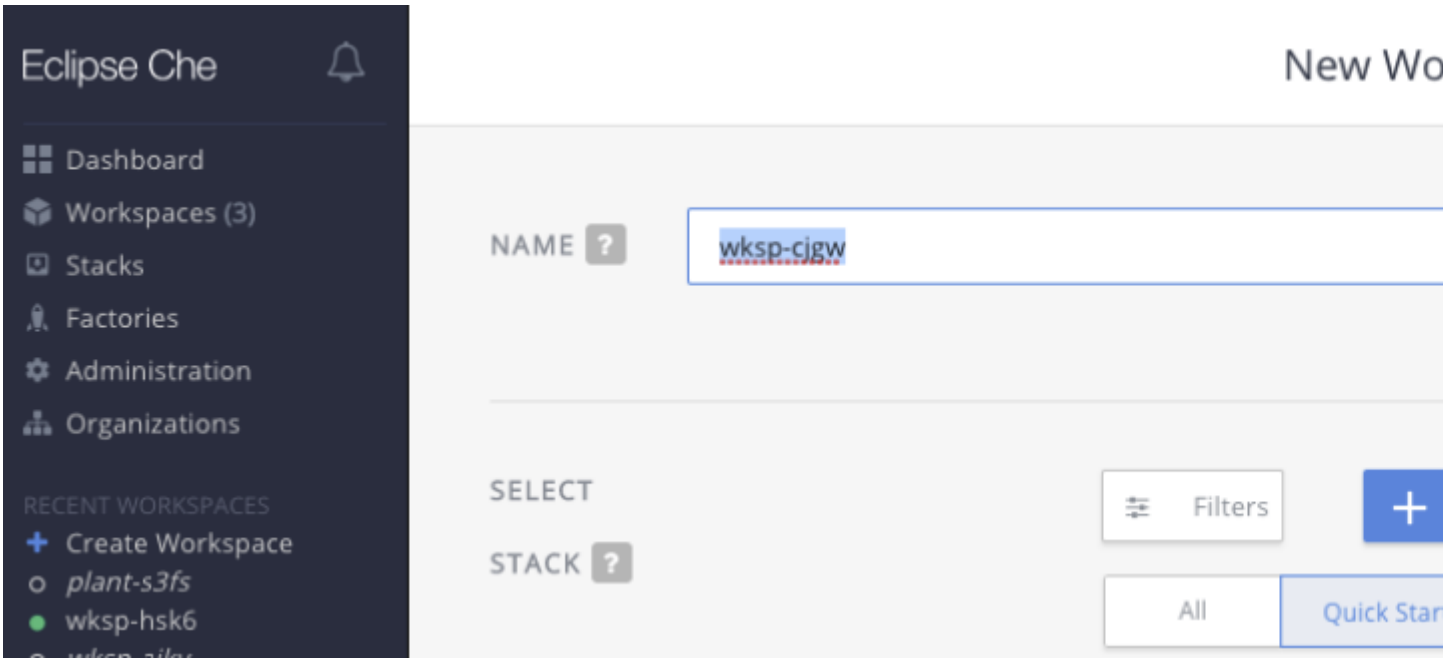

Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
Create Workspace
maap-stable
wksp-hszr
dps-dev
maap-dev
File Edit View Run Kernel Git Data Search
/ dps_plot /
Name
docker
topo_a41_r9
algorithm_config.yaml
plot_analysis.py
plot.py
run_plot.sh
Untitled.ipynb
Jobs
from maap.maap
maap = MAAP()
import ipycmc
Import EarthD

Caveat: This call uses the MAAP Python library. Make sure you import it before running the inserted code. You can do this by clicking on the blue "MAAP" text just below your notebook name (circled in red).

4.9.9 How Do I Rename My Workspace?

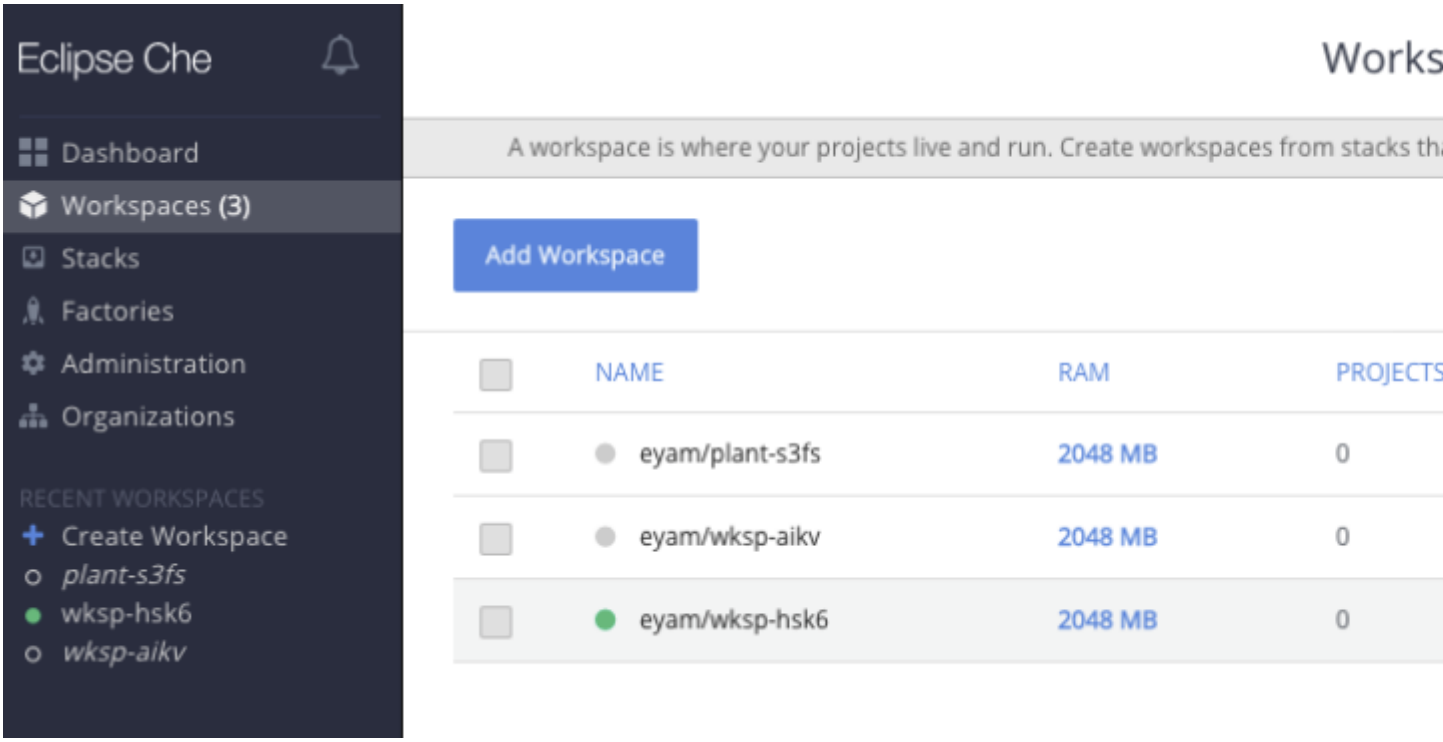
Option 1: At Workspace Creation

You can replace the auto-generated workspace name during creation.

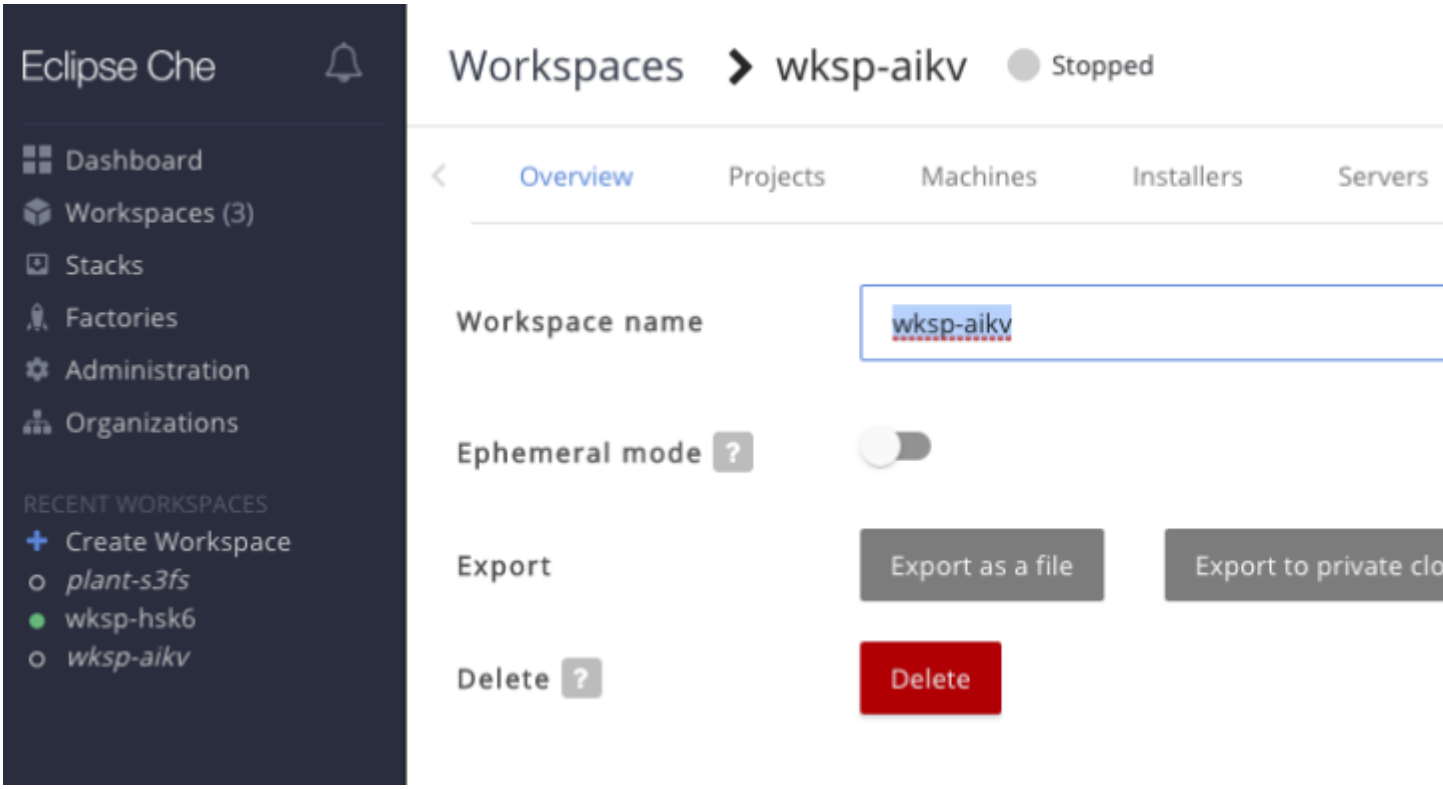


Option 2: Edit Existing Workspace

1. In the workspaces tab under the Che side panel, select the workspace you want to rename.



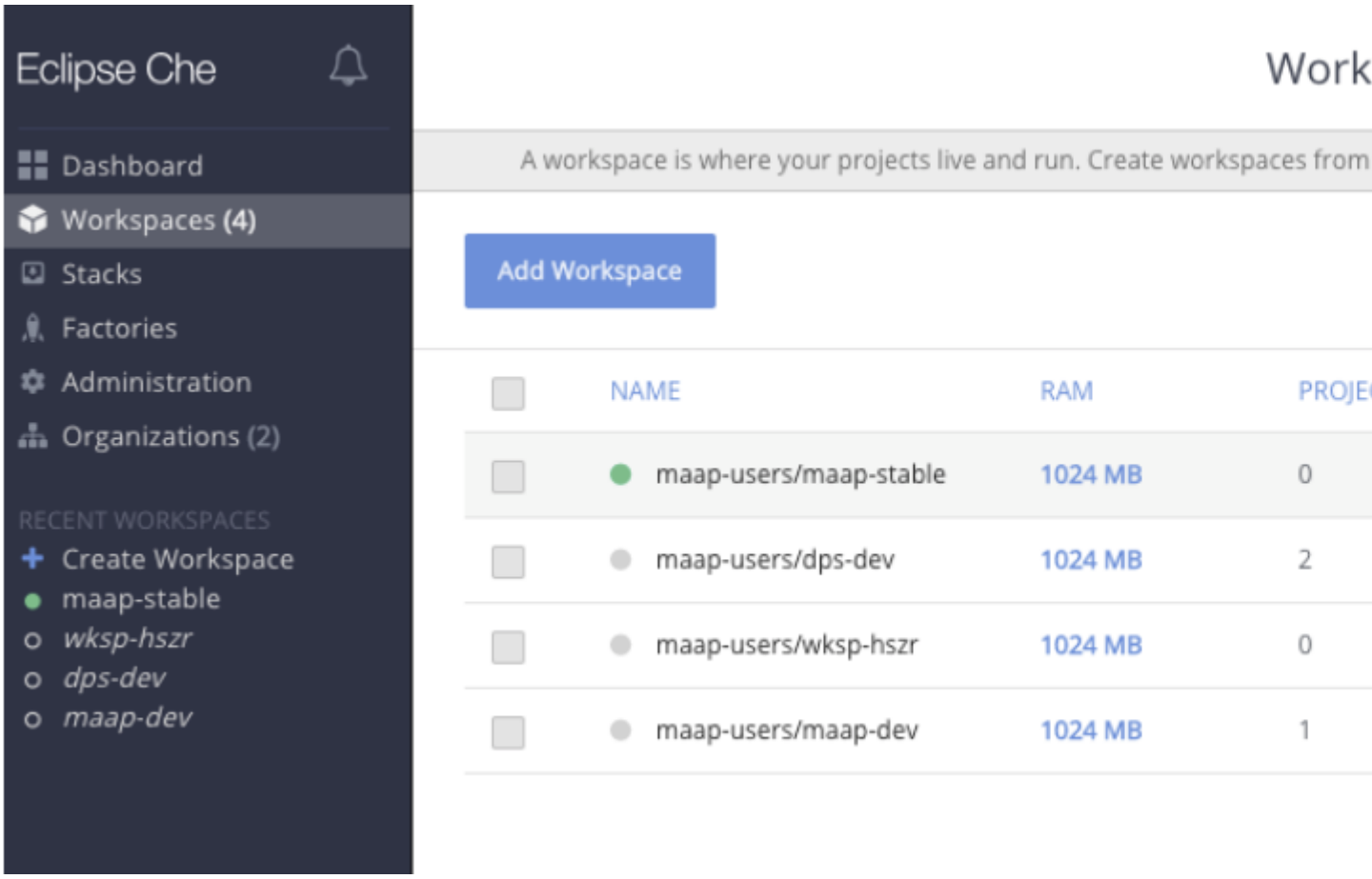
2. Under the Overview section, you can replace the Workspace name field with whatever you wish to name it.



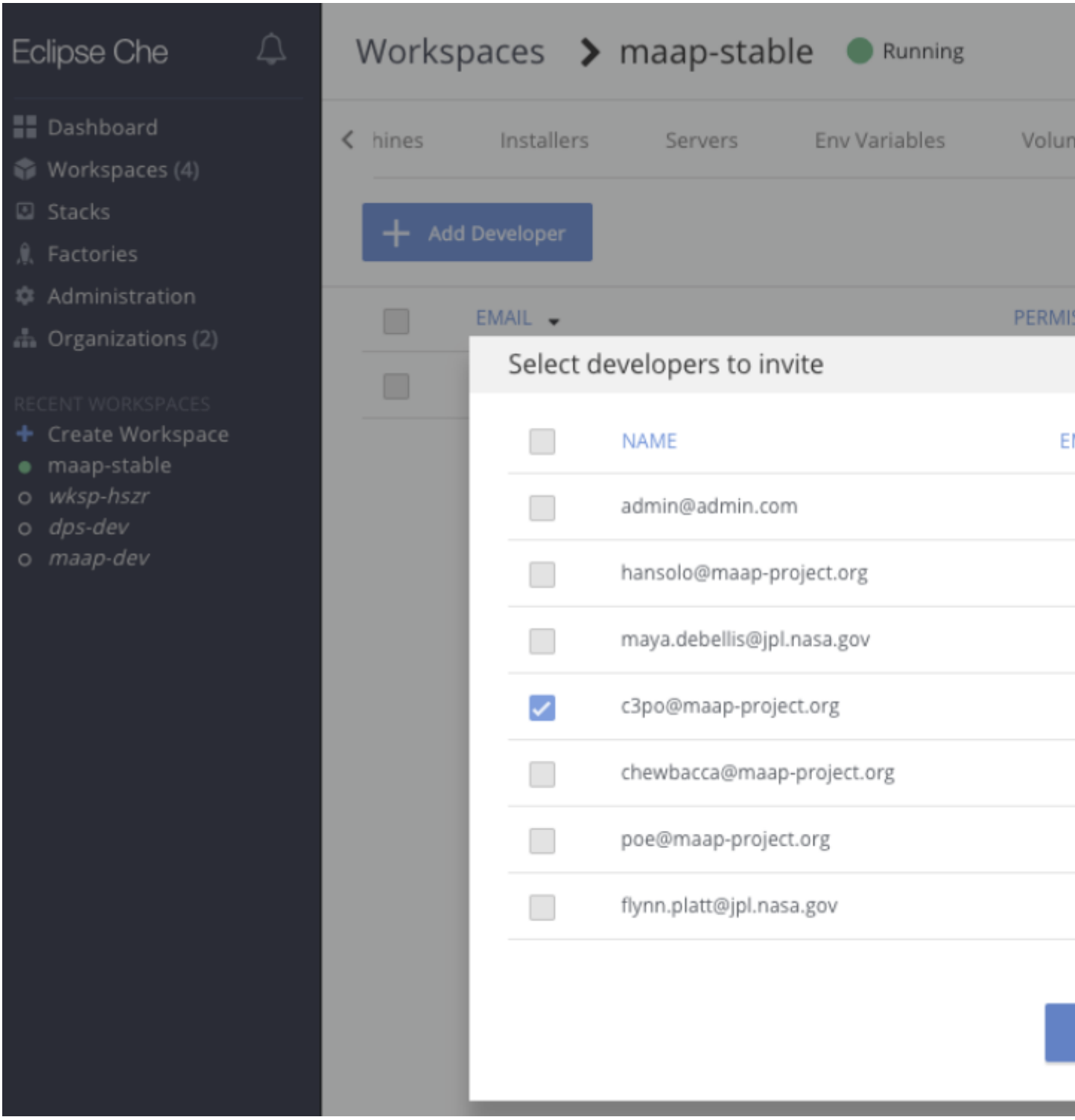
Caveat: no special characters like space, dollar sign, etc; name should start/end only with digits, Latin letters, or underscores

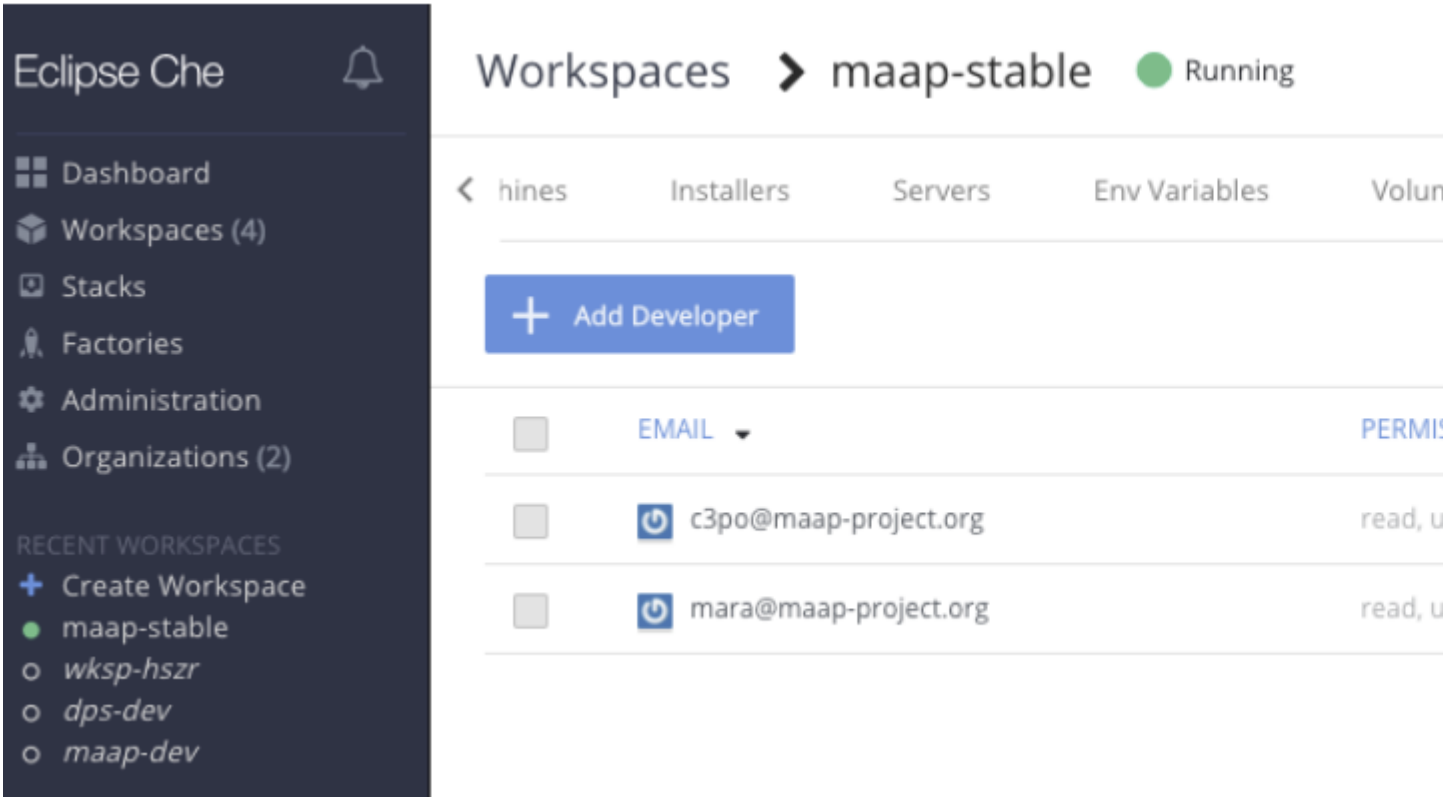
4.9.10 How Do I Share My Workspace With Another MAAP User?

1. In the workspaces tab under the Che side panel, select the workspace you want to share.



2. Under the Share section, click the button + Add Developer. Then check the box next to the email of the person/people you wish to share the workspace with and click *Share*. Only emails registered with MAAP can be used to share workspaces.


Eclipse Che
Dashboard
Workspaces (4)
Stacks
Factories
Administration
Organizations (2)
RECENT WORKSPACES
Create Workspace
maap-stable
wksp-hszr
dps-dev
maap-dev
Workspaces
maap-stable
Running
hines
Installers
Servers
Env Variables
Add Developer
EMAIL
Select developers to invite
NAME
admin@admin.com
hansolo@maap-project.org
maya.debellis@jpl.nasa.gov
c3po@maap-project.org
chewbacca@maap-project.org
poe@maap-project.org
flynn.platt@jpl.nasa.gov



3. Note: If the workspace was opened under an organization, the workspace can only be shared with members of that organization and not any other MAAP user.

CHAPTER 5

Indices and tables

- genindex
- modindex
- search