NDP Data Challenges

Participant Tutorial

1. Log in / Register

NATIONAL DATA PLATFORM
v1.0.0
Release Notes

Catalog About Events and Press

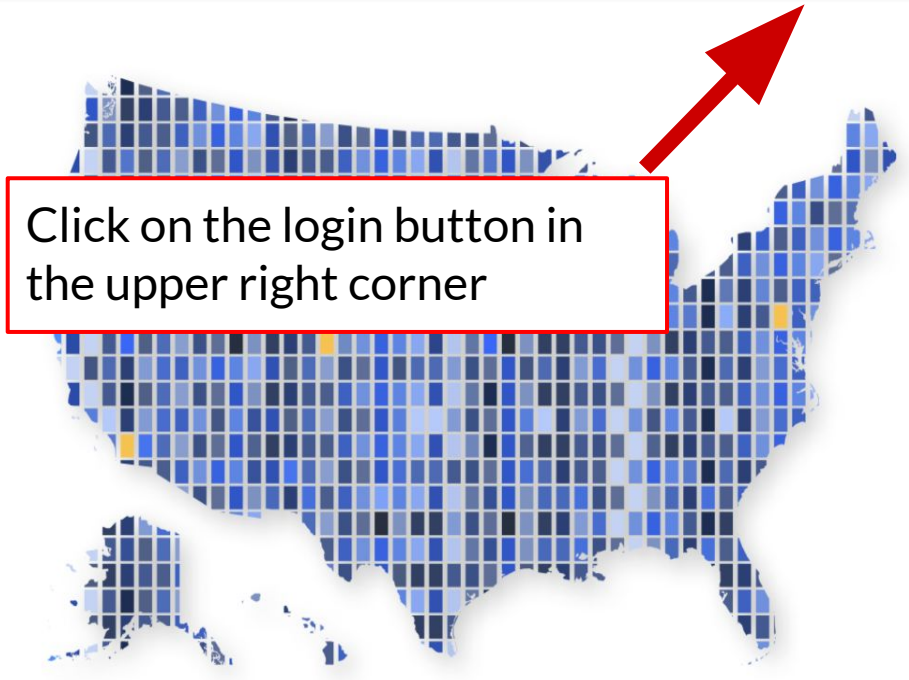
Get Started with NDP | Log in/Register

Open Data, Equitable Access and AI Services for All

Building the nation's federated data ecosystem.
Explore data. Run analyses. Transform AI education.

Explore our catalog of datasets >

SAN DIEGO SUPERCOMPUTER CENTER
SCI www.sci.utah.edu
EarthScope Consortium
UC San Diego
THE UNIVERSITY OF UTAH
University of Colorado Boulder



Click on the login button in the upper right corner

5491
data collections and livestreams

5
data and AI services

348
registered users

You are using v0.1 alpha version of the NDP platform. Please report any issues here

NATIONAL DATA PLATFORM

Sign in to your NDP Account

Email

Password

Remember me

Forgot Password?

Sign In

CILogon

ORCID

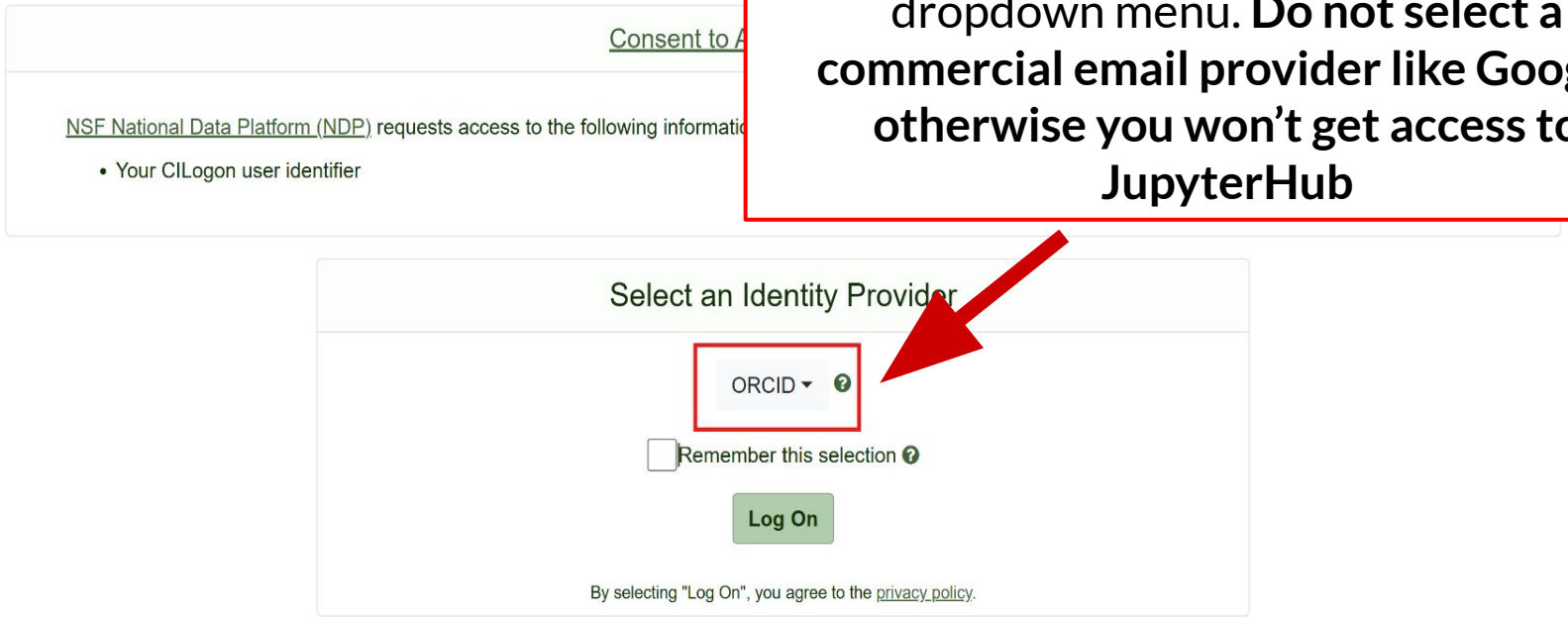
Click on CILogon

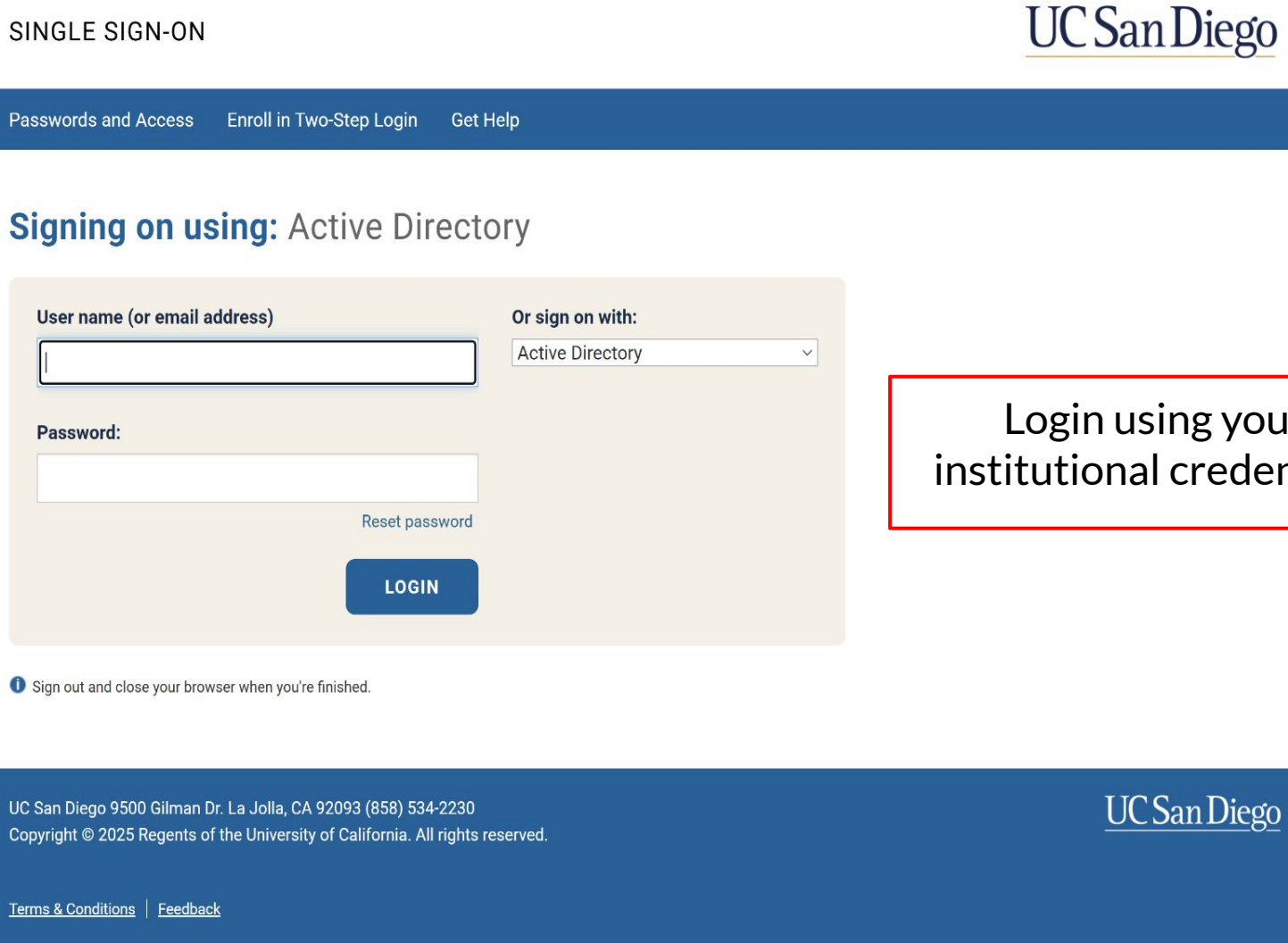
Contact - ndp@sdsc.edu

The National Data Platform was funded by NSF 2335609 under CI, CISE Research Resources programs. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funders.

CILogon

Select your institution/university from the dropdown menu. **Do not select a commercial email provider like Google, otherwise you won't get access to JupyterHub**




SINGLE SIGN-ON
UC San Diego
Passwords and Access
Enroll in Two-Step Login
Get Help
Signing on using: Active Directory
User name (or email address)
Or sign on with:
Active Directory
Password:
Reset password
LOGIN
Sign out and close your browser when you're finished.
Login using your institutional credentials
UC San Diego 9500 Gilman Dr. La Jolla, CA 92093 (858) 534-2230
Copyright © 2025 Regents of the University of California. All rights reserved.
UC San Diego
Terms & Conditions
Feedback

2. Discover the Data Challenge

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard

My Uploads

Workspace

Catalogs

Education Hub

Educator Portal

Learner Portal

Explore

Open Learning

POPs

My Dashboard

2

PV

Welcome back, Pedro Antonio Ramonetti Vega!

LATEST ANNOUNCEMENTS

January 22-23, 2024	2-day all hands team meeting at the San Diego Supercomputer Center!
January 19, 2024	More details ironed out. Alpha version of NDP is live!
January 12, 2024	User dashboard drafted. JupyterHub integrated.
January 5, 2024	String search integrated.

Quick Explore

Upload

JupyterHub

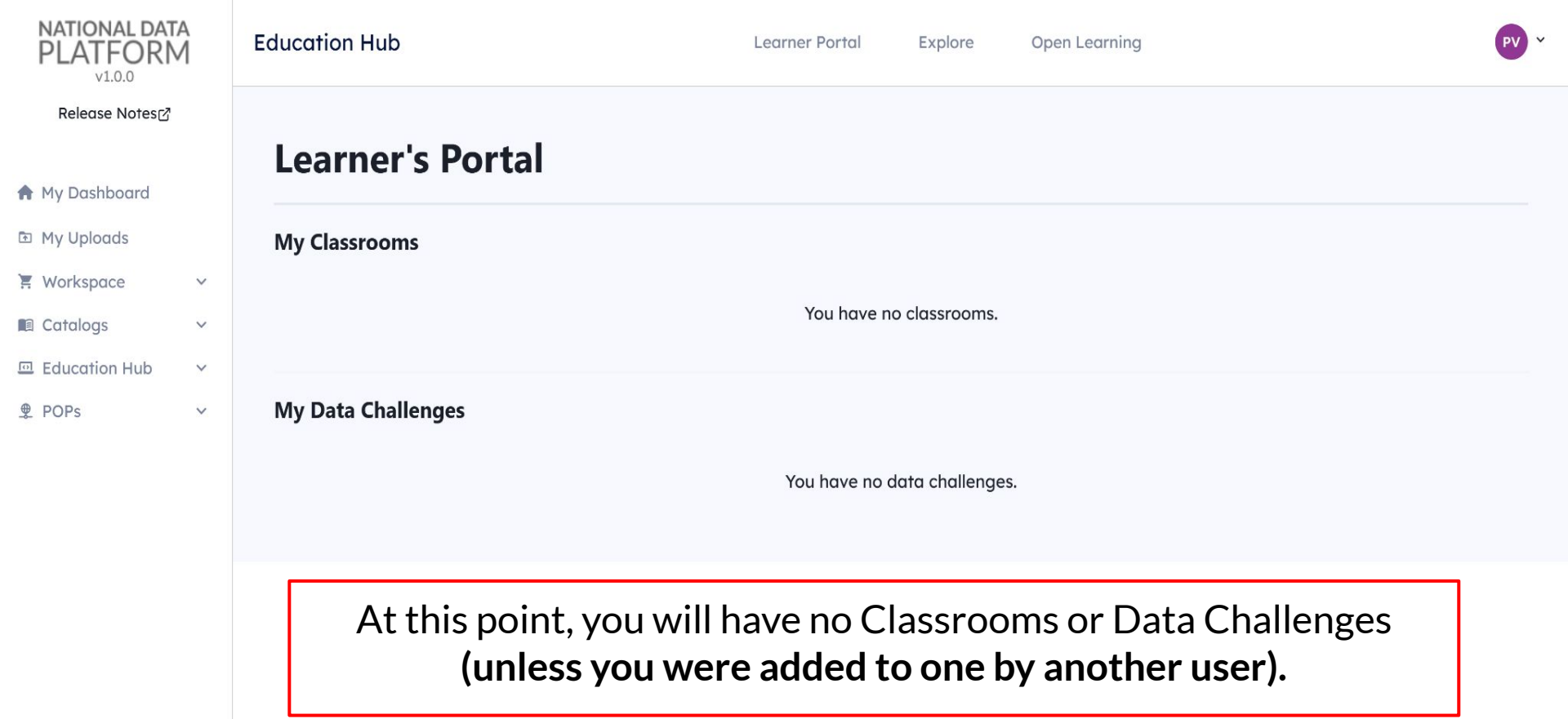
Education Gateway

Click on Education Hub, followed by Learner Portal

Your Profile

PV

Pedro Antonio Ramonetti Vega
pramonettivega@ucsd.edu



At this point, you will have no Classrooms or Data Challenges **(unless you were added to one by another user).**

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Education Hub | Learner Portal | Explore | Open Learning | PV

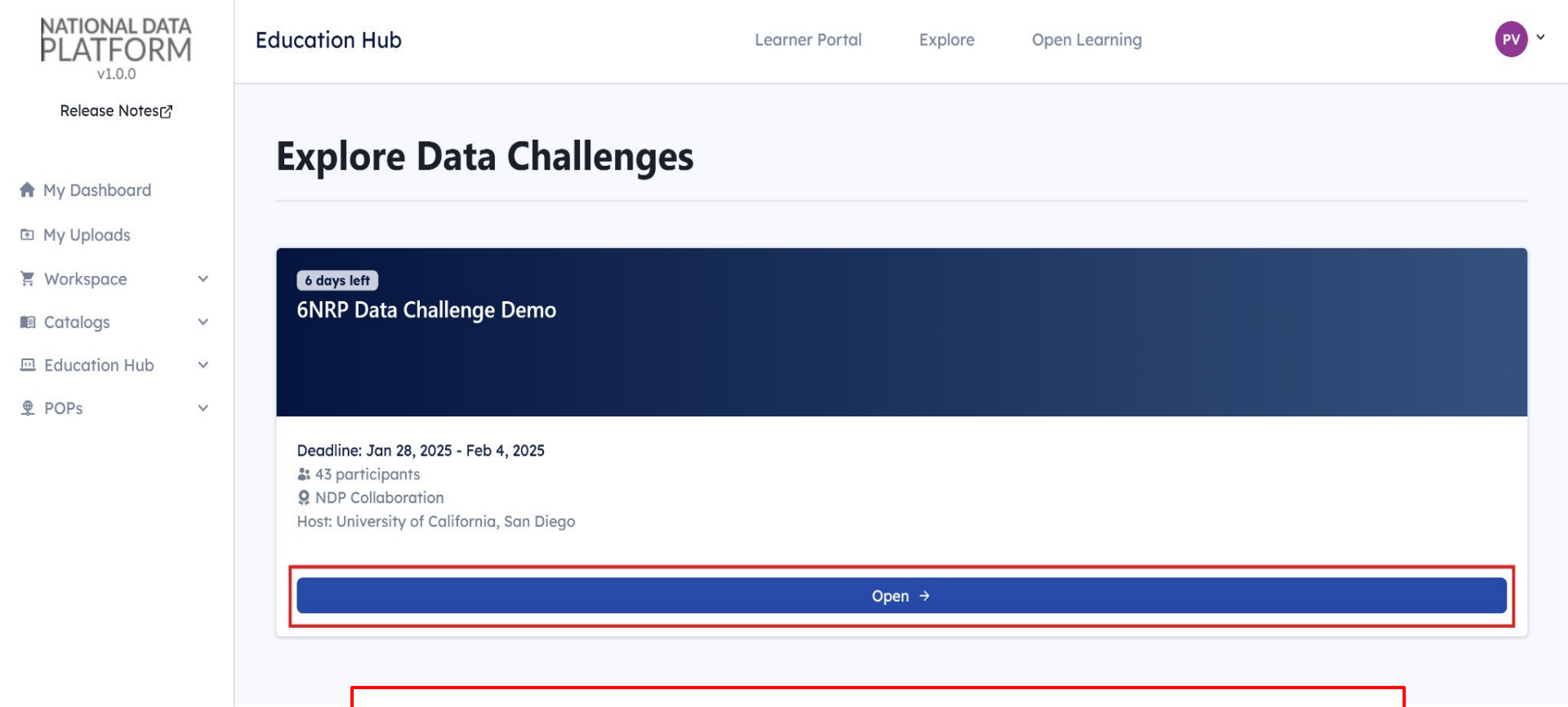
Learner's Portal

My Classrooms

Click on Explore

My Data Challenges

You have no data challenges.



Identify the Data Challenge of your interest and click on Open

← Go Back

6NRP Data Challenge Demo

University of California, San Diego

This Data Challenge Demo aims to showcase the user experience and capabilities of the National Data Platform in hosting challenges that utilize big data and high-performance computation.

Prize: NDP Collaboration

Join Challenge

Overview | Rules and Eligibility | FAQ | Modules | Dataset Details

Background

The National Data Platform (NDP) had been developed as a federated and extensible data ecosystem to promote collaboration, innovation and equitable use of data using existing and future national cyberinfrastructure (CI) capabilities.
The Education Hub aims to democratize access to advanced computational tools, resources, and AI-ready data, empowering the development of an AI-ready workforce. It enables instructors to deliver resource-intensive courses, facilitate data challenges, and share their expertise through hands-on, open le

Explore all the information provided in the Data Challenge

This de ve fire models.

The primary objective of this challenge is to leverage the extensive Forest Inventory and Analysis (FIA) database to predict tree size and species composition, providing valuable insights for advancing forest management and fire modeling efforts.

3. Join the Data Challenge / Create a Team

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

← Go Back

6NRP Data Challenge Demo

University of California, San Diego

This Data Challenge Demo aims to showcase the user experience and capabilities of the National Data Platform in hosting challenges that utilize big data and high-performance computation.

Prize: NDP Collaboration

Join Challenge

Overview | Rules and Eligibility | FAQ | Modules | ataset Details

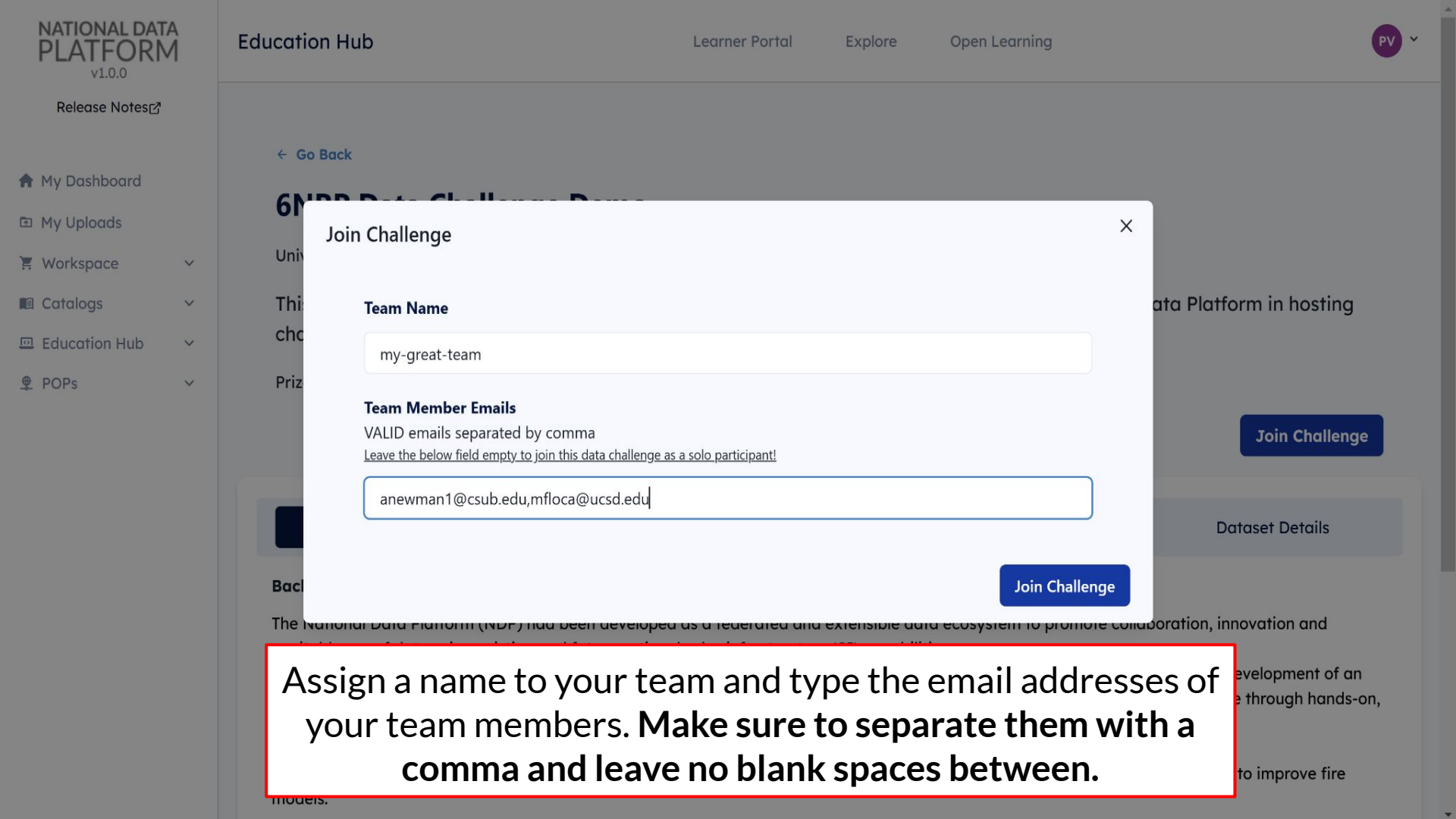
Background

The National Data Platform (NDP) had been developed as a federated and extensible data ecosystem to promote collaboration, innovation and equitable use of data using existing and future national cyberinfrastructure (CI) capabilities.
The Education Hub aims to democratize access to advanced computation development of an AI-ready workforce. It enables instructors to deliver resource-intensive co se through hands-on, open learning resources

Click on Join Challenge

This demo data challenge is designed to address a critical need of our research teams: the development of more accurate tree lists to improve fire models.

The primary objective of this challenge is to leverage the extensive Forest Inventory and Analysis (FIA) database to predict tree size and species composition, providing valuable insights for advancing forest management and fire modeling efforts.


NATIONAL DATA PLATFORM
v1.0.0
Release Notes
Education Hub
Learner Portal
Explore
Open Learning
PV
My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs
Go Back
Join Challenge
Dataset Details
Join Challenge
Team Name
my-great-team
Team Member Emails
VALID emails separated by comma
Leave the below field empty to join this data challenge as a solo participant!
anewman1@csub.edu,mfloca@ucsd.edu
Join Challenge
Assign a name to your team and type the email addresses of your team members. **Make sure to separate them with a comma and leave no blank spaces between.**

4. Review your modules

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Your Team | Overview | Rules and Eligibility | FAQ | Dataset Details

Deadline: February 3, 2025

0 DAYS LEFT

Team Details

my-great-team Edit

Once you join a Data Challenge, your modules will appear at the bottom of *Your Team* page. To view the content of each module, click on the *View* button.

Modules

Module 1 - Exploring FIA Database

UC San Diego

In this module, you will explore the Forest Inventory and Analysis (FIA) Database. The primary goal is to help you become familiar with this dataset, laying the groundwork for developing a model to predict various tree...

View → JupyterHub

Module 3 - Solution Development

UC San Diego

This module is set as a base workspace for the participants to start working on the solution of the Data Challenge. No github repository or preparatory material is provided.

View → JupyterHub

Module 2 - Base Solution

UC San Diego

In this module, you will implement a Random Forest model, which serves as the baseline solution for the Data Challenge. This example is provided to demonstrate the expected output of the challenge and offer participants a...

View → JupyterHub

Read Modules Carefully

Read carefully the instructions of each module prior to launching them on JupyterHub. Each module should come with specific instructions on how to:

- Reserve your JupyterHub server computing resources
- Clone the attached repository (if provided)
- Install dependencies
- Download your data to JupyterHub

5. Launching JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Datasets

FIA Database - California

Once you have read the module's instructions, you can go to JupyterHub by clicking on the JupyterHub button at the bottom of the module

Scripts

https://github.com/pramonettivega/6nrp-demo.git

Go to JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

Your Team | Overview | Rules and Eligibility | FAQ | Dataset Details

Deadline: February 3, 2025

3 DAYS LEFT

Team Details

my-great-team Edit

anewman1@csub.edu
mfloca@ucsd.edu
pramonettivega@ucsd.edu

This button is also located in the cover of each module

Modules

Module 1 - Exploring FIA Database

UC San Diego

In this module, you will explore the Forest Inventory and Analysis (FIA) Database. The primary goal is to help you become familiar wit this dataset, laying the groundwork for developing a model to predict various tree...

View → JupyterHub

Module 3 - Solution Dev lopment

UC San Diego

This module is set as a base wor pace for the participants to start working on e solution of the Demo Data Challenge. No repository or preliminary material is prov .

View → JupyterHub

Module 2 - Base Solution

UC San Dieg

In this modul you will implement a Random Forest model, w erves as the baseline solution for th Challenge. This example is provided to demon te the expected output of the challenge and offer participants a...

View → JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

- My Dashboard
- My Uploads
- Workspace
- Catalogs
- Education Hub
- POPs

My Dashboard

2 PV

Welcome back, Pedro Antonio Ramonetti Vega!

LATEST ANNOUNCEMENTS

January 22-23, 2024	2-day all hands team meeting at the San Diego Supercomputer Center!
January 19, 2024	More details ironed out. Alpha version of NDP is live!
January 12, 2024	User dashboard drafted. JupyterHub integrated.
January 5, 2024	String search integrated.

Or in your Dashboard

Quick Explore

- Explore Data Catalog
- Upload Data to Catalog
- JupyterHub
- Education Gateway

Your Profile

PV

Pedro Antonio Ramonetti Vega
pramonettivega@ucsd.edu

Setup resources as indicated in your module

NDP JupyterHub Server Options

Available resources page

Region

Any

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

/dev/shm for pytorch

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

Enter your custom image URL here, including the tag. For example: jupyter/r-notebook:latest

Timeout (in seconds): once a server has been successfully spawned, time to wait until it actually starts

1200

Reserving unauthorized resources (such as reserving GPU's without prior authorization) might ban you from the platform

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

/dev/shm for pytorch

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

Enter your custom image URL here, including the tag.

Timeout (in seconds): once a server has been successfully spawned, time to wait until it actually starts

1200

Architecture

amd64

Note: *Please stop your server after it is no longer needed, or in case you want to laun ent content image*
In order to stop the server from running Jupyter Lab, go to File > Hub Control Panel > server

Note: */home/jovyan/work/_User-Persistent-Storage_CephBlock_ is the persistent volu e directory, make sure to save your work in it, otherwise it will be deleted*

Start

After setting up your resources, click on start to launch your server

NDP JupyterHub Server Options

Available resources page

Error: HTTP 401: Unauthorized (Your session has expired. Please log out and log in again.)

Region

Any

Zone

Any

GPUs

You may encounter this message. This means your access credentials within JupyterHub have expired.

16

GPU type

Any

☐ **/dev/shm for pytorch**

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

NDP JupyterHub Server Options

Available resources page

Error: HTTP 401: Unauthorized (Your session has expired. Please log out and log in again.)

To solve this problem, simply logout using this button, and login again through CILogon. Once you do that, you will be able to start your server

Region

Any

Zone

Any

GPUs

0

Cores

1

RAM, GB

16

GPU type

Any

/dev/shm for pytorch

Select Pre-Built Image (Pre-Built Image Guide):

Minimal NDP Starter Jupyter Lab

Or Bring Your Own Image (JupyterLab Compatible):

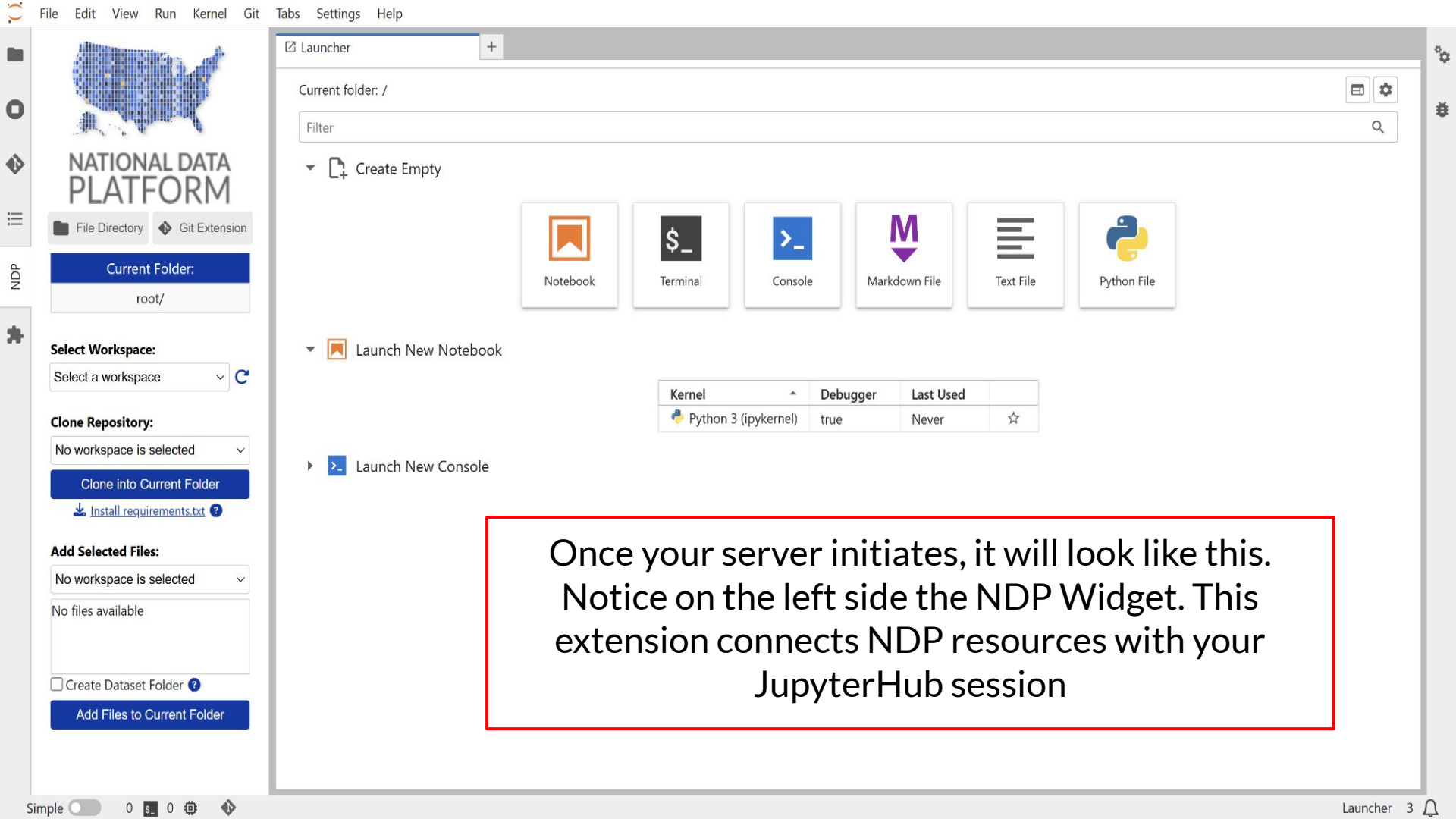
Your server is starting up.

You will be redirected automatically when it's ready for you.

2025-01-29T21:35:52Z [Normal] AttachVolume.Attach succeeded for volume "pvc-f54bb41b-2b4b-4efd-8c20-e43441ec0215"

Event log

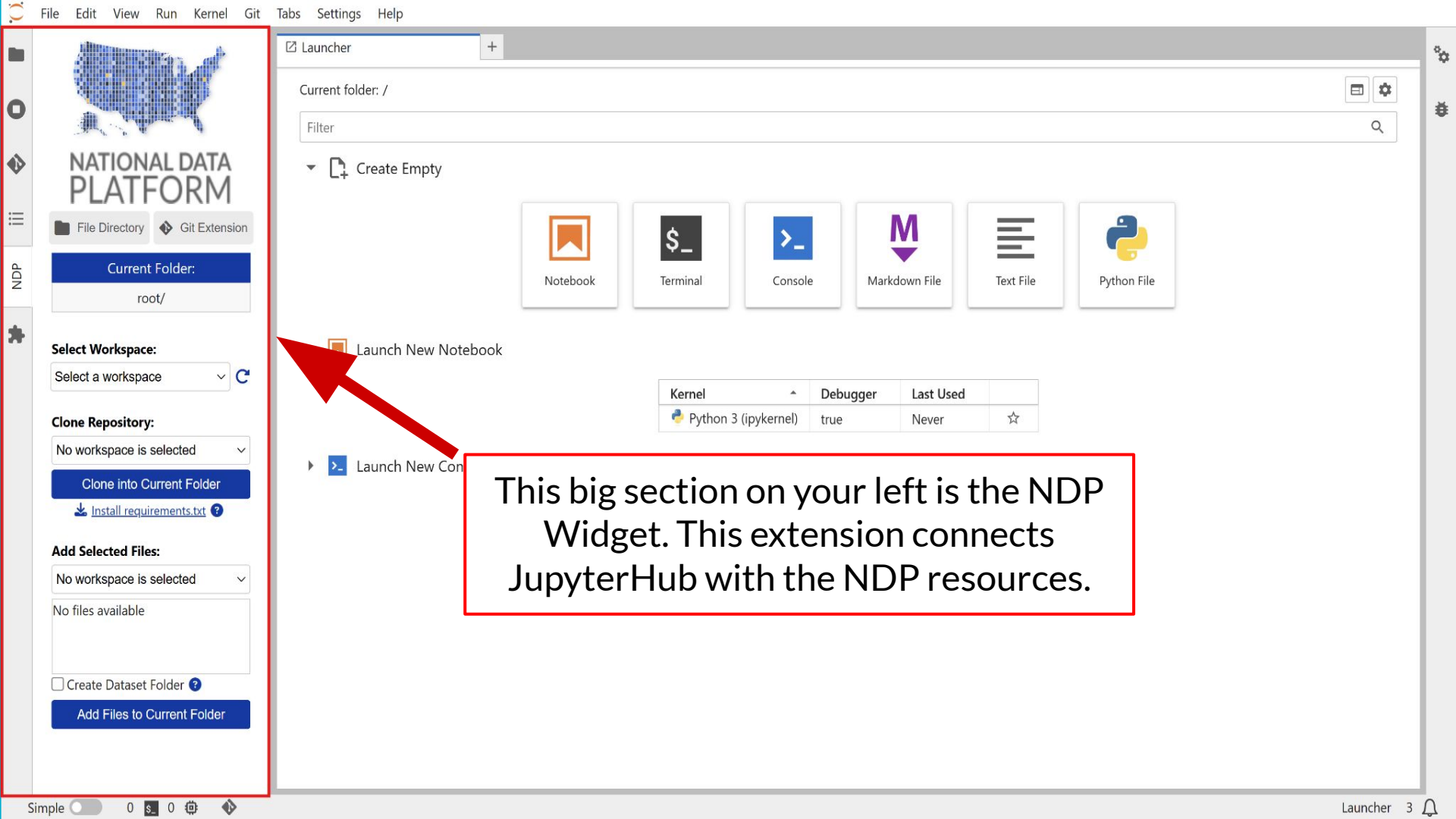
Wait for your server to initiate

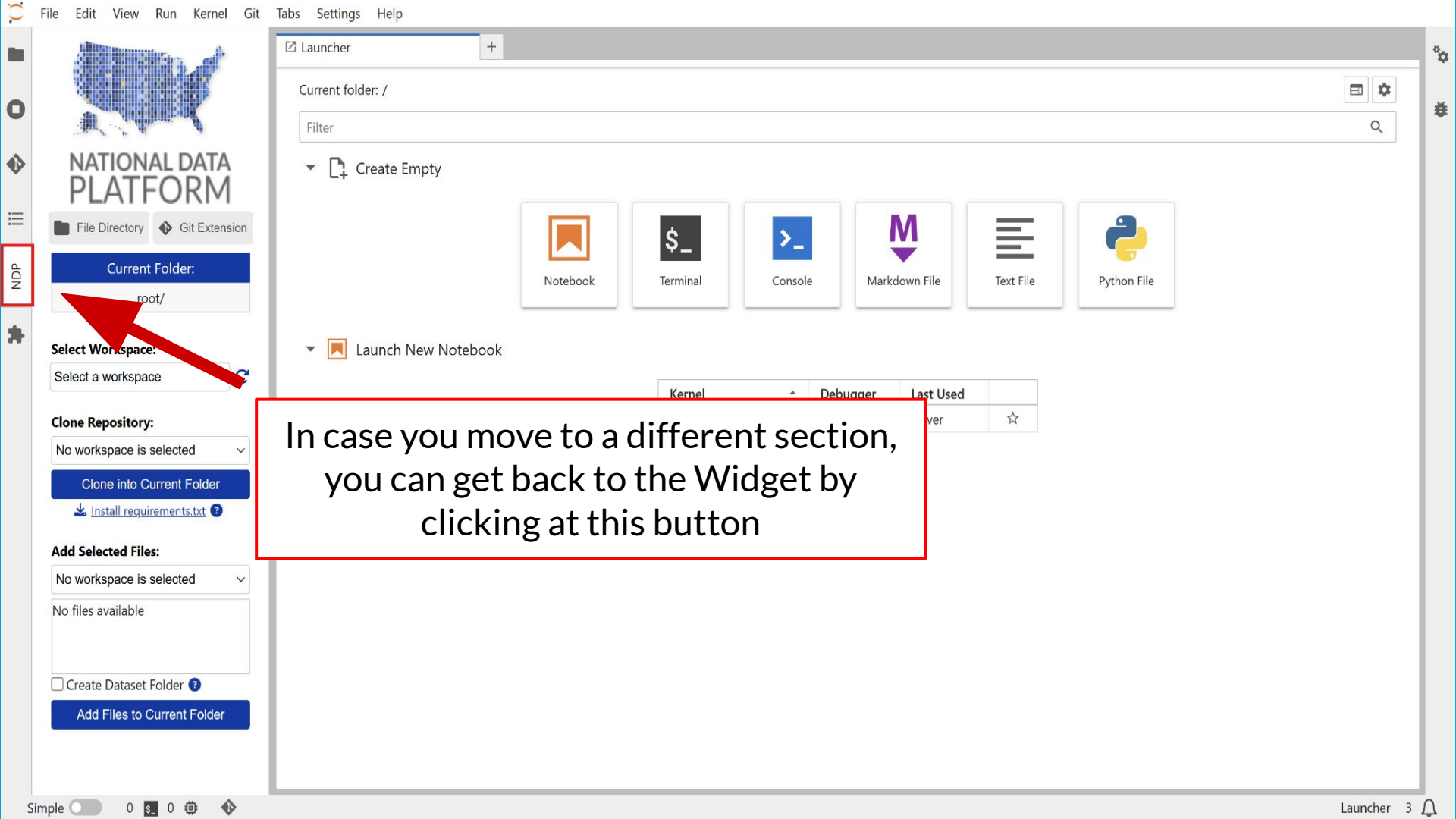

File Edit View Run Kernel Git Tabs Settings Help
Launcher
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Once your server initiates, it will look like this. Notice on the left side the NDP Widget. This extension connects NDP resources with your JupyterHub session
Simple
Launcher 3

6. JupyterHub sections

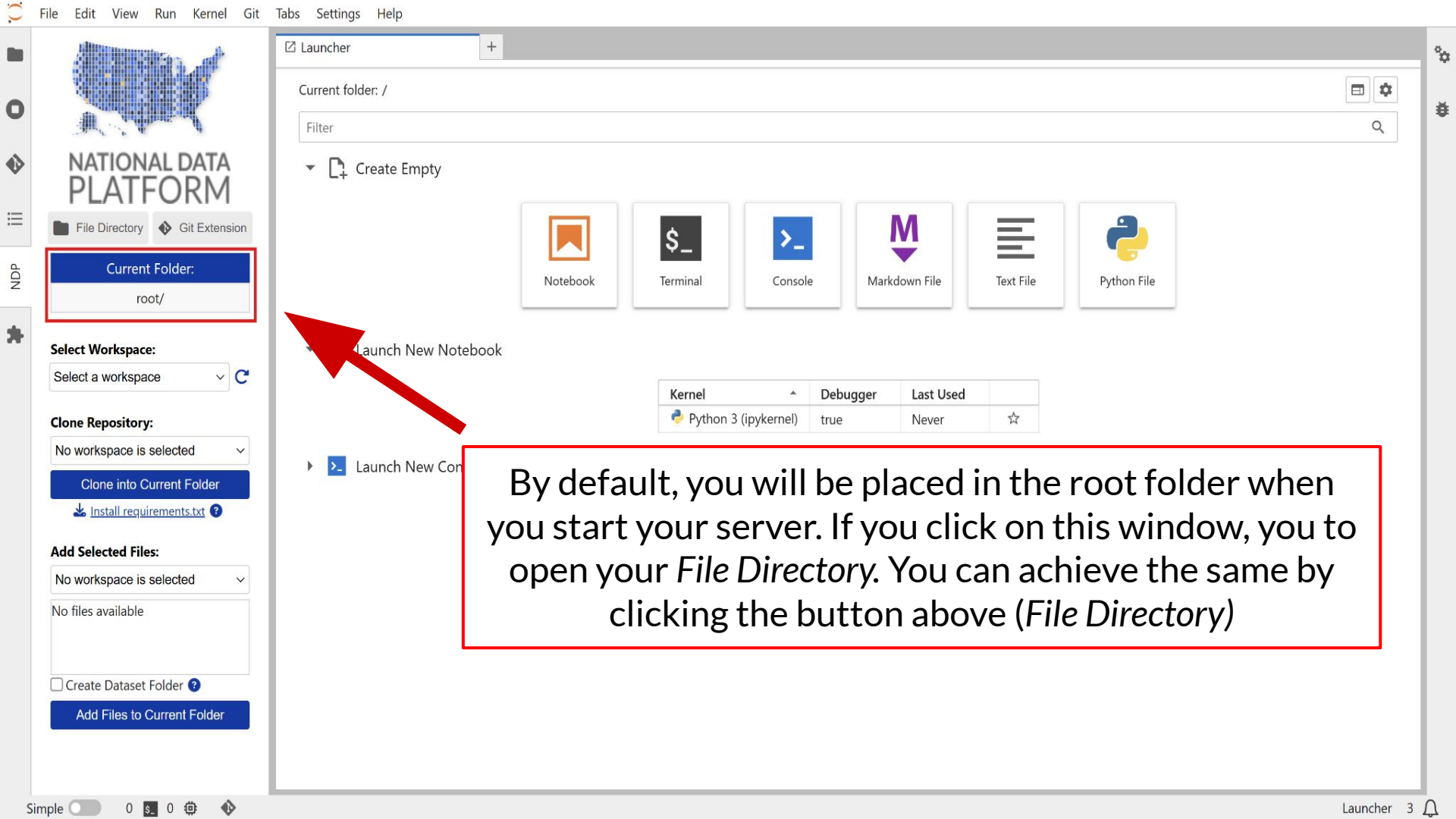
In the next slides, we will take a look at the different sections and pieces of your JupyterHub server

6.a. NDP Widget

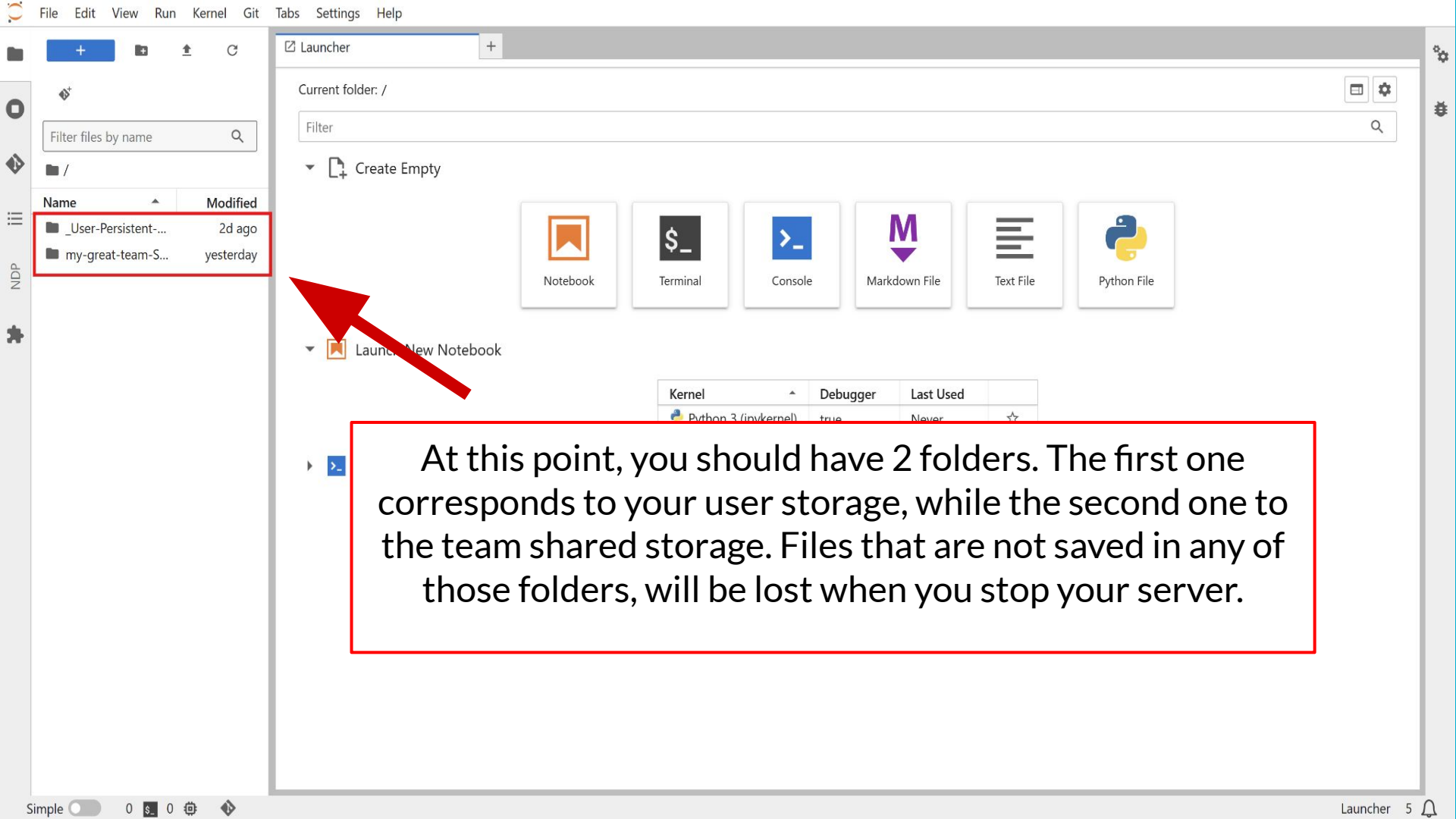

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
NDP
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Con
This big section on your left is the NDP Widget. This extension connects JupyterHub with the NDP resources.
Simple
Launcher 3


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
NDP
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
In case you move to a different section, you can get back to the Widget by clicking at this button
Simple
Launcher 3

6.b. File Manager


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Con
By default, you will be placed in the root folder when you start your server. If you click on this window, you to open your *File Directory*. You can achieve the same by clicking the button above (*File Directory)*
Simple
Launcher 3

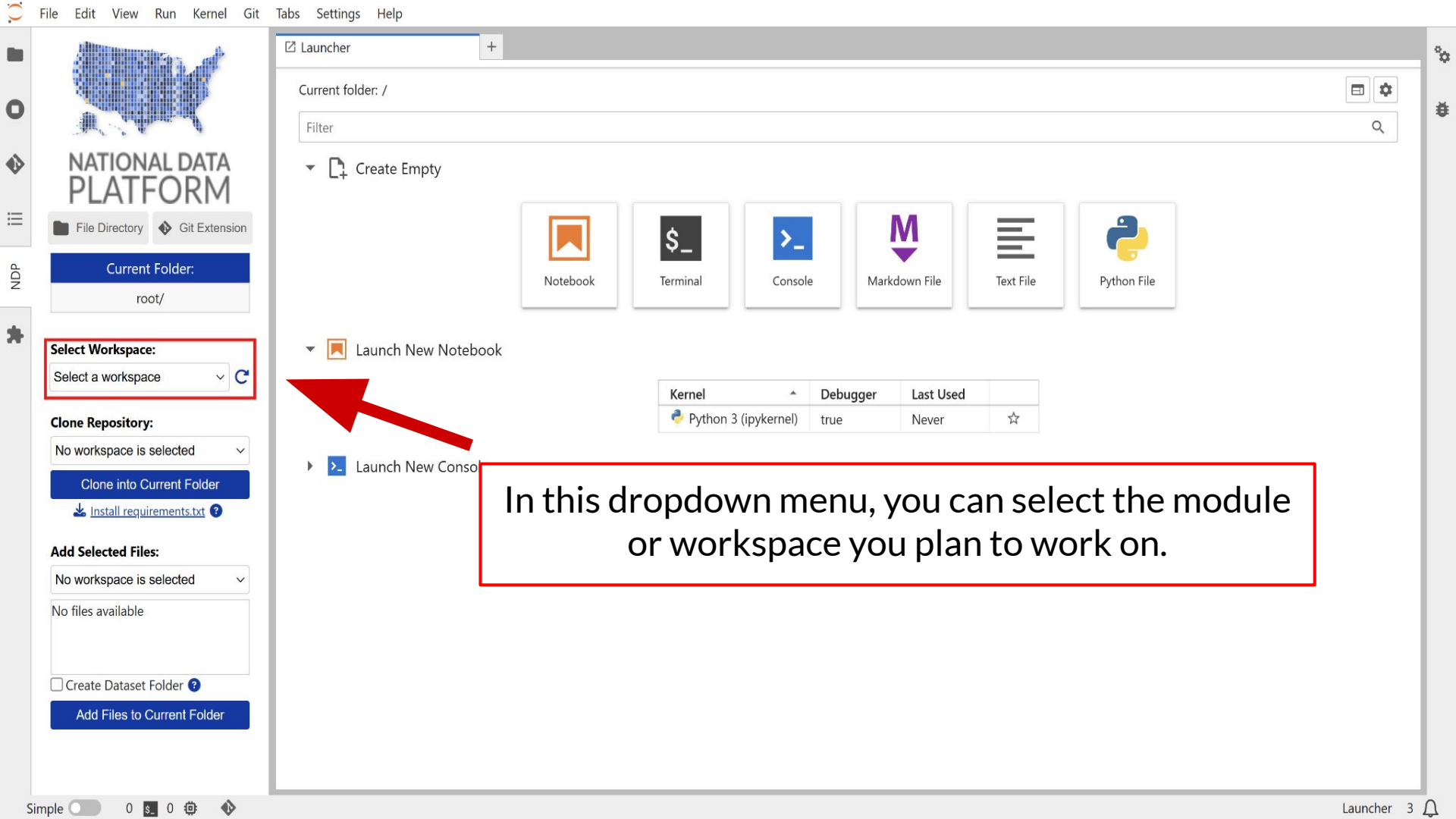
6.c. User Storage and Shared storage


File Edit View Run Kernel Git Tabs Settings Help
Launcher
Current folder: /
Filter files by name
/
Name
Modified
_User-Persistent-...
2d ago
my-great-team-S...
yesterday
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
At this point, you should have 2 folders. The first one corresponds to your user storage, while the second one to the team shared storage. Files that are not saved in any of those folders, will be lost when you stop your server.
Simple
Launcher 5

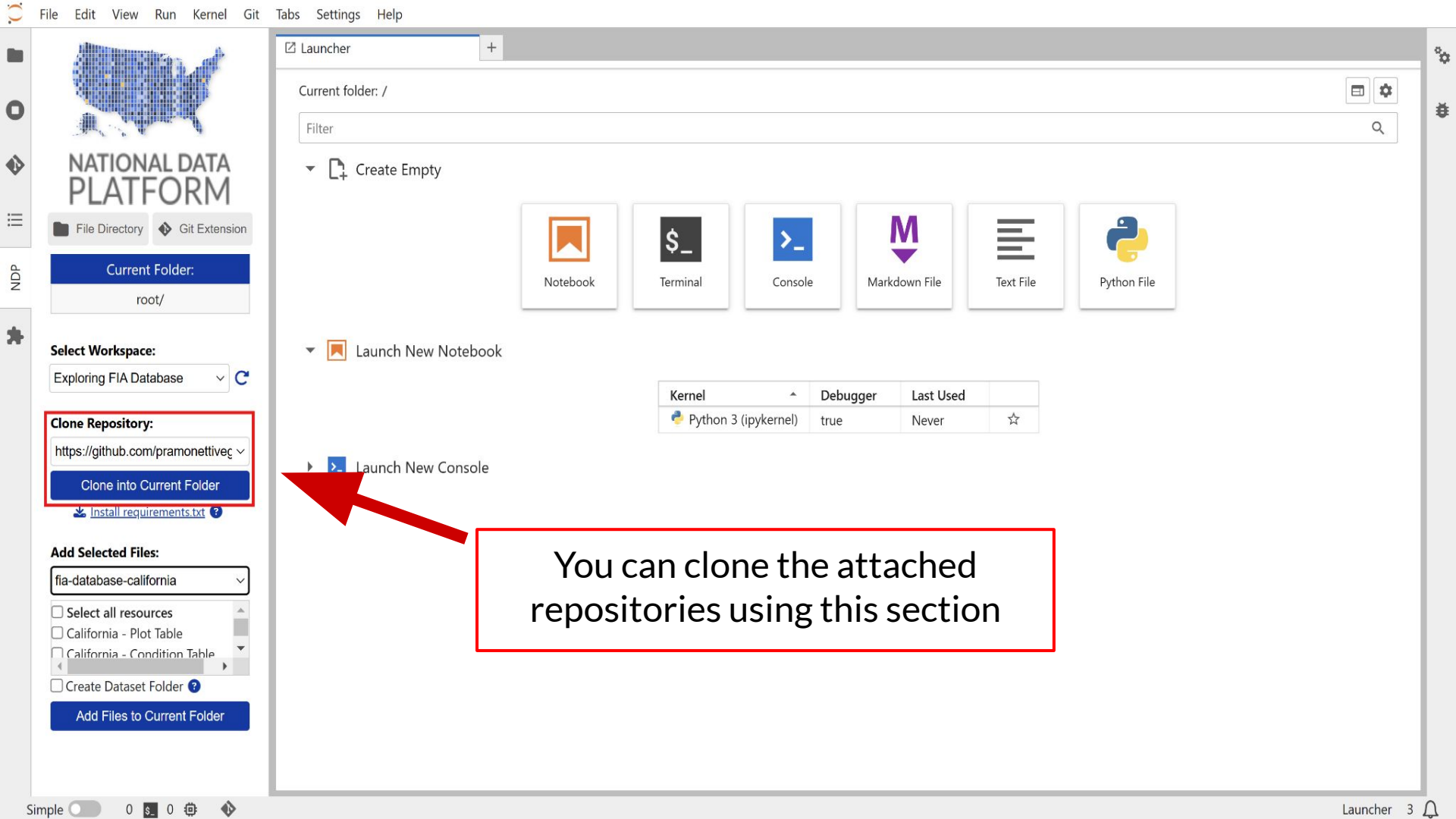
Good practices about shared storage

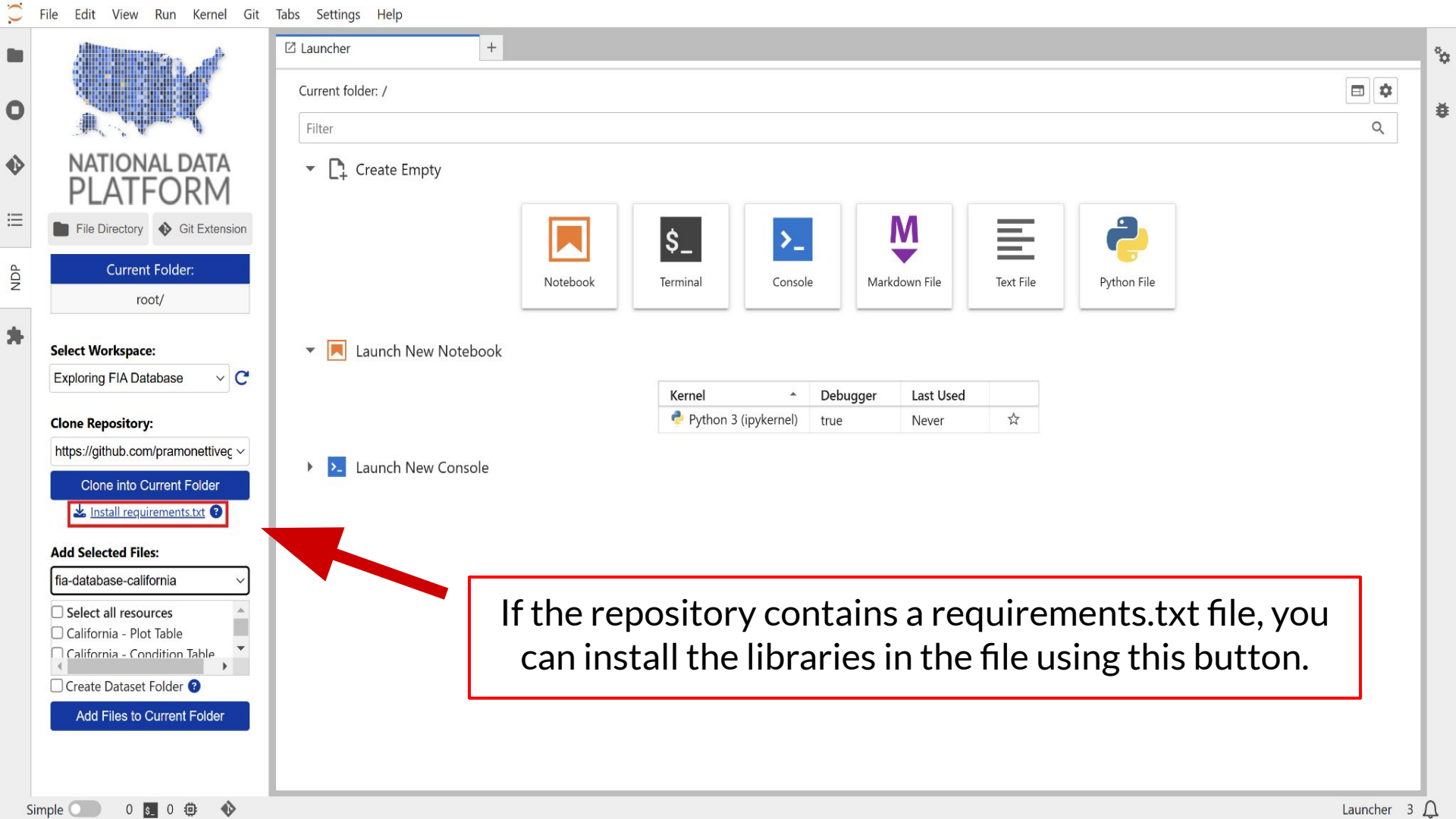
- Always communicate with your team members about the files they place/remove from this folder.
- Do not work on the same file at the same time or you will run into an overwrite conflict.
- Shared storage is limited, so use it to share frequently accessed files or data derived from your workflow.

6.e. Modules/Workspaces List

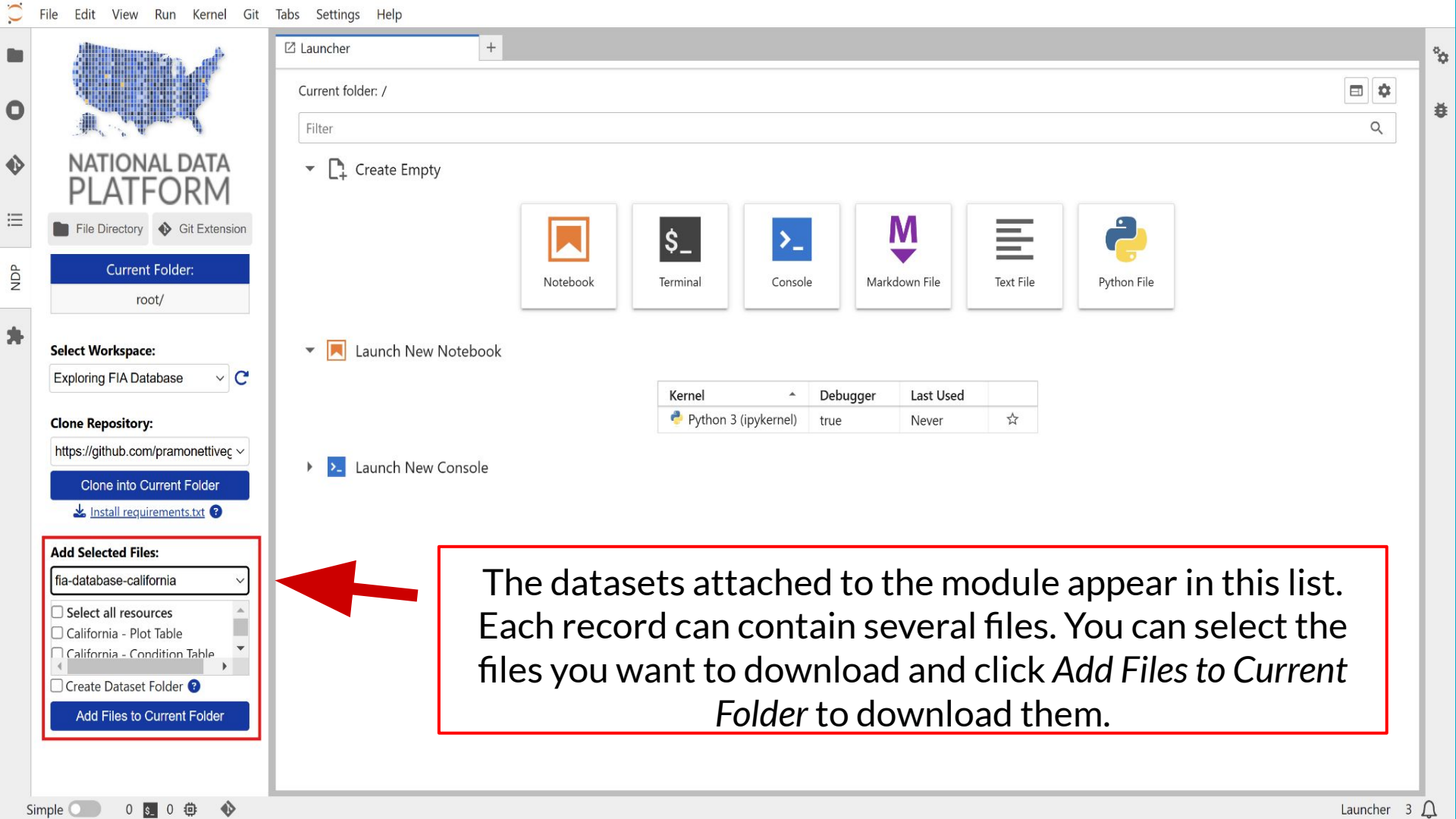

File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Select a workspace
Clone Repository:
No workspace is selected
Clone into Current Folder
Install requirements.txt
Add Selected Files:
No workspace is selected
No files available
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Conso
In this dropdown menu, you can select the module or workspace you plan to work on.
Simple
Launcher 3

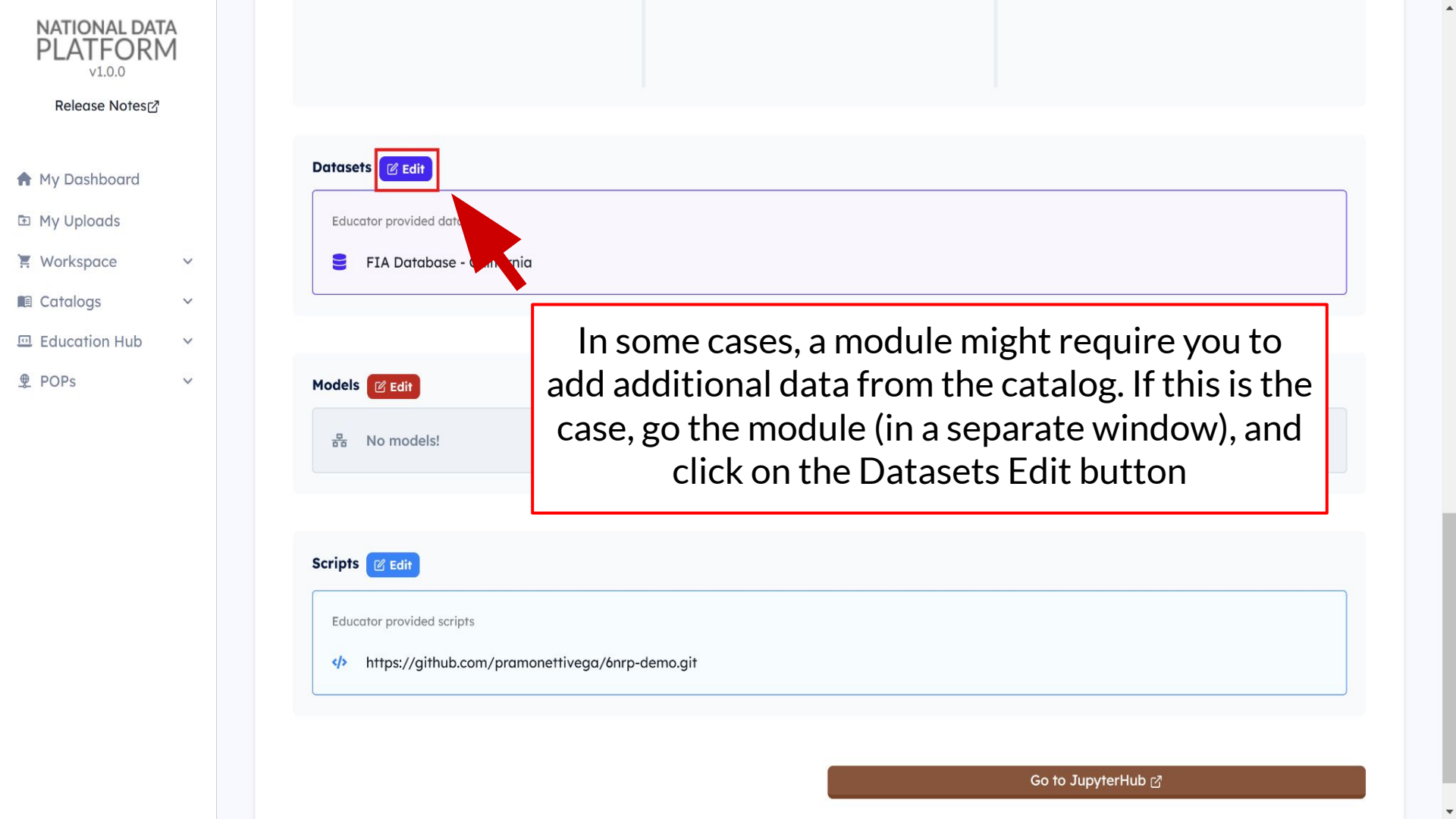
6.f. Repositories List


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
You can clone the attached repositories using this section
Simple
Launcher 3


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
If the repository contains a requirements.txt file, you can install the libraries in the file using this button.
Simple
Launcher 3

6.g. Datasets List


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
The datasets attached to the module appear in this list. Each record can contain several files. You can select the files you want to download and click *Add Files to Current Folder* to download them.
Simple
Launcher 3


NATIONAL DATA PLATFORM
v1.0.0
Release Notes
My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs
Datasets
Edit
Educator provided dat
FIA Database -
nia
In some cases, a module might require you to add additional data from the catalog. If this is the case, go the module (in a separate window), and click on the Datasets Edit button
Models
Edit
No models!
Scripts
Edit
Educator provided scripts
https://github.com/pramonettivega/6nrp-demo.git
Go to JupyterHub

NATIONAL DATA PLATFORM
v1.0.0

Release Notes

My Dashboard
My Uploads
Workspace
Catalogs
Education Hub
POPs

Edit Datasets

Module datasets

FIA Database - California

You have not added any datasets.

Search data catalog

sierra

Search | Clear

1 - 3 of 97 Data Collections and Streams

1 2 ... 33

Sierra Nevada Sawtimber Volume

This metric expresses the amount of total existing, aboveground, live tree stem biomass measured in dry weight tons per acre for the Sierra Nevada region. This metric can be used to assess the sawtimber volume present at t...

View More | + Add

Sierra Nevada Biomass Volume

...ry weight tons per acre) from ...for the Sierra Nevada region....

View More | + Add

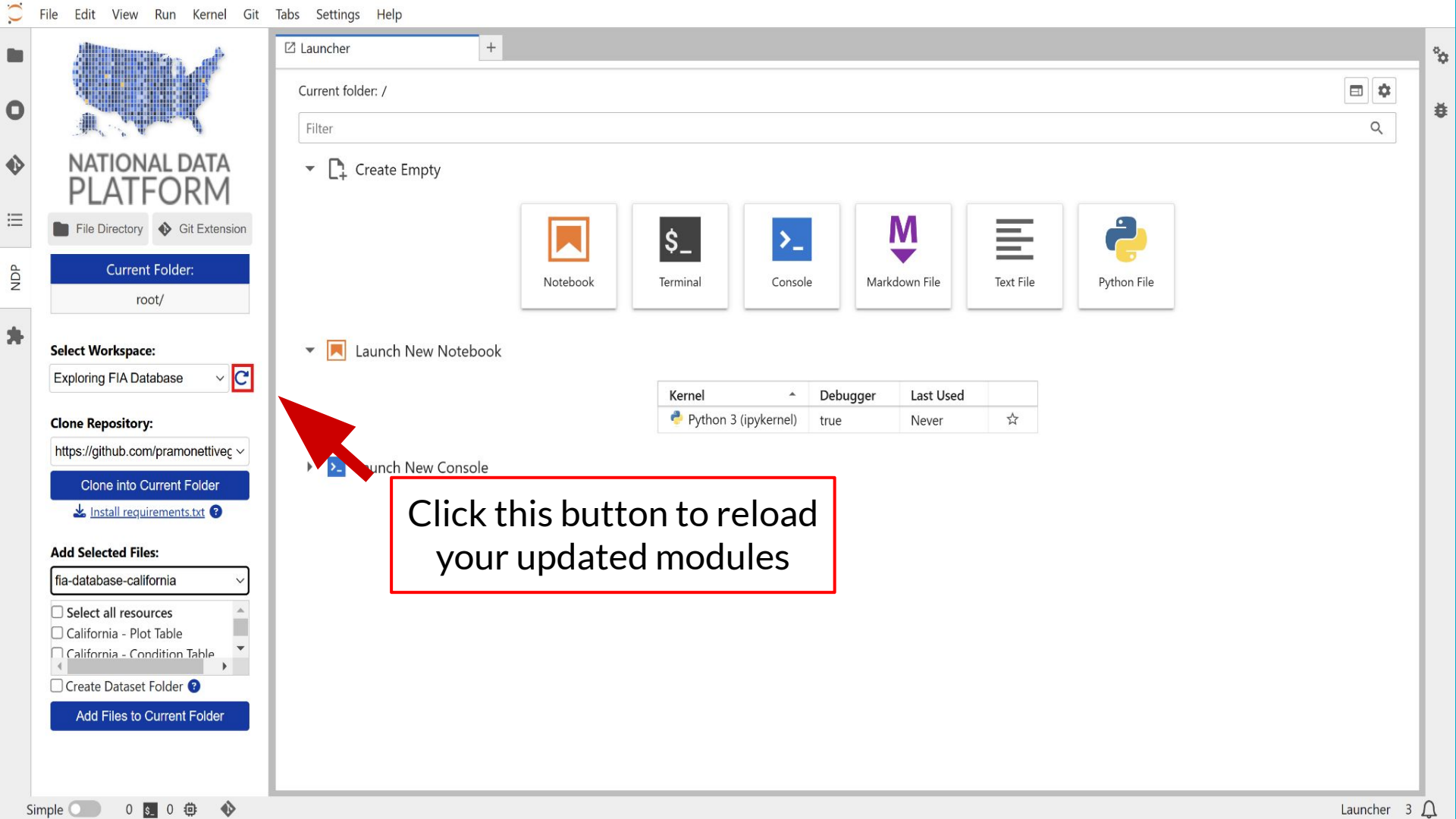
Natural Conifer Regeneration Probability

This metric is intended to be used to identify areas where reforestation may be necessary if stakeholders want to reestablish coniferous forests following fire. Conifers in our region generally lack the capacity to resprout after fi...

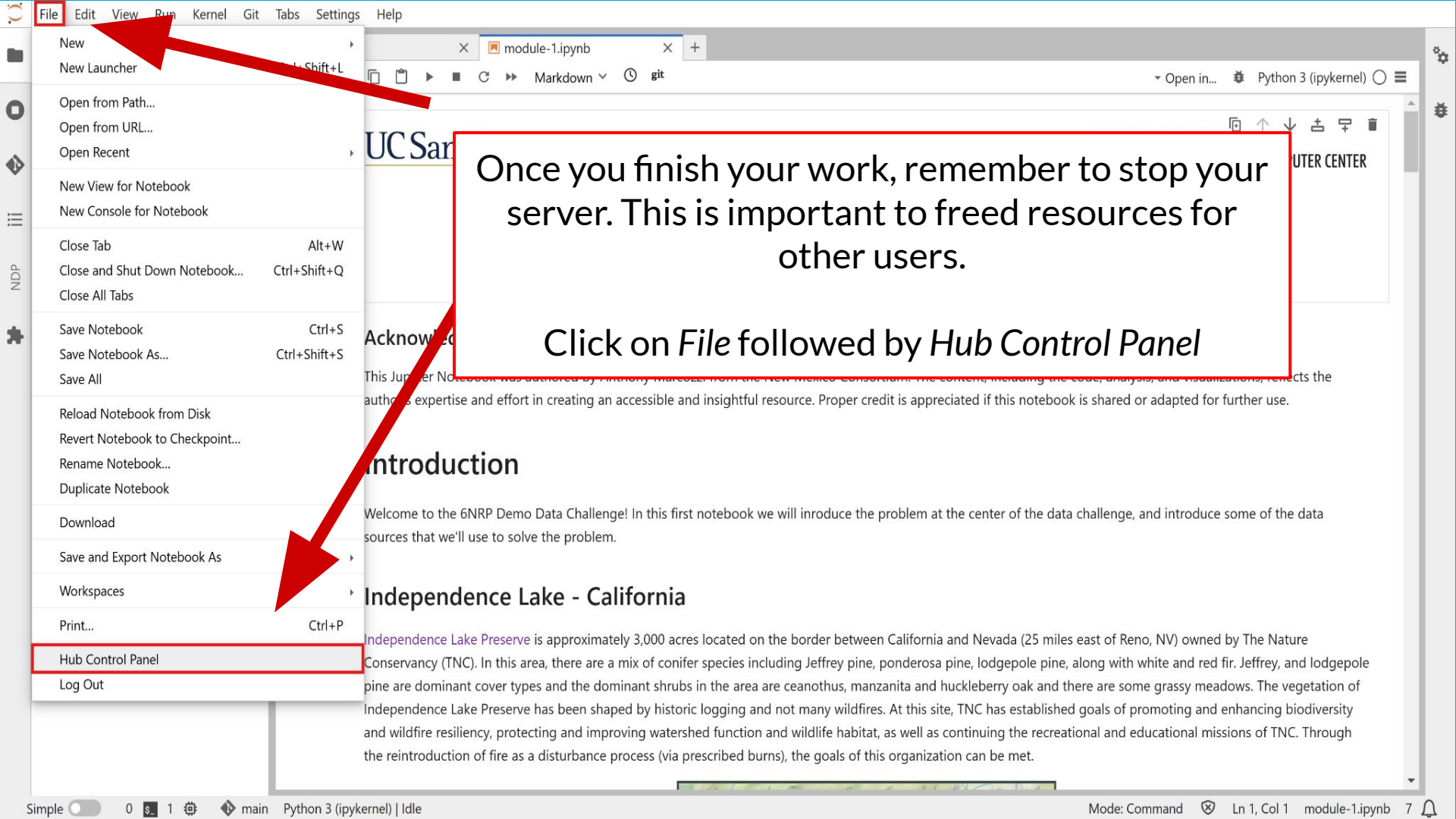
View More

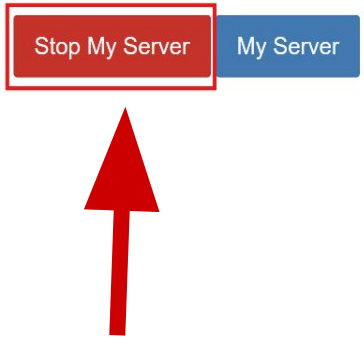
Search for the dataset of your interest and add it to your module

Go to JupyterHub


File Edit View Run Kernel Git Tabs Settings Help
NATIONAL DATA PLATFORM
File Directory
Git Extension
Current Folder:
root/
Select Workspace:
Exploring FIA Database
Clone Repository:
https://github.com/pramonettiveg
Clone into Current Folder
Install requirements.txt
Add Selected Files:
fia-database-california
Select all resources
California - Plot Table
California - Condition Table
Create Dataset Folder
Add Files to Current Folder
Launcher
Current folder: /
Filter
Create Empty
Notebook
Terminal
Console
Markdown File
Text File
Python File
Launch New Notebook
Kernel
Debugger
Last Used
Python 3 (ipykernel)
true
Never
Launch New Console
Click this button to reload your updated modules
Simple
Launcher 3

7. Stop your server


File
Edit
View
Run
Kernel
Git
Tabs
Settings
Help
New
New Launcher
Open from Path...
Open from URL...
Open Recent
New View for Notebook
New Console for Notebook
Close Tab Alt+W
Close and Shut Down Notebook... Ctrl+Shift+Q
Close All Tabs
Save Notebook Ctrl+S
Save Notebook As... Ctrl+Shift+S
Save All
Reload Notebook from Disk
Revert Notebook to Checkpoint...
Rename Notebook...
Duplicate Notebook
Download
Save and Export Notebook As
Workspaces
Print... Ctrl+P
Hub Control Panel
Log Out
module-1.ipynb
Markdown
Open in...
Python 3 (ipykernel)
Once you finish your work, remember to stop your server. This is important to freed resources for other users.
Click on File followed by Hub Control Panel
Introduction
Welcome to the 6NRP Demo Data Challenge! In this first notebook we will inroduce the problem at the center of the data challenge, and introduce some of the data sources that we'll use to solve the problem.
Independence Lake - California
Independence Lake Preserve is approximately 3,000 acres located on the border between California and Nevada (25 miles east of Reno, NV) owned by The Nature Conservancy (TNC). In this area, there are a mix of conifer species including Jeffrey pine, ponderosa pine, lodgepole pine, along with white and red fir. Jeffrey, and lodgepole pine are dominant cover types and the dominant shrubs in the area are ceanothus, manzanita and huckleberry oak and there are some grassy meadows. The vegetation of Independence Lake Preserve has been shaped by historic logging and not many wildfires. At this site, TNC has established goals of promoting and enhancing biodiversity and wildfire resiliency, protecting and improving watershed function and wildlife habitat, as well as continuing the recreational and educational missions of TNC. Through the reintroduction of fire as a disturbance process (via prescribed burns), the goals of this organization can be met.
Simple
main
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
module-1.ipynb



And finally click on *Stop Server*. Wait for the red button to disappear as a confirmation.