Food and Agriculture Organization of the United Nations

SEPAL
System for earth observation data access, processing & analysis for land monitoring

Peatlands restoration monitoring working session manual

Bogor, Indonesia
19 – 23 August

Contact:
remi.dannunzio@fao.org
erik.lindquist@fao.org
yelena.finegold@fao.org

Table of Contents

Objective

The objective of this exercise is to create a time stack of soil moisture maps using free and open tools.

This user manual will guide the workshop participants through soil moisture content mapping tools and time series analysis of vegetation indices.

Use of the terminal in SEPAL

One of the ways SEPAL provides access to cloud computing infrastructure is through Amazon Web Services (AWS) cloud computing.
In SEPAL you can select from a variety of cloud computers of various sizes.

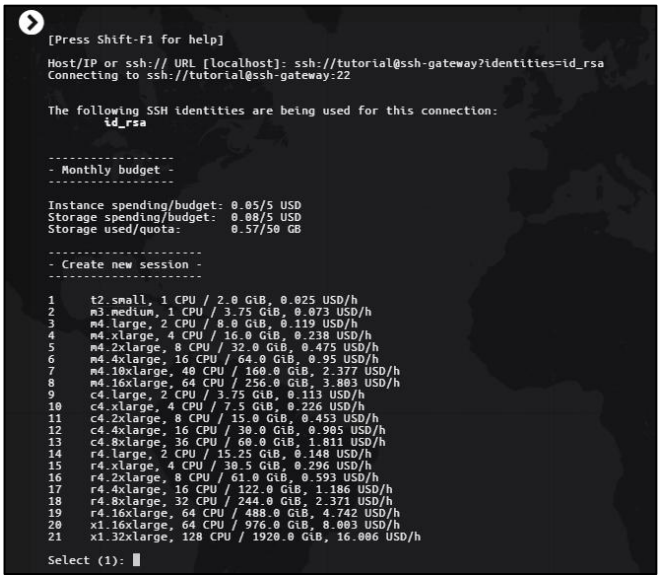
These cloud computers are referred to as **instances** and when they are switched on, you are running a **session**. An instance will automatically start if you launch an application in the processing options. The instance that automatically starts is always the smallest available one, the t2small, which has 1 GB of CPU and 2 GB of RAM power, for a small price of 0.03 USD per hour.

If the session is not active for more than 15 minutes (i.e. you are not using any of the processing tools nor are you actively running anything in the terminal), the instance will be automatically shutdown.

Using the terminal to start an instance

Go to the terminal 

If you don't have any instance running you should see the option to start a new session and the list of instances (see image below)



What type of instance do I need ?

It is good practice to adapt the type of instance to your needs.

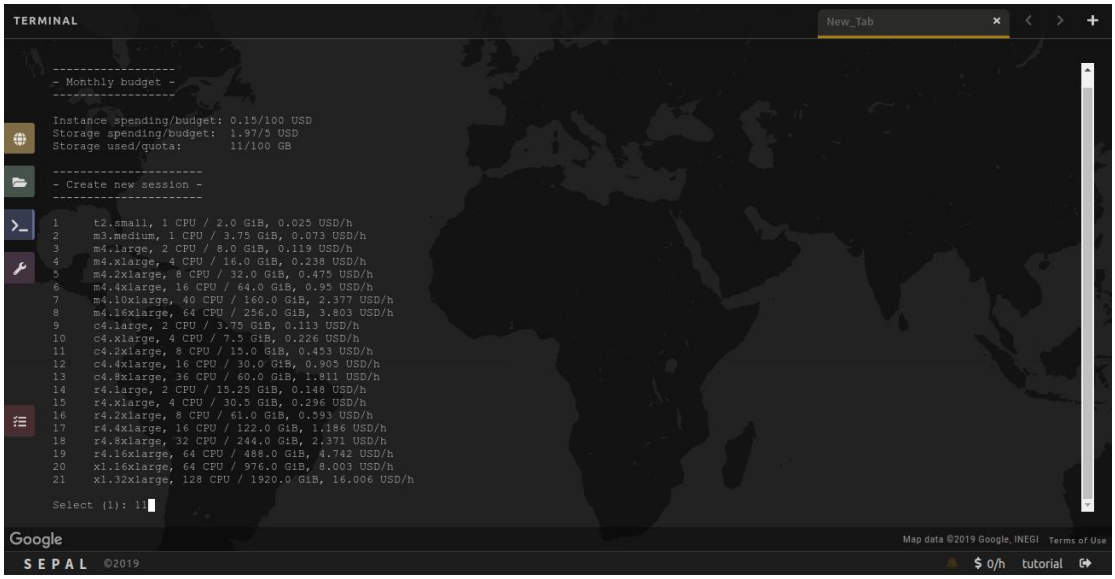
If you are running the soil moisture content processes in Jupyter notebook, you do not need much processing power because the bulk of the processing is sent to Google Earth Engine. You can use a t2.small, instance #1 or m3.medium, instance #2.

If you want to run calculations in R with big objects (national scale raster), you will require a lot of RAM (at least 16GB, instance #4)

The time series analysis application is computationally intensive. You will need an instance with more CPUs and less memory because the algorithm is parallelized among many CPUs and does not require a huge amount of memory. Running a larger instance saves you time and money in your SEPAL account, depending on the size of the area. If the area is large, processing can take days to complete if using a small instance.

For running the time series analysis test data set, you can first try selecting the c4.xlarge. If you are planning on running the time series analysis application on your own area of interest, consider using a larger instance, such as c4.4xlarge or c4.8xlarge

To select the instance you need to type the number corresponding to the instance you would like to start. For example if you are starting the m3.medium, type **2** in the terminal and then hit the ENTER key on the keyboard. The command line only works by typing text in the terminal.



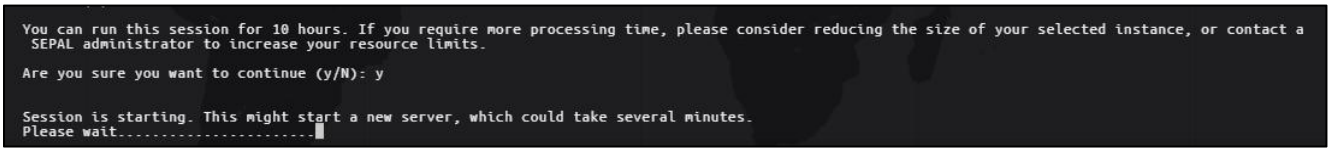
You will then see text about the amount of time you can leave your session running according to your user resources.

i. If your user resources are limited you will be asked if you would like to proceed with starting this instance.

ii. If you would like to proceed, type **y** (which stands for yes) in the terminal. Otherwise, if you would like to select another instance type **n** (which stands for no)**.**

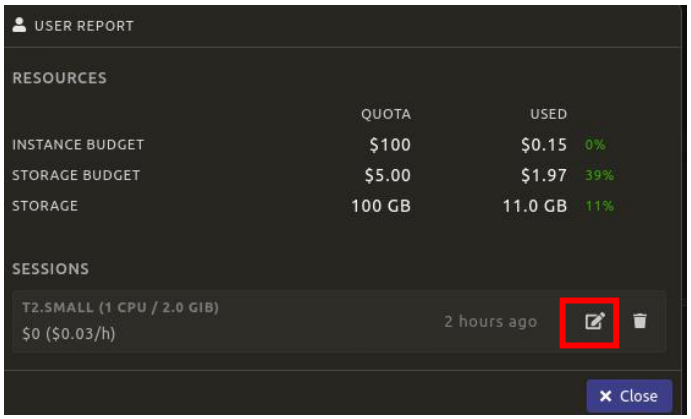
iii. If you have plenty of resources you will not see this text.

It can take a minute to start an instance. You will see text that says ‘Please wait….’

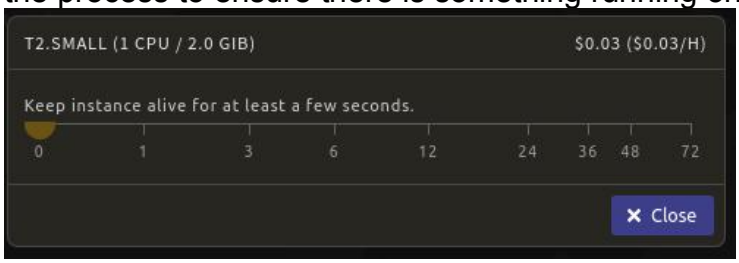


When the session has completed loading you will see your username@numbersandletters$

Keeping an instance active



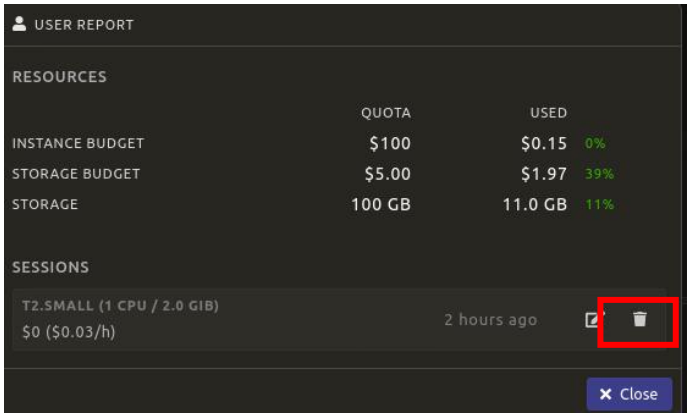
Click on the edit symbol 
Use the slider to choose the amount of time to keep the instance alive. You can keep the instance alive for up to 72 hours. Only use this feature if you are using a process that you know needs to an active instance, such as the time series analysis application or soil moisture mapping. Make sure to check on the process to ensure there is something running on the instance.



Shut down a existing instance

You can manually shutdown an instance in your user resources, which you can access by clicking on the instance budget.

Click the  next to the running instance to shutdown the existing instance.



What can I do in the Terminal ?

SEPAL is running under a Linux server with the latest long term release Ubuntu operating system and the corresponding Bash Shell. You can run all the basic Linux commands in the terminal. In addition, several geospatial processing libraries are available such as GDAL/OGR (www.gdal.org) or the Orfeo Toolbox (www.orfeo-toolbox.org) and you can run processing commands from the terminal.
You can also clone, update and push git repositories directly in the terminal as illustrated in the below examples.
Clone a repository:

```
git clone https://github.com/yfinegold/ws_idn_20190819
```

Go to a repository:

```
cd    ~/ws_idn_20190819/
```

Update a repository:

```
git pull
```

Connecting your Google account to SEPAL

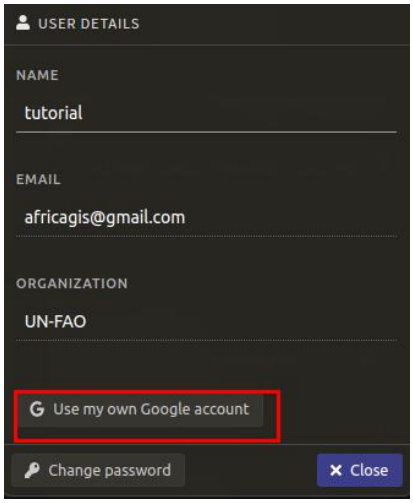
Some of the core functionality of SEPAL utilizes Google Earth Engine (GEE). To take full advantage of SEPAL you need to connect your Google account to your SEPAL account.

First, make sure you have a Google account and it is whitelisted in GEE.
If you do not have a GEE enabled account, sign up here:
https://signup.earthengine.google.com/

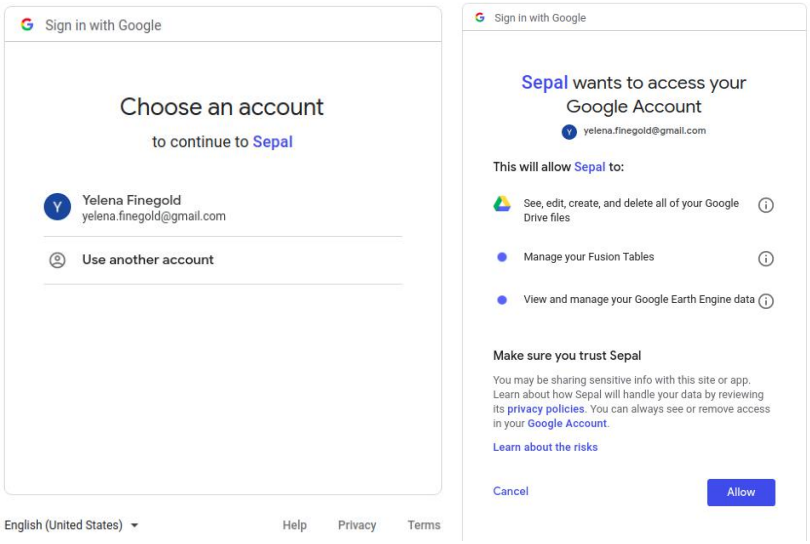
After you get an email confirming your access to Google Earth Engine you can connect your SEPAL account to Google.

Click on your user name in the bottom right corner

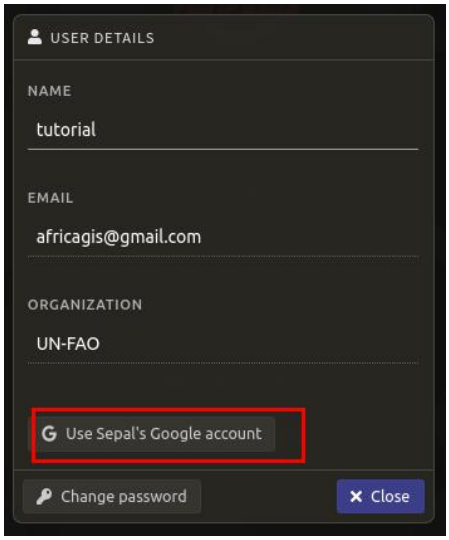
If you see ‘Use my own Google account’, your Google account is not yet connected to SEPAL. In the following steps you will connect your account. First, click on ‘Use my own Google account’. This will redirect you to a Google sign in page.



Select the Google account that is already approved for use of GEE. Sign into your account with your Google password. Then click on ‘Allow’ to grant SEPAL access to your Google account.



After clicking ‘Allow’, you will be automatically redirected to the SEPAL page. You can check your Google account connection by clicking on your user name again. It should now say ‘Use SEPAL’s Google account’. Do not click on this button, it will disconnect the connection between your personal SEPAL account and your personal Google account.



Soil moisture content mapping

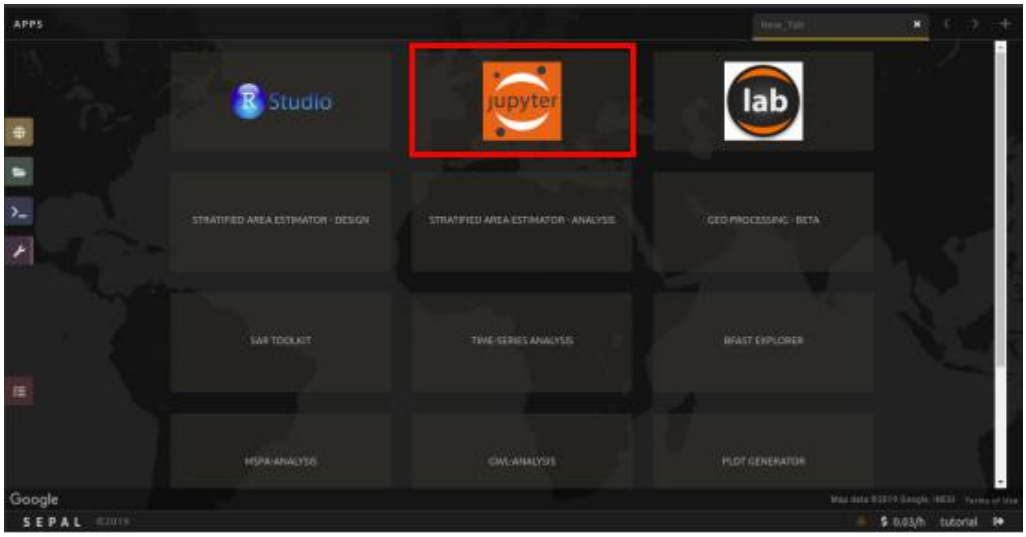
The soil moisture content module creates soil moisture maps using the PYSMM process developed by, Felix Greifeneder, read more about PYSMM here: https://pysmm.readthedocs.io/en/latest/.

The estimation of soil moisture is based on a Support-Vector-Regression machine learning approach. The model training was performed based on in-situ data from the International Soil Moisture Network (ISMN). PYSMM all processing steps for spatial and temporal mapping of surface soil moisture are fully executed online on GEE - and then the data sets are transferred to your SEPAL account.

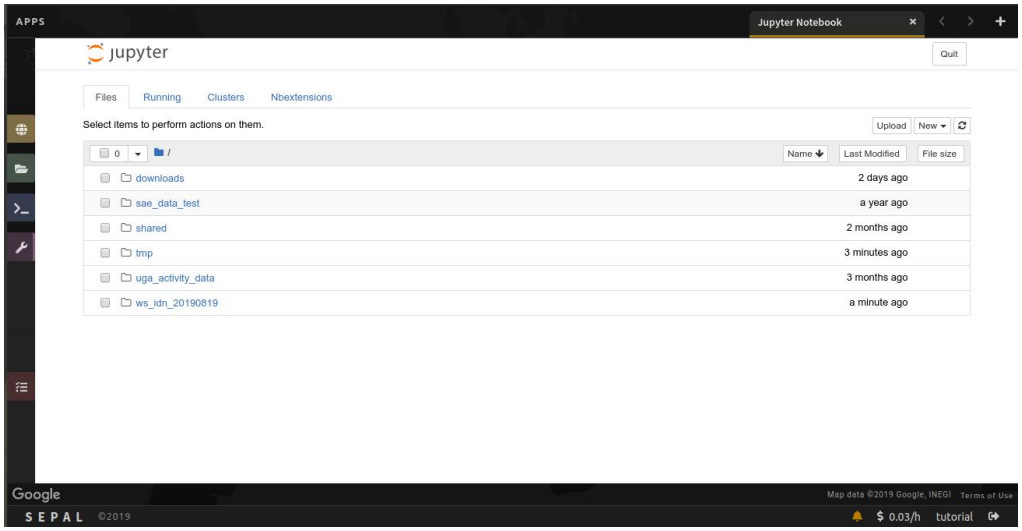
Jupyter notebooks interfaces have been created to processing the soil moisture maps for an area of interest. The area needs to be uploaded as a Google Earth Engine asset. To upload an asset follow the instructions here: https://developers.google.com/earth-engine/importing

The PYSMM modules are part of the ws_idn_20190819 folder which was downloaded using the git clone command in the command line. It will be necessary to update this folder periodically using the instructions in the Terminal section of this manual.

To access the modules click on the processing tab  in SEPAL. Then click on Jupyter notebooks.

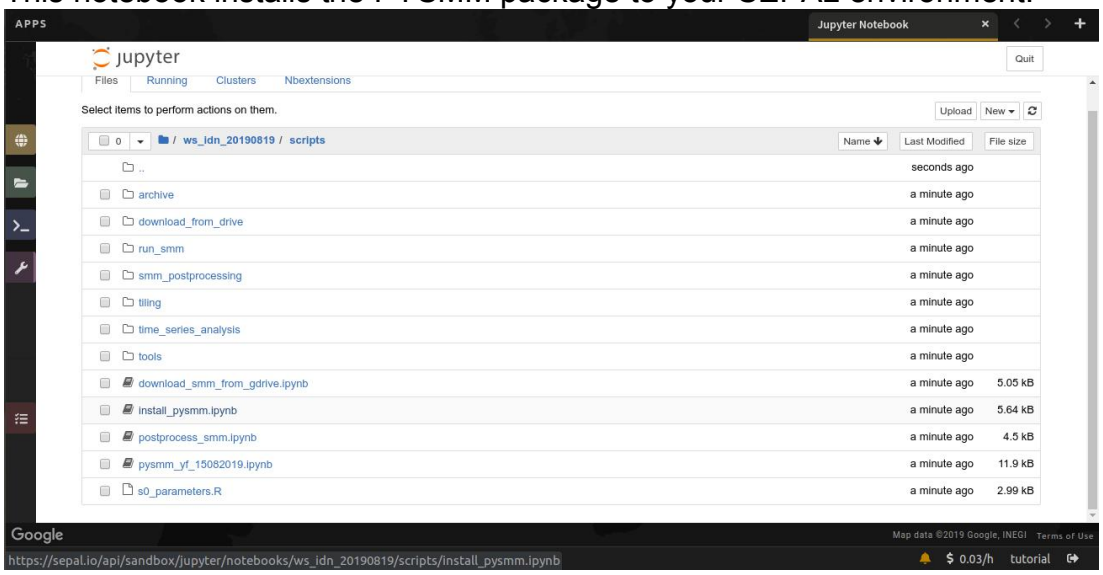


Then navigate to the ws_idn_20190819 folder by clicking on it. Then click on the scripts folder.

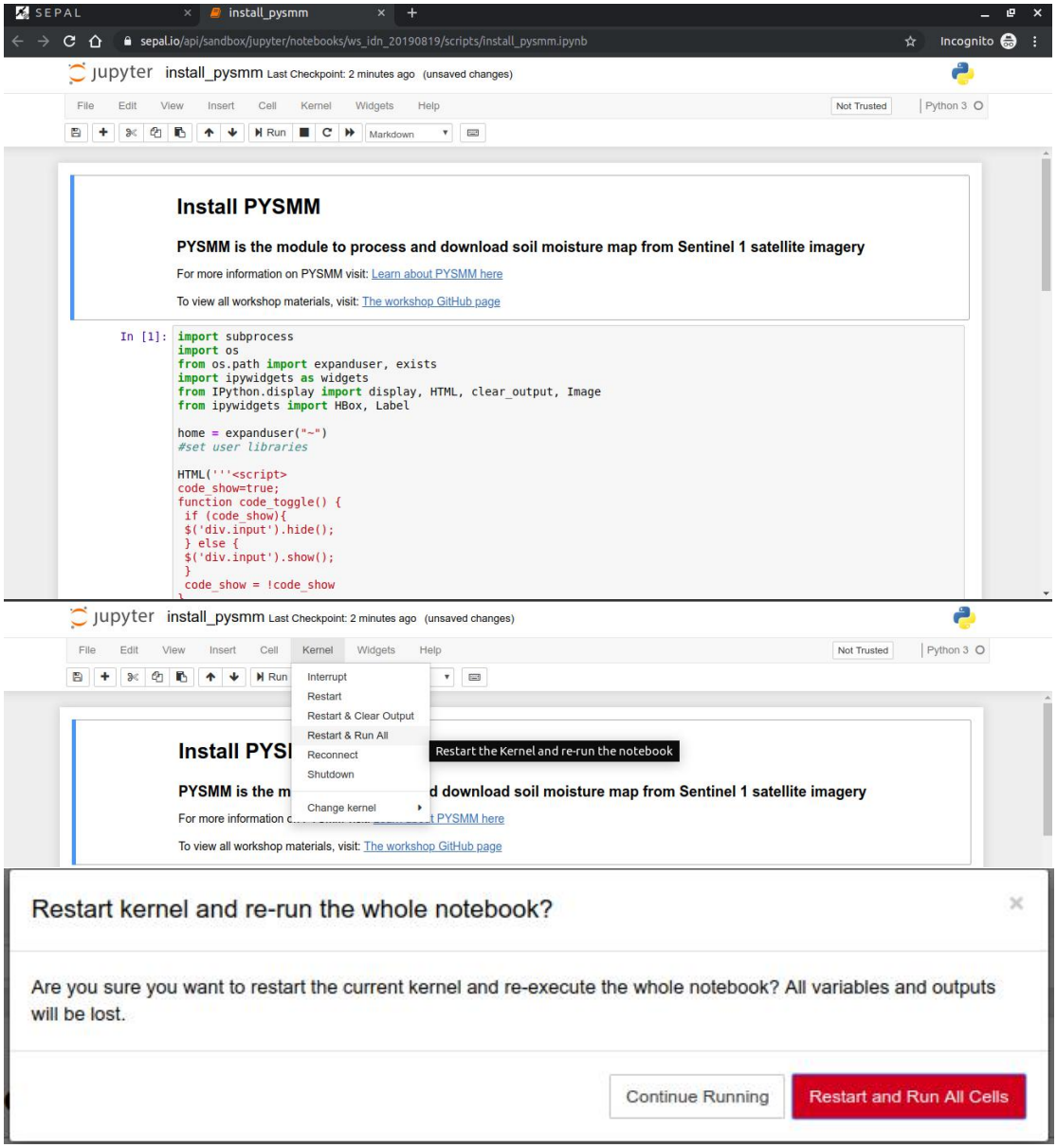


Installing PYSMM

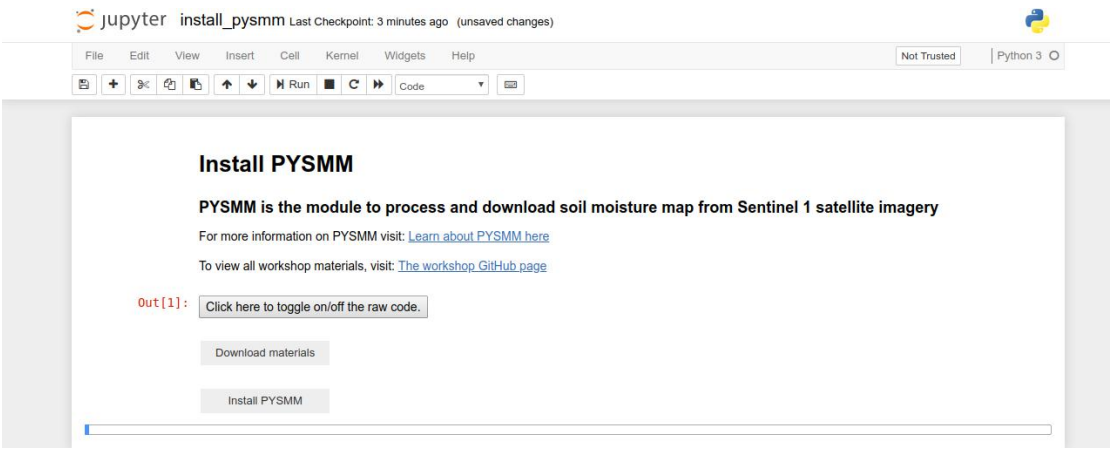
Inside the scripts folder there are python notebooks, which end with the .ipynb extension. Click on the install_pysmm.ipynb to open the first python notebook. This notebook installs the PYSMM package to your SEPAL environment.



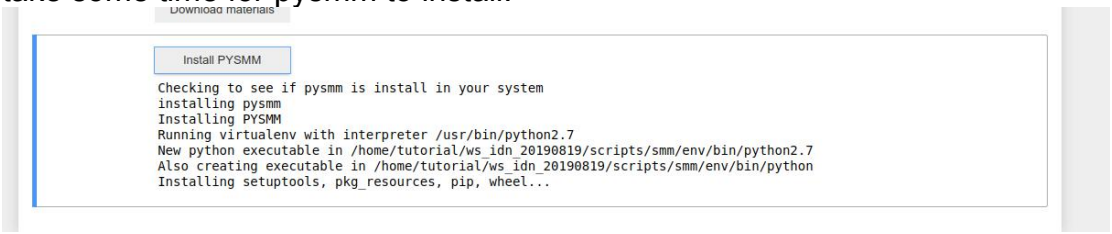
The jupyter notebook will open in a new tab. First you need to restart the kernel and run all. To go this click on ‘Kernel’ and then click on ‘Restart & Run All’. Wait a moment for the notebook to initialize.



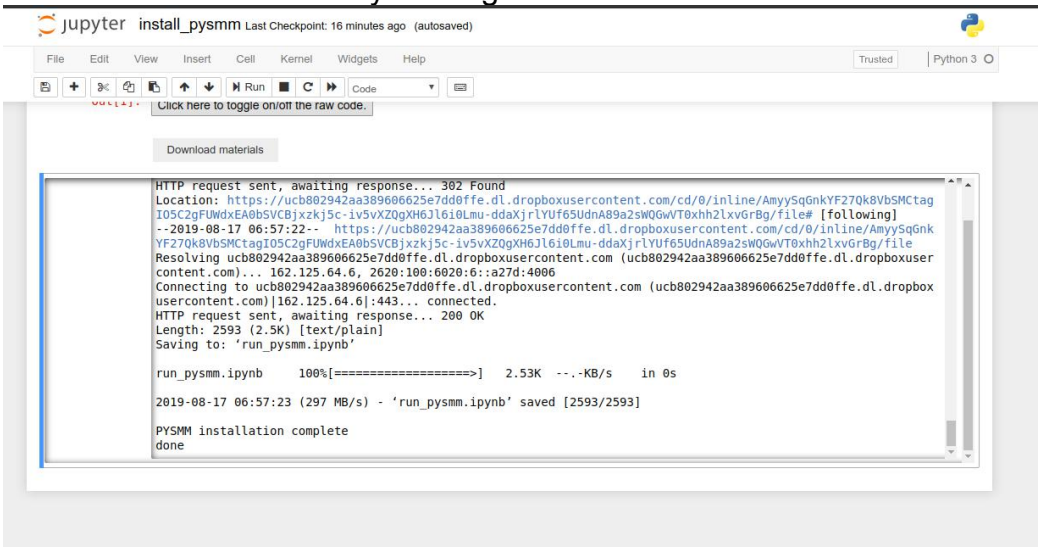
Once the notebook is initialized, the code should not be visible and only the user interface remains. There is the option to toggle the code on and off. If you are interested in seeing the code you can toggle the raw code, although this is not necessary.
To install pysmm click on the ‘Install PYSMM’ button. You will only need to install pysmm one time.



You will see the installation text appear with updates about the process. It will take some time for pysmm to install.

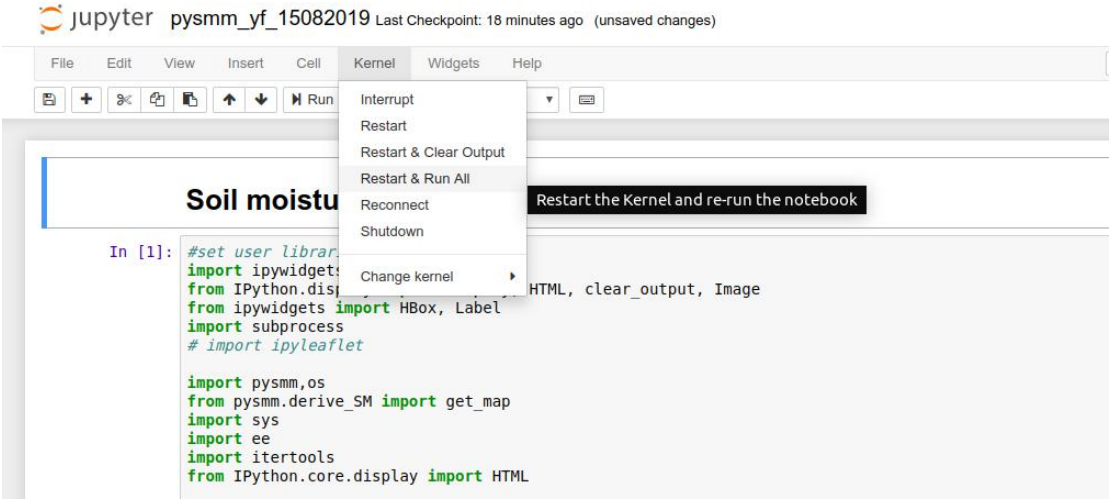


When the installation is completed, scroll down to the bottom of the installation updates and it will say, ‘done’. When you complete the installation you can shut down the notebook by clicking file->close and halt.

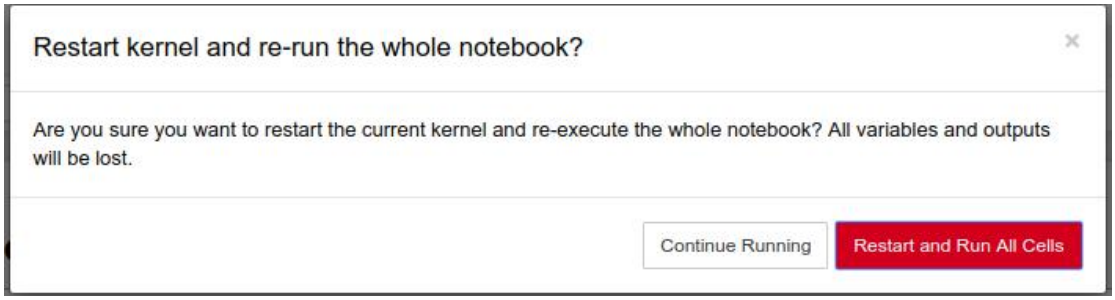


Processing soil moisture

The python notebook to process soil moisture is called run_pysmm_ui.ipynb. Open the notebook from jupyter notebooks and reset the interface by selecting 'Kernel'->'Restart & Run All'

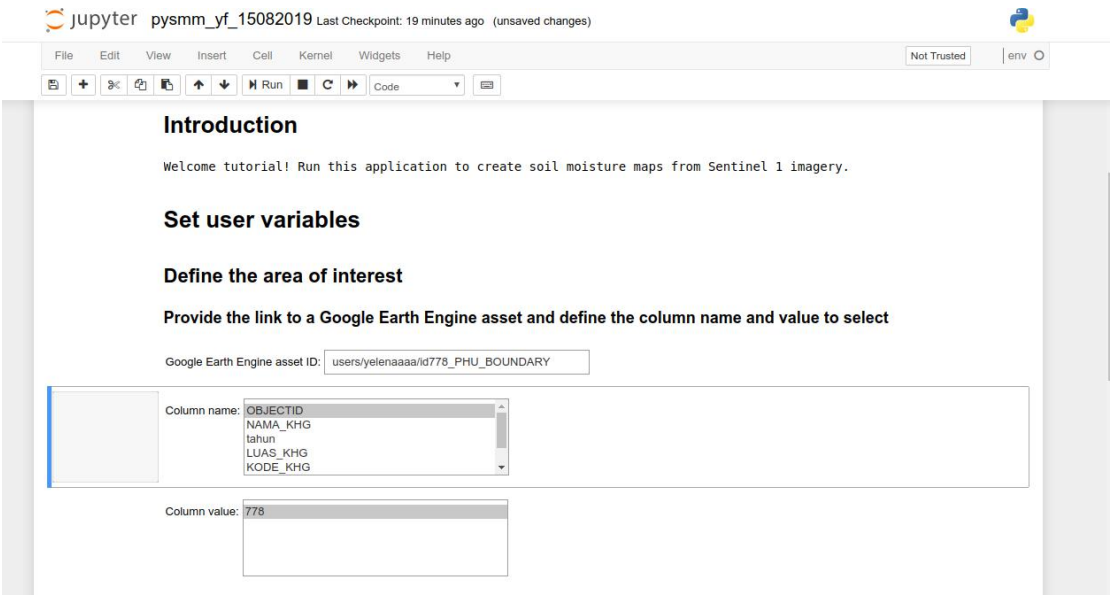


Click on Restart and Run All Cells button

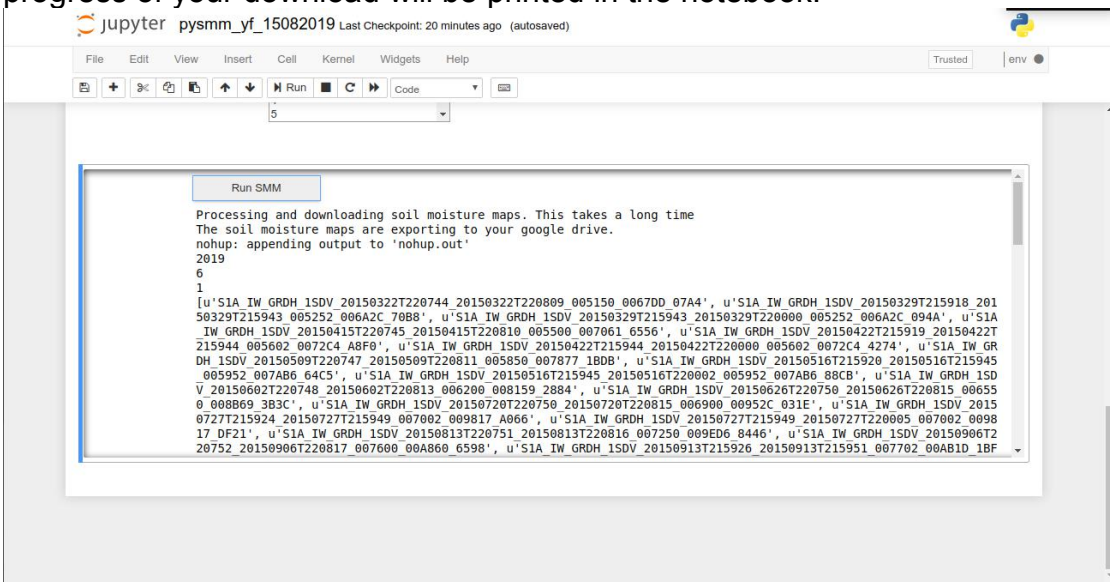


The area of interest is specified using a Google Earth Engine table asset. A Google Earth Engine table asset is vector data, such as a shapefile, uploaded to your GEE account. You can change the area of interest by pasting your GEE asset id into the text box. If you change the asset id, click on the white space left of the column name box and type CTRL and ENTER at the same time to read your area of interest file and load the custom column names. The area of interest is filtered by rows in the attribute table. To select the column value, first select the row you want to chose the value from. Then click left of the column value box and type CTRL and ENTER at the same time to read your area of interest file, selected column names and load the values from that column.

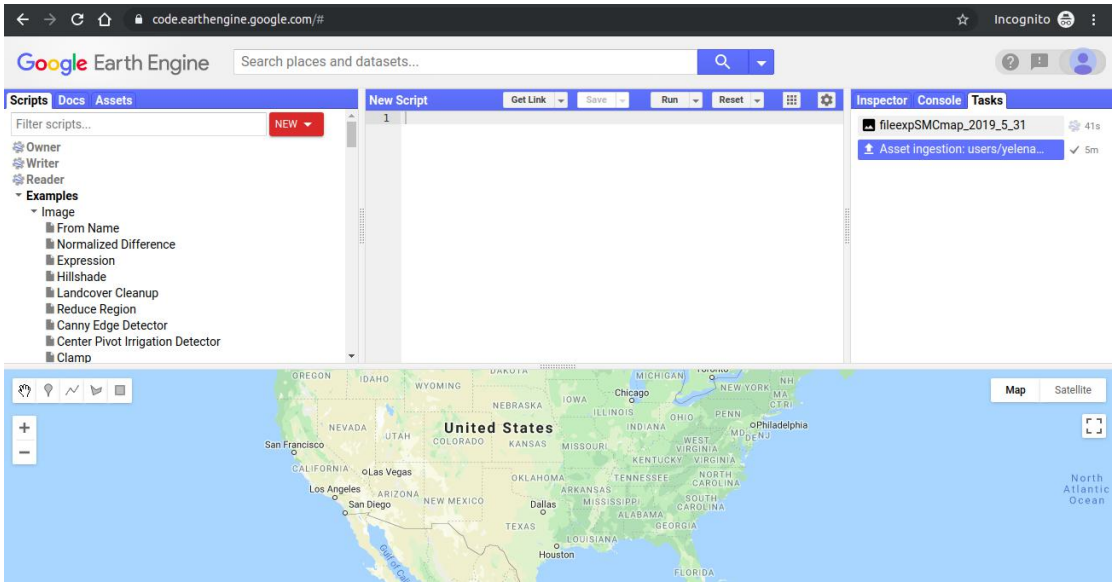
Next select the date parameters for downloading the soil moisture maps.



After all the parameters have been customized you can run the soil moisture module. This will send the command to Google Earth Engine to run the PYSMM process. Your data will be downloaded into your Google drive. The progress of your download will be printed in the notebook.



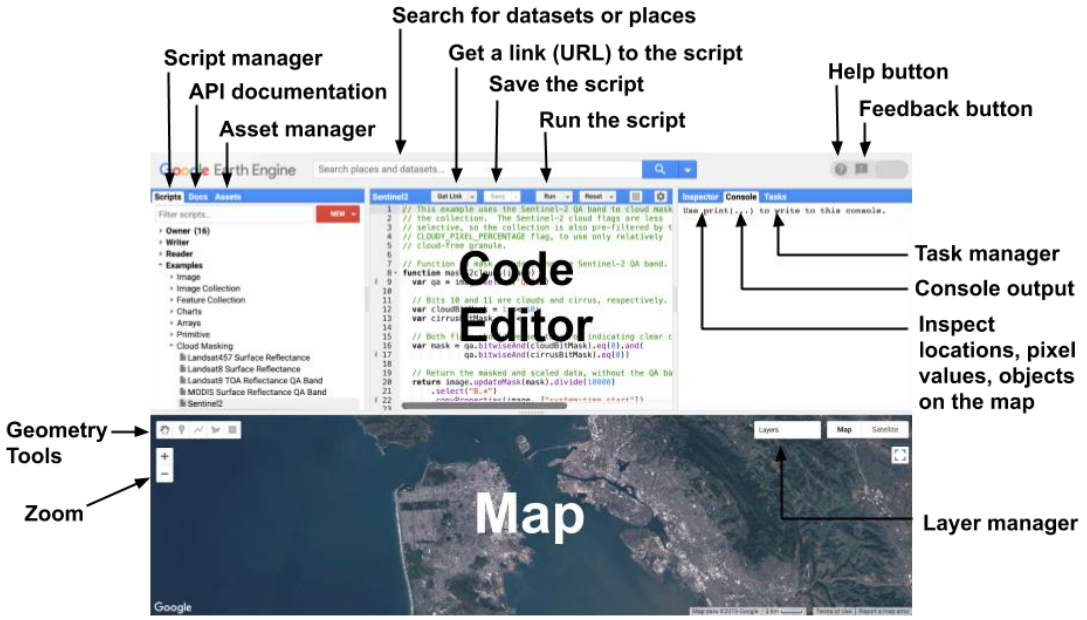
Another way to check on the status of your download is to go to https://code.earthengine.google.com. Click on the 'Tasks' tab in the section on the right. You should see the process running with the spinning gear. When the download completes you will see a blue check mark. Check periodically on your download to make sure all the dates specified are being downloaded.



Optional Exercise 1

While waiting for the soil moisture maps to download lets explore the availability of Sentinel 1 imagery over Indonesia. Click on this link to see the number of acquisitions of Sentinel 1 globally: https://code.earthengine.google.com/6c919eaa51cb77507e373af8eca3fbc7

Here are the components of Google Earth Engine:



For more information about using Google Earth Engine, check out: https://developers.google.com/earth-engine/playground

Guiding questions

Choose an area that you work in. How many Sentinel-1 images are available over your area of interest?

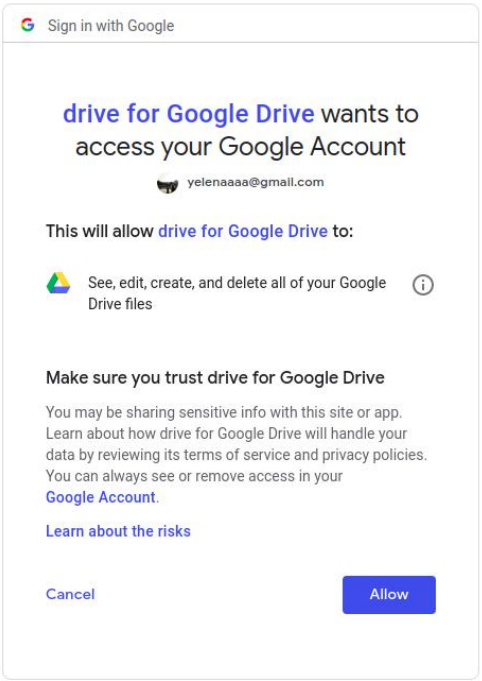
How many Sentinel-1 images are available over a location in Germany?

Downloading the soil moisture maps from Google Drive to your SEPAL account

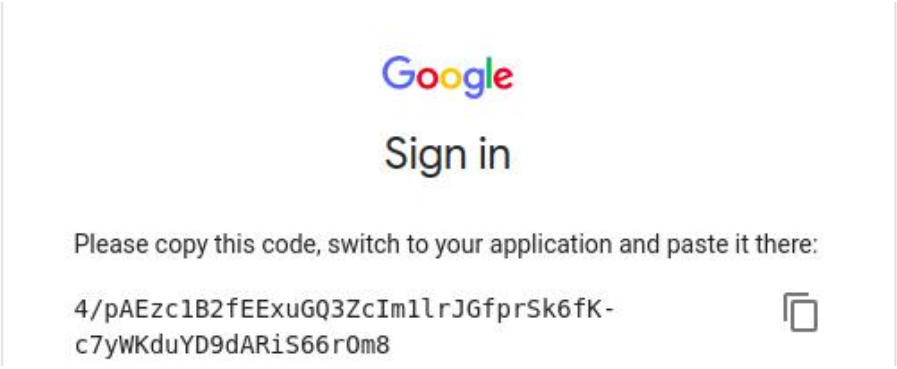
After the download of the time series of soil moisture content maps in complete, check your Google Drive, you will see newly downloaded SMCmaps in your home folder. To copy these images between your cloud storage in Google and your cloud storage in SEPAL is very easy.

In Jupyter notebooks open up the download_smm_from_gdrive.ipynb Jupyter notebook in the scripts folder.
Click on Kernel>Restart and Run All as in the previous notebooks.

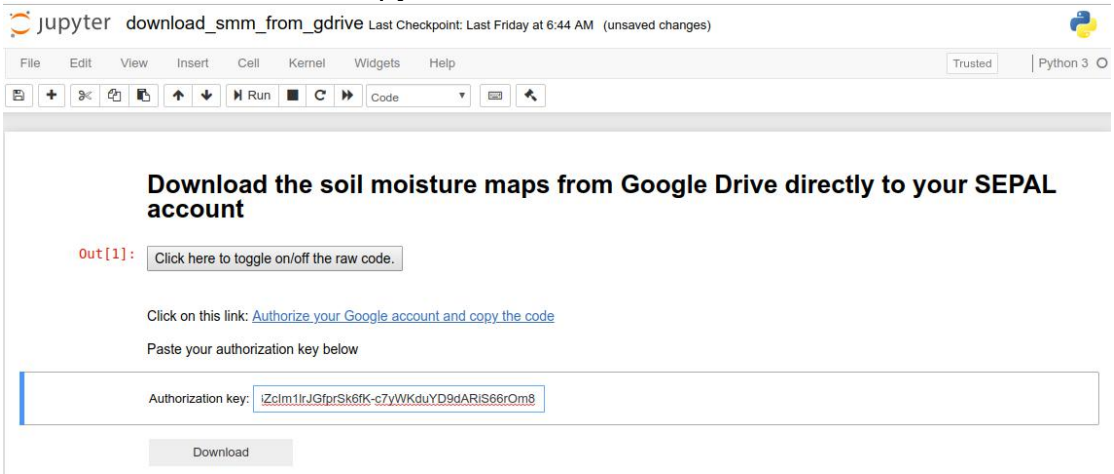
First click on the link to authenticate your Google account.This must be the same Google account you connected to SEPAL and where you soil moistre maps are downloaded.
Click allow.



Copy the authorization code by clicking on the button to the left of the code or by using copy.



Paste this code into the Jupyter notebook.



Then click Download to start the data transfer from Google drive to your SEPAL account.
After the data download completes you can use tools available in SEPAL to process and analyze these soil moisture maps.

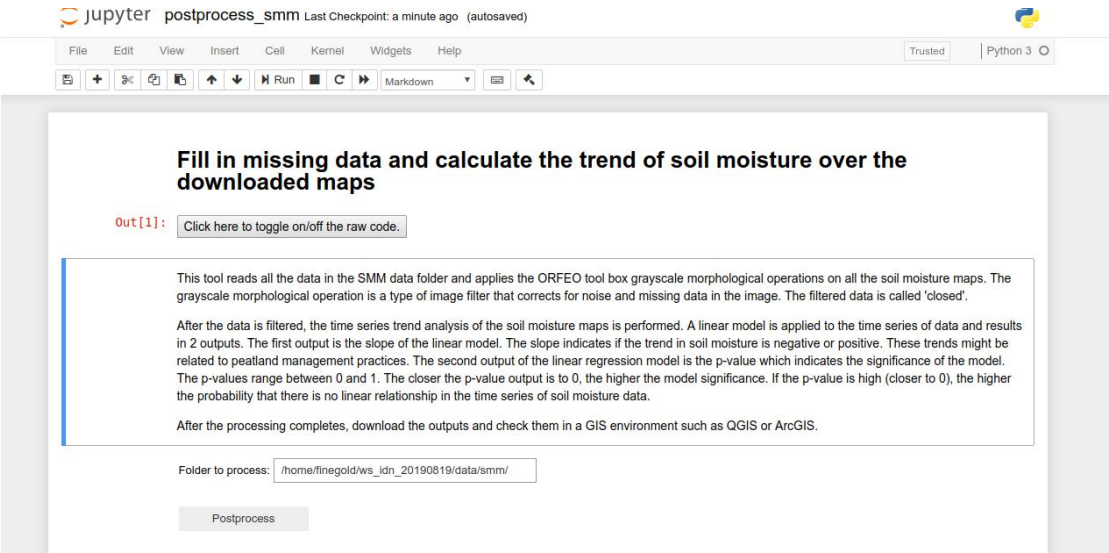
Post-processing and analyzing the soil moisture maps

After your download of the soil moisture content is complete we can apply a robust methodology for image filtering to fill no data gaps. Due to speckle in the Sentinel-1 imagery the soil moisture maps contain some noise and no-data values which are corrected for in some extent using grayscale morphological operation from ORFEO toolbox, a free and open source image processing tool. To read more about the parameterization of the Orfeo toolbox tool, read: https://www.orfeo-toolbox.org/CookBook/Applications/app_GrayScaleMorphologicalOperation.html.

After filtering the image and filling in no data gaps in all the downloaded soil moisture maps in the specified folder, this Jupyter notebook applies a linear regression over the time series of soil moisture map to help understand if there is an increasing or decreasing trend of soil moisture content and the significance of this trend. A linear model is applied to the time series of data and results in 2 outputs. The first output is the slope of the linear model. The slope indicates if the trend in soil moisture is negative or positive. These trends

might be related to peatland management practices. The second output of the linear regression model is the p-value which indicates the significance of the model. The p-values range between 0 and 1. The closer the p-value output is to 0, the higher the model significance. If the p-value is high (closer to 0), the higher the probability that there is no linear relationship in the time series of soil moisture data.

To run the post-processing module specify the folder you where your soil moisture maps are located and then click the ‘Postprocess button’.



After the post-processing completes download the data from SEPAL to your personal computer for analysis and interpretation of the results.

Download/Upload data from/to SEPAL to/from your computer

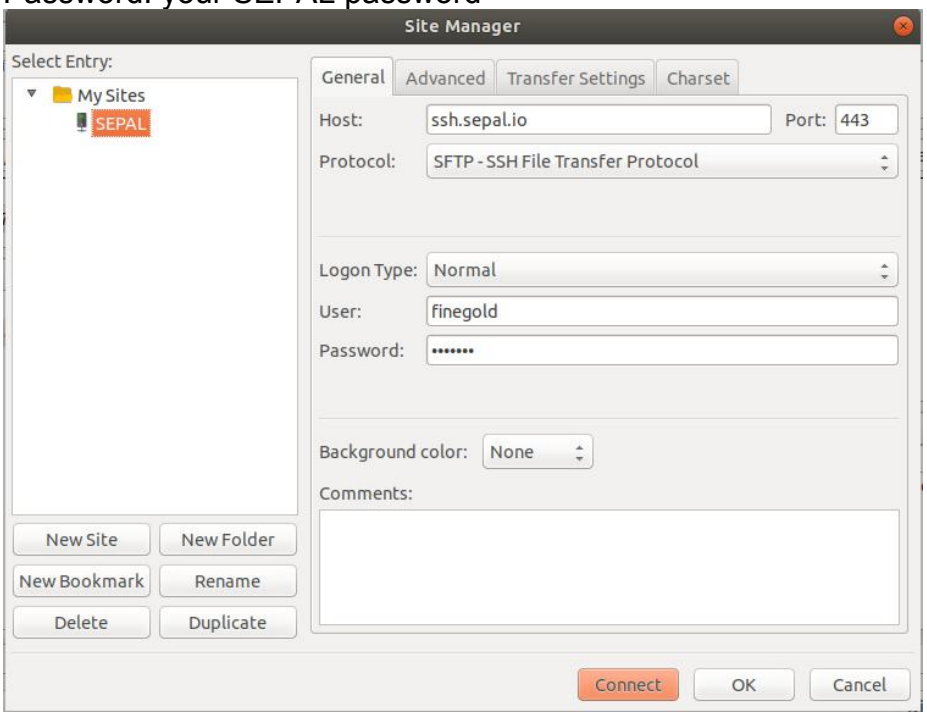
Connect FileZilla to your SEPAL account

Download FileZilla from this link

Accessing files in SEPAL is easy using FileZilla. To use FileZilla, open the application and connect to the SEPAL server by selecting Menu File --> Site Manager in the menu tab.

In the site manager pop-up click on the ‘New Site’ button. Use the screenshot below as a guide for filling in the form:
Host: ssh.sepal.io
Port: 443
Protocol: SFTP – SSH File Transfer Protocol
Logon Type: Normal

User: your SEPAL username
Password: your SEPAL password

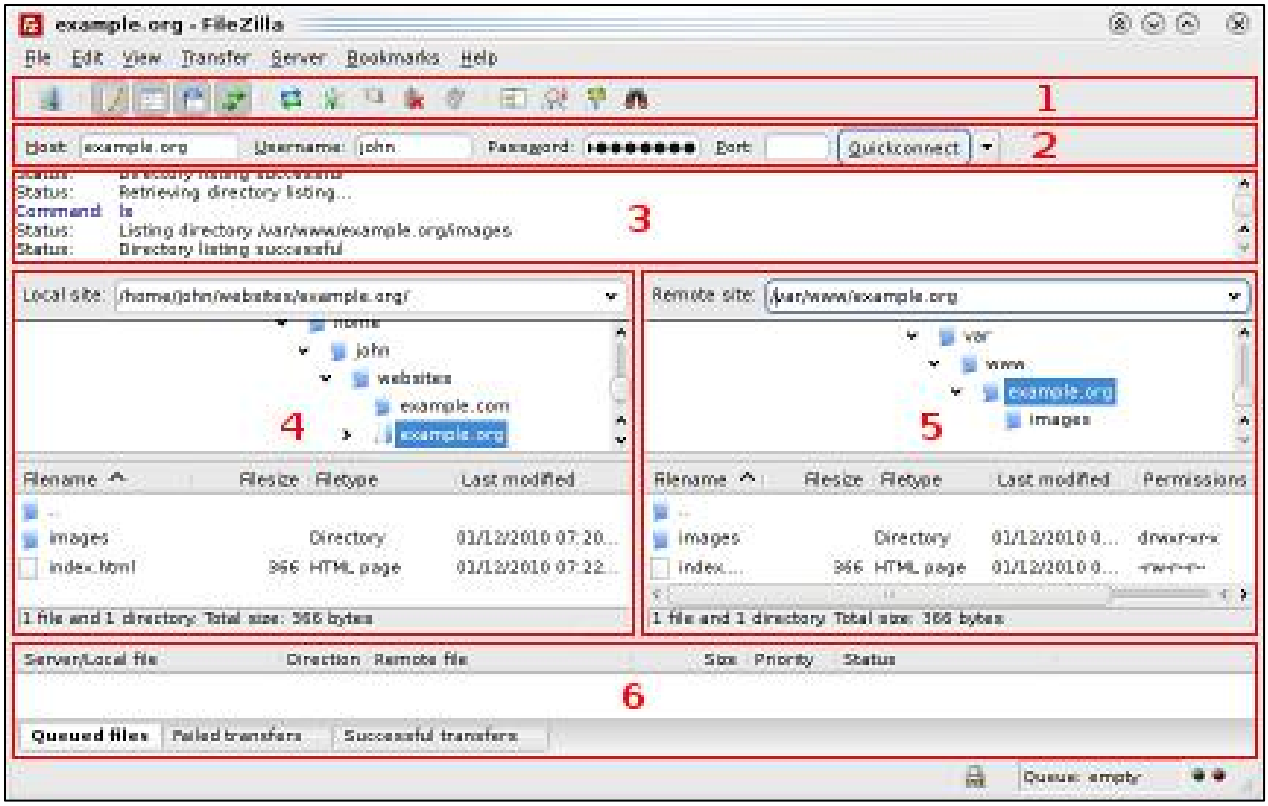


Click Connect and your SEPAL drive will appear in the right panel, parallel to the panel on the left with the files on your computer.

Using FileZilla

Here is a quick introduction from the FileZilla online tutorial corresponding to the image below:

i. the toolbar (1)
ii. quick connect bar (2),
iii. the message log (3) displays transfer and connection related messages. Below, you can find the file listings.
iv. The left column (local pane, 4) displays the local files and directories, i.e. the stuff on the PC you're using FileZilla on. Make sure you are located in the right folder. Ideally, you should have a folder on your computer that will store all data related to this process, point to that folder
v. The right column (server pane, 5) displays the files and directories on the server you are connected to. Both columns have a directory tree at the top and a detailed listing of the currently selected directory's contents at the bottom. You can easily navigate either of the trees and lists by clicking around like in any other file manager.
vi. At the bottom of the window, the transfer queue (6) lists the to-be-transferred and already transferred files.


example.org - FileZilla
File Edit View Transfer Server Bookmarks Help
1
Host: example.org Username: john Password: Port: Quickconnect
2
Status: Retrieving directory listing...
Command: ls
Status: Listing directory /var/www/example.org/images
Status: Directory listing successful
3
Local site: /home/john/websites/example.org/
Remote site: /var/www/example.org
4
5
Filename Filesize Filetype Last modified
images Directory 01/12/2010 07:20...
index.html 366 HTML page 01/12/2010 07:22...
1 file and 1 directory. Total size: 366 bytes
Filename Filesize Filetype Last modified Permissions
images Directory 01/12/2010 0... drwxr-xr-x
index... 366 HTML page 01/12/2010 0...
1 file and 1 directory. Total size: 366 bytes
Server/Local file Direction Remote file Size Priority Status
6
Queued files Failed transfers Successful transfers
Queue: empty