NATIONAL DATA PLATFORM

# National Data Platform

## CICI National Scale Data Fabric Presentation

August 21, 2025

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

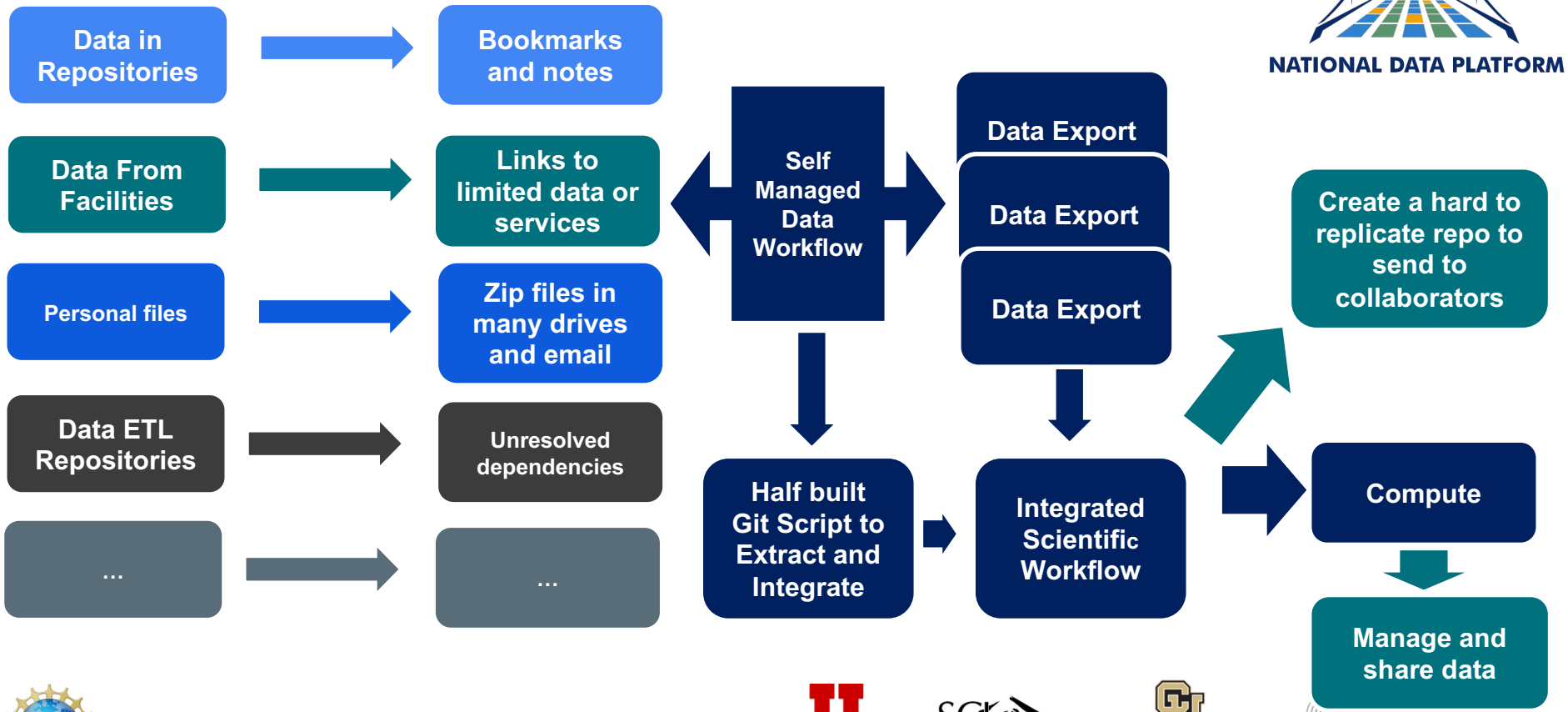
THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# Working with data can be frustrating!




NATIONAL DATA PLATFORM

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# Why is data so challenging to use?

NATIONAL DATA PLATFORM

Fragmented access to data, models and services

Complexity of formats, interoperability and context

Disconnect between producing & sharing data

Lack of user experiences enabling collaboration

Hard to customized service deployment practices

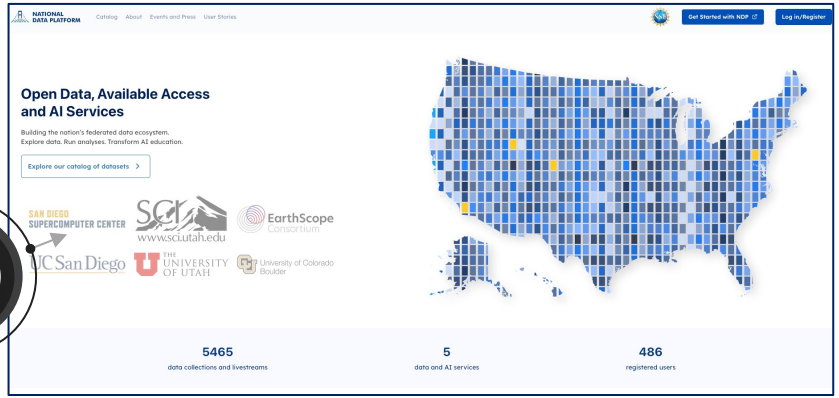
Need for scalable or specialized compute readiness

Disconnect between research & education capacity

# National Data Platform tackles the challenges around using and sharing data!

A **broad**, **federated** and **extensible** data ecosystem to promote collaboration, innovation and customizable use of data on top of existing and future national cyberinfrastructure capabilities.

NATIONAL DATA PLATFORM


https://www.nationaldataplatform.org/


**Centralized portal** for discovery, access and use workspaces for research and education

A scalable **platform** for using, developing and deploying composable services and application workflows at **distributed endpoints**

#2333609

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

# NDP Flow of Capabilities

NATIONAL DATA PLATFORM

**Discover**

Find data, services and models in catalog

**Catalog**

Add additional workflow inputs to catalog

**Endpoint Deployment**

Find data and models in catalog

**Collect**

Create a workspace with to begin analysis

**Collaborate**

Invite collaborators to shared workspace & storage

**Compute**

Execute workflow on the cloud

**Store**

Export final data product

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# What does NDP help with?

NATIONAL DATA PLATFORM

## NDP Facilitates Data Cataloging to Make Data FAIR and Used at Scale.

Searchable catalogs with standardized metadata and vocabularies integrate data from different sources — research, government agencies, and facilities. Data agents for impact and usage tracking across datasets.

## NDP Provides Collaborative Workspaces and Customizable Services to Use Data.

Once you find data, you can explore it and analyze it right on the platform on a browser — no need to download huge files or set up special software. You can also bring your own data, and create near data services, models, and workflows.

## NDP Connects Users to National Cyberinfrastructure and Cloud Resources.

Users can access existing services or deploy their own services through a standardized software stack to any data or computing resource, so they can process big data effectively— even from a laptop.

## NDP Enables Development and Deployment of AI-integrated Workflows for Science and Society.

NDP provides intuitive user experiences and extensible services for users to build AI-powered solutions by making data ready for training, building, and testing machine learning models as part of their workflows.

## NDP Offers Tools to Create Classroom and Data Challenge Learning Experiences.

Through partnerships with educators and open course materials, NDP supports hands-on learning in data science, AI, and scientific computing.

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

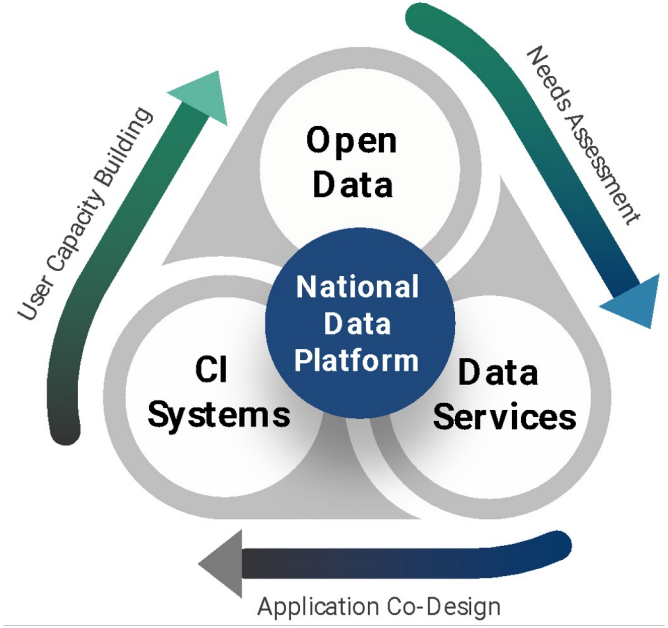
SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# Our Use and User-Inspired Co-Design Approach

NATIONAL DATA PLATFORM




**Current NDP User Personas:**

- **Researcher:** find and user data in research workflows: run data/compute- intensive analyses; collaborate across institutions; publish reproducible findings
- **Researcher-Educator:** integrate open datasets and tools into research-driven course materials; publish reproducible notebooks; train students on real-world data
- **Data Provider-Steward:** ensure datasets are FAIR and usable; track usage; support documentation and metadata standards
- **Student-Learner:** learn data science methods using real data; build and share projects; contribute to team-based work
- **Research Software Engineers:** build near data experiences using the standardized endpoints; create service stacks deployable on compute by user community; create custom branded hubs on top of NDP

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

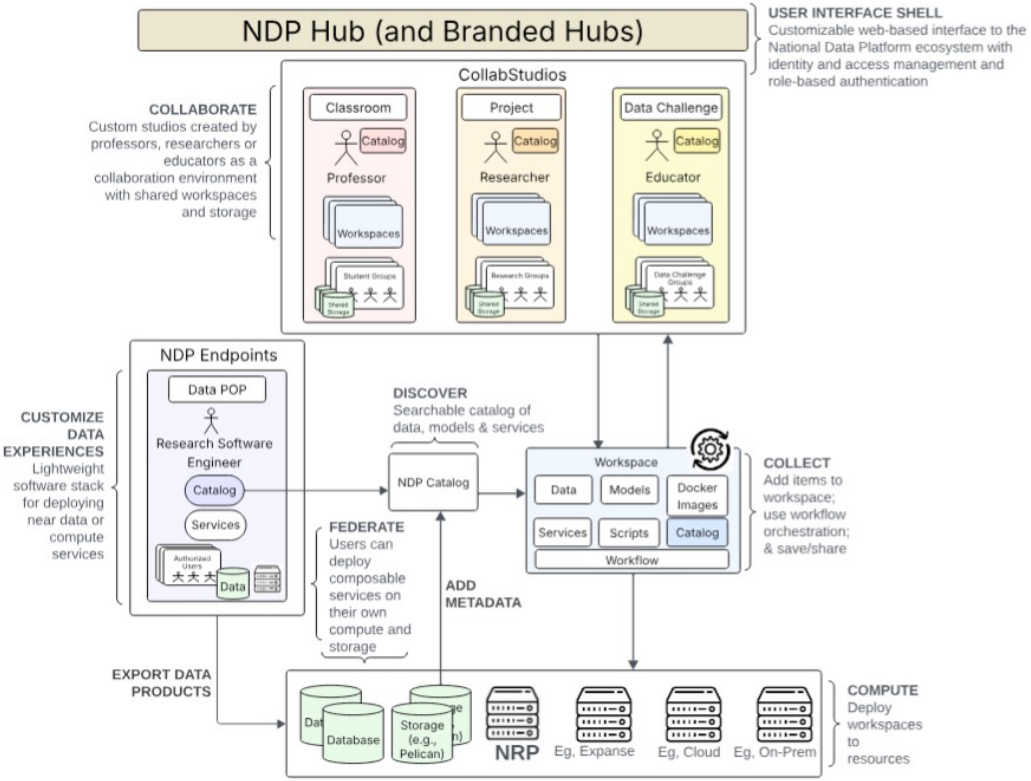
THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# How does NDP do it?




NATIONAL DATA PLATFORM

- Browser-based UI
- Keycloak authentication and authorization
- FAIR data catalog with multifaceted search
- Standard Jupyter widget service linking workspaces to infrastructure
- Customizations through the platform:
  - -- Robust data ingestion pipeline
  - -- Deployable customizable near data endpoints with tracking
  - -- Personal workspaces
  - -- Collaborative Studios for classroom, project and data challenges
  - -- Bring your own compute resources
  - -- OSDF/Pelican namespace integration
  - -- Hub shells for branded external hubs

# NDP and CICI Projects: What can you do on NDP?

NATIONAL DATA PLATFORM

DATA CATALOG INTEGRATION

AI AND SCIENTIFIC WORKFLOWS

CREATE CUSTOM SHELLS

NDP ENDPOINT DEPLOYMENT NEAR YOUR DATA

OFFER MATERIALS FOR CLASSROOM CAPABILITIES and CREATE DATA CHALLENGES


**İlkay ALTINTAŞ**
ialtintas@ucsd.edu

**Manish PARASHAR**
manish.parashar@utah.edu



NSF #2333609


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

# NDP Flow of Capabilities

NATIONAL DATA PLATFORM

**Discover**

Find data, services and models in catalog

**Catalog**

Add additional workflow inputs to catalog

**Endpoint Deployment**

Find data and models in catalog

**Collect**

Create a workspace with to begin analysis

**Collaborate**

Invite collaborators to shared workspace & storage

**Compute**

Execute workflow on the cloud

**Store**

Export final data product

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

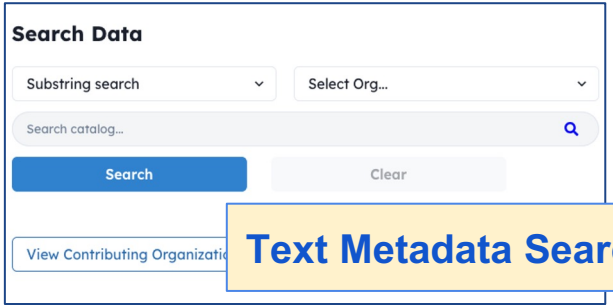
University of Colorado Boulder

EarthScope Consortium

# Data Discovery and Usage Exploration

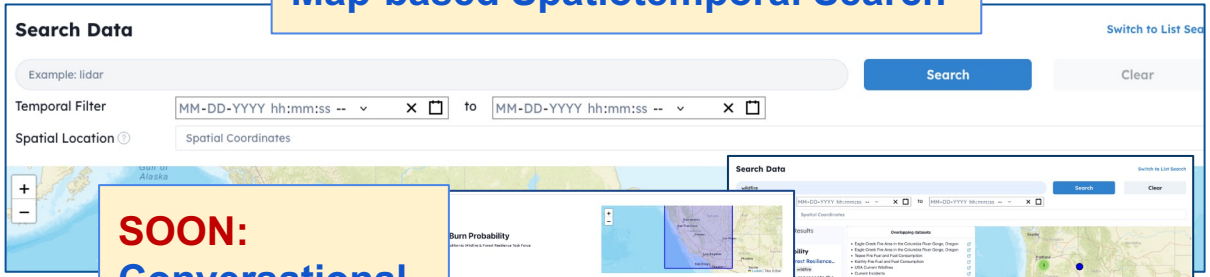
NATIONAL DATA PLATFORM

## Multimodal Search and Exploration Interfaces



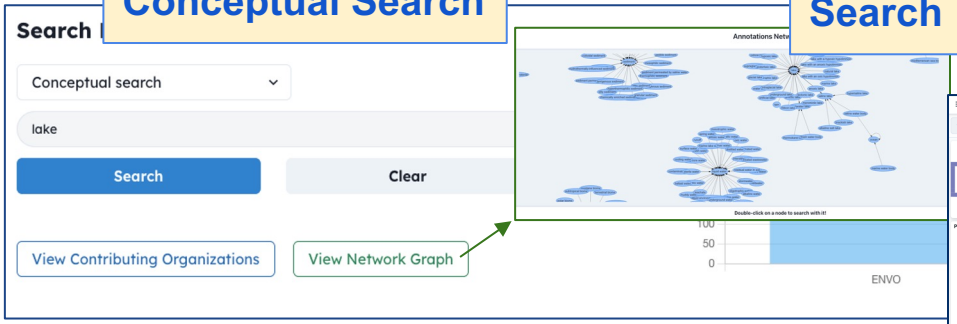

**Text Metadata Search**

**Map-based Spatiotemporal Search**

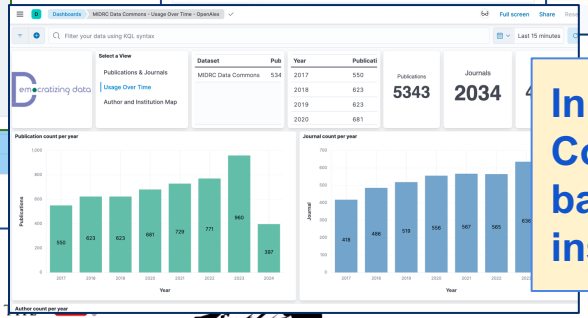



**SOON: Conversational Search**

**Conceptual Search**






**In progress: Contextual Search based on usage insights**

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

UNIVERSITY OF UTAH

www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# NDP Flow of Capabilities

NATIONAL DATA PLATFORM

**Discover**

Find data, services and models in catalog

**Catalog**

Add additional workflow inputs to catalog

**Endpoint Deployment**

Find data and models in catalog

**Collect**

Create a workspace with to begin analysis

**Collaborate**

Invite collaborators to shared workspace & storage

**Compute**

Execute workflow on the cloud

**Store**

Export final data product

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

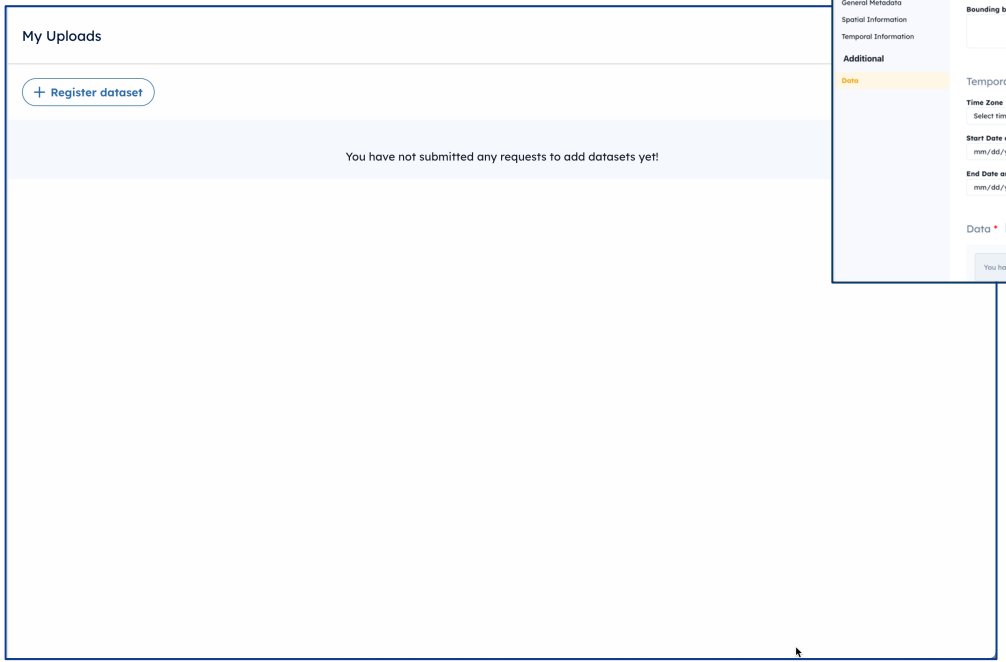
University of Colorado Boulder

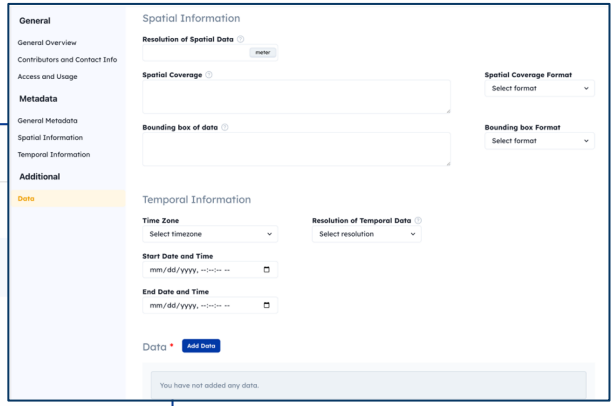
EarthScope Consortium

# Data Ingestion and Cataloging

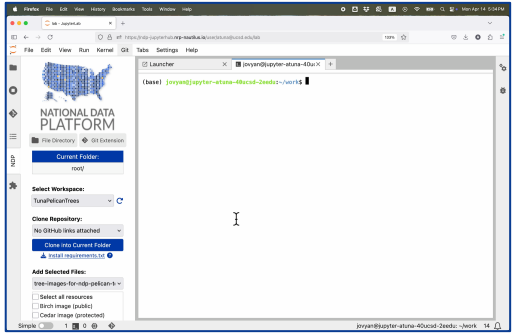
NATIONAL DATA PLATFORM

## Formal Data Ingestion into the NDP Catalog




## OSDF Integration




#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# NSF HydroFrame and Hydrogen Digital Twin NAIRR Pilot

**Contacts:** Reed Maxwell, Princeton University and Laura Condon, University of Arizona

- Hydrologic simulations and datasets, currently listed at https://hf-hydrodata.readthedocs.io/en/latest/available_datasets.html
- **NDP catalogs** are waiting for final data steward approval for public access
  - Textual, conceptual, and map-based spatiotemporal search capabilities
- Linking data to **AI workflows through NDP JupyterHub on NRP**
  - Physics-guided machine learning and digital twin applications
  - Application currently a NAIRR Pilot - https://nairrpilot.org/projects/demo/hydrogen
- **NDP Data Challenge** using this dataset is being discussed

NATIONAL DATA PLATFORM

**A national framework for hydrologic modeling and scientific discovery**

Our goal is to make national hydrologic simulations and datasets more accessible. We have a variety of tools available to access data, build models and learn more about our national watersheds and hydrology.

https://hydroframe.org

NSF#2134892

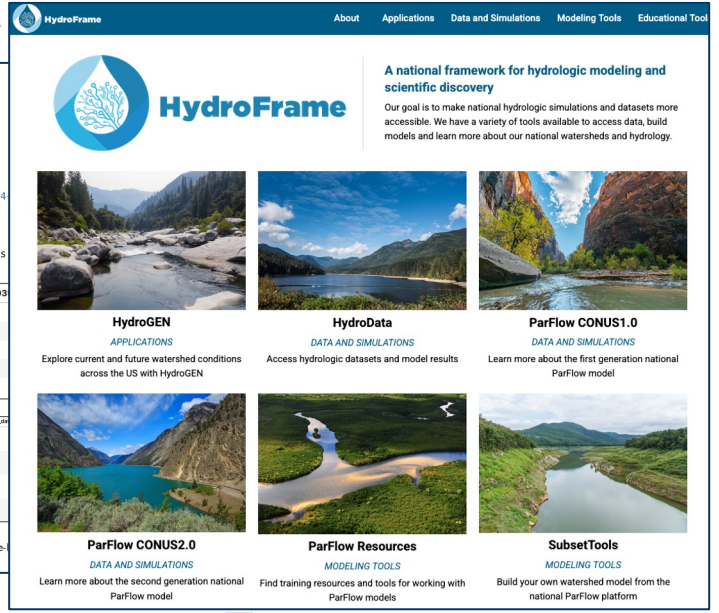
```
import hf_hydrodata

point_parameters = {'dataset': 'usgs_nwis',
                    'variable': 'water_table_depth',
                    'temporal_resolution': 'daily',
                    'aggregation': 'mean',
                    'latitude_range': (38.749, 41.485),
                    'longitude_range': (-106.207, -100.695),
                    'date_start': '2003-10-01', 'date_end': '2004
                    }

point_data = hf_hydrodata.get_point_data(point_parameters)
point_metadata = hf_hydrodata.get_point_metadata(point_parameters
```

| | date | 393358103454800 | 393902103554000 | 393902103554001 | 393 |
|---|---|---|---|---|---|
| 0 | 2003-10-01 | 2.941320 | 3.392424 | 3.617976 | |
| 1 | 2003-10-02 | 2.932176 | 3.386328 | 3.605784 | |
| 2 | 2003-10-03 | 2.950464 | 3.392424 | 3.605784 | |
| 3 | 2003-10-04 | 2.953512 | 3.389376 | 3.608832 | |
| 4 | 2003-10-05 | 2.956560 | 3.392424 | 3.605784 | |

| | site_id | site_name | site_type | agency | state | latitude | longitude | first_date_data_available | last_date_da |
|---|---|---|---|---|---|---|---|---|---|
| 0 | 393358103454800 | SCO060S703EAB DTX5 BEAVER CREEK | groundwater well | USGS | CO | 39.566111 | -103.763333 | 2000-02-24 | |
| 1 | 393902103554000 | SCO050S696BED DTX9 MUDDY CREEK | groundwater well | USGS | CO | 39.650596 | -103.927778 | 2000-04-26 | |
| 2 | 393902103554001 | SCO050S696BD1 DTX10A MUDDY CREEK | groundwater well | USGS | CO | 39.650596 | -103.927778 | 2000-04-26 | |
| 3 | 393902103554003 | SCO050S696BD3 DTX11 MUDDY CREEK | groundwater well | USGS | CO | 39.650596 | -103.927778 | 2000-04-26 | |

Figure 1: Image of example site-level point observations DataFrame and select site-l
returned by the provided example function calls.

https://joss.theoj.org/papers/10.21105/joss.06623




#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

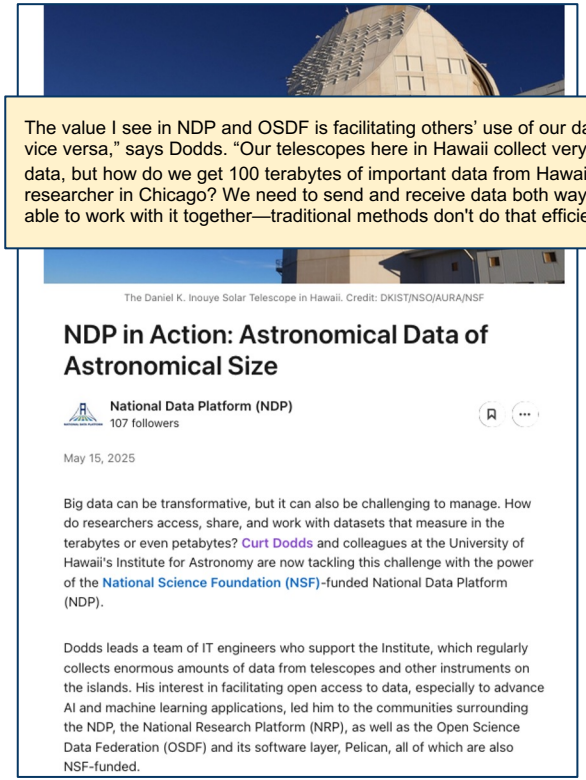
EarthScope Consortium

# Spectropolarimetric Inversion in Four Dimensions with Deep Learning

**Contact:** Curt Dodds, University of Hawaii

SPIn4D, NSF#2008344

NATIONAL DATA PLATFORM



The Daniel K. Inouye Solar Telescope in Hawaii. Credit: DKIST/NSO/AURA/NSF

**NDP in Action: Astronomical Data of Astronomical Size**

National Data Platform (NDP)
107 followers

May 15, 2025

Big data can be transformative, but it can also be challenging to manage. How do researchers access, share, and work with datasets that measure in the terabytes or even petabytes? Curt Dodds and colleagues at the University of Hawaii's Institute for Astronomy are now tackling this challenge with the power of the National Science Foundation (NSF)-funded National Data Platform (NDP).

Dodds leads a team of IT engineers who support the Institute, which regularly collects enormous amounts of data from telescopes and other instruments on the islands. His interest in facilitating open access to data, especially to advance AI and machine learning applications, led him to the communities surrounding the NDP, the National Research Platform (NRP), as well as the Open Science Data Federation (OSDF) and its software layer, Pelican, all of which are also NSF-funded.

https://www.linkedin.com/pulse/ndp-action-astronomical-data-size-national-data-platform-lfwnc/

The value I see in NDP and OSDF is facilitating others' use of our data, and vice versa," says Dodds. "Our telescopes here in Hawaii collect very unique data, but how do we get 100 terabytes of important data from Hawaii to a researcher in Chicago? We need to send and receive data both ways and be able to work with it together—traditional methods don't do that efficiently."

- Data from telescopes and other instruments as well as valuable solar simulation data **stored on OSDF** via Pelican interface
  - Through 2021-2023, the SPIn4D team modeled and ran solar simulations using 10 million CPU hours on the NSF's Cheyenne supercomputer, producing a massive dataset of 110 terabytes.
- **NDP catalogs** for multimodal data is being developed as public and private catalogs
  - Textual, conceptual, and map-based spatiotemporal search capabilities
  - 13 TB of data was cataloged and published to NDP users
  - Jupyter notebooks to explore the data was published to NDP users
- Linking data to **AI workflows through NDP JupyterHub on NRP**
  - Deep learning based models are being developed to be published to NDP
  - Application is currently a NAIRR Pilot - https://nairrpilot.org/projects/demo/spin4d
- **NDP Education Hub** will be used to deliver online, interactive course content using this data to introduce students to solar spectropolarimetry

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# Biologging for Marine Animal Behavior and Physiology

**Contact:** Jessica Kendall-Bar, UC San Diego

NATIONAL DATA PLATFORM

- High-resolution multimodal data (5-10TB of video, sensor data on electrophysiology, GPS) **stored on OSDF** via Pelican interface
- **NDP catalogs** for multimodal data is being developed as public and private catalogs
  - Textual, conceptual, and map-based spatiotemporal search capabilities

Dr. Jessica Kendall-Bar shares lessons on environmental data visualizations.

**NDP in Action: Tracking the Data Lifecycle from Field to Impact**

National Data Platform (NDP)
132 followers

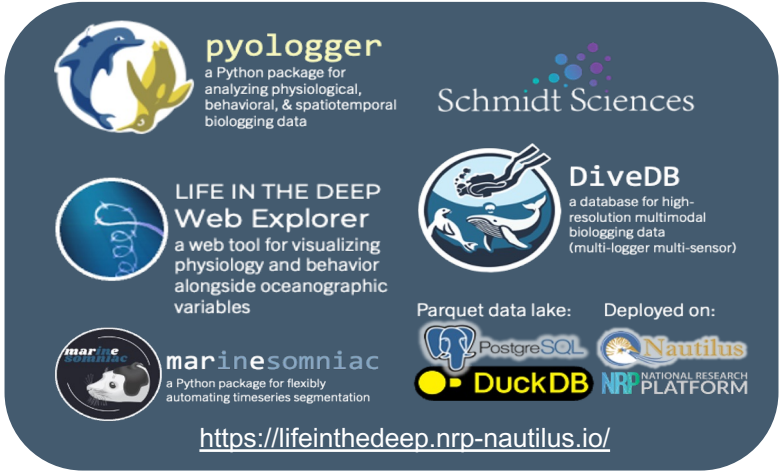
August 14, 2025

The journey of Dr. Jessica Kendall-Bar's data may start on the back of an elephant seal in the deep ocean, but its impact can stretch all the way to government policy and clinical research. But keeping track of the data, from collection and analysis to publication and real-world applications, is another story altogether—one the NSF National Data Platform (NDP) was built to tell.

As a postdoctoral research fellow in animal ecophysiology at Scripps Institution of Oceanography at UC San Diego, Kendall-Bar studies the interactions

https://www.linkedin.com/pulse/ndp-action-tracking-data-lifecycle-from-field-impact-xpdac

- Linking data to **AI workflows through NDP JupyterHub on NRP**
  - Using the marinesomniac library for timeseries segmentation
- **NDP Data Challenge** using this dataset is being planned for 2026
- Current link to application running with NRP as default endpoint is https://lifeinthedeep.nrp-nautilus.io/




#2333609

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

# NDP Flow of Capabilities

NATIONAL DATA PLATFORM

**Discover**

Find data, services and models in catalog

**Catalog**

Add additional workflow inputs to catalog

**Endpoint Deployment**

Find data and models in catalog

**Collect**

Create a workspace with to begin analysis

**Collaborate**

Invite collaborators to shared workspace & storage

**Compute**

Execute workflow on the cloud

**Store**

Export final data product

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

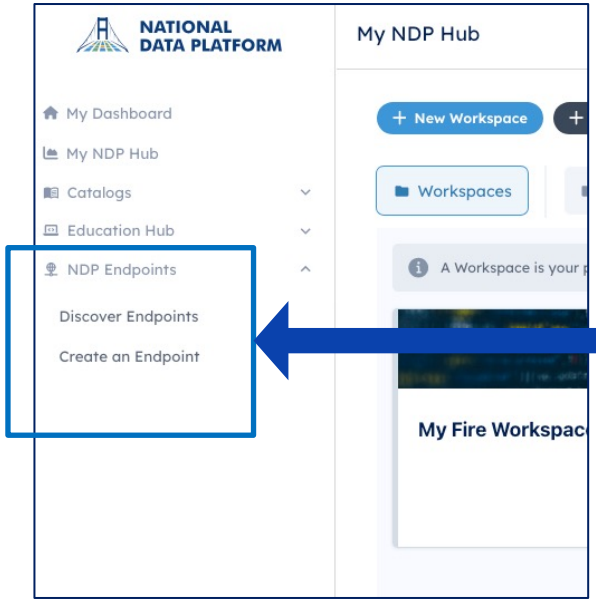
SCI www.sci.utah.edu

University of Colorado Boulder

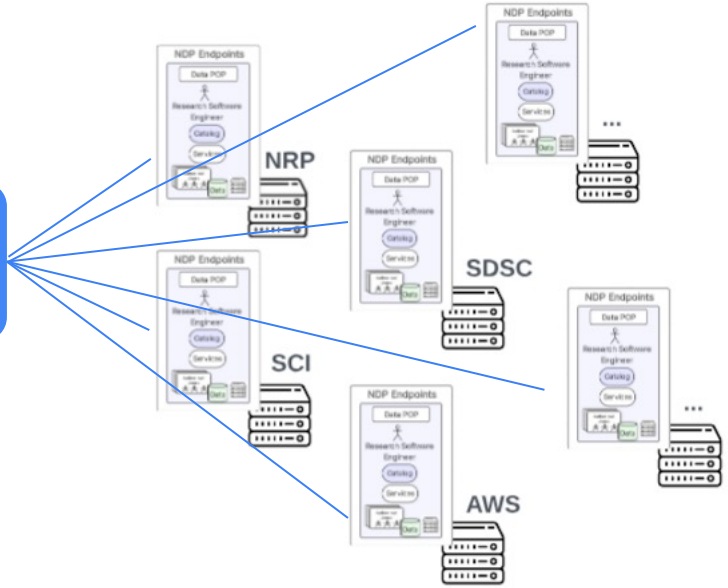
EarthScope Consortium

# Endpoint Deployment on Composable Resources and Federation

NATIONAL DATA PLATFORM

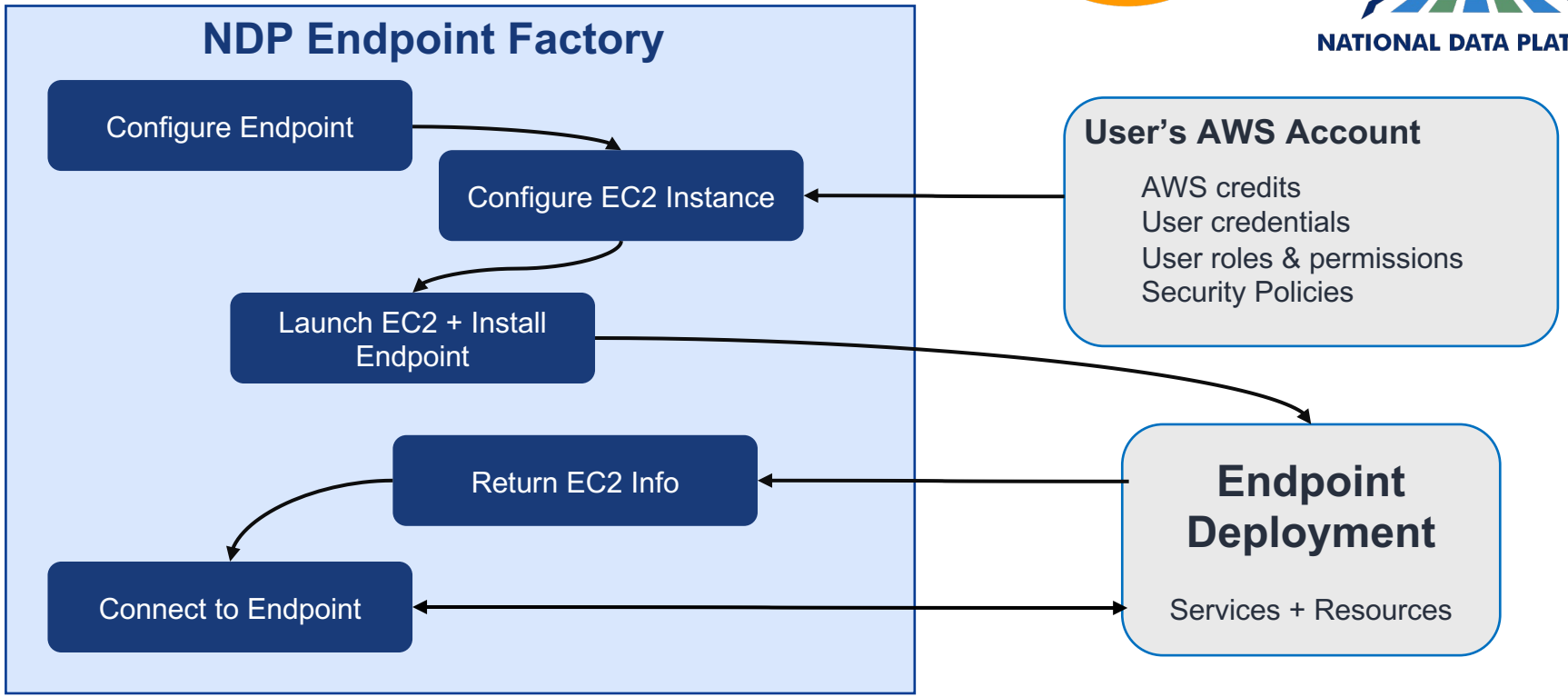



Endpoint Federation



# Launching Endpoints on aws

NATIONAL DATA PLATFORM




#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

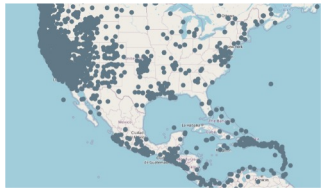
EarthScope Consortium

# EarthScope Data Streaming Workflows

**Contact:** Charles Meertens, University of Colorado and Dave Mencin, EarthScope Consortium

NATIONAL DATA PLATFORM

EarthScope Consortium



**Live Streamed & Archived Sensor Data**

3D Global Navigation Satellite System (GNSS)

Ground displacement measurements

High temporal rate (1 Hz)

~ 1000 stations (geo-located)

Subscription

SciDx Science Data Exchange

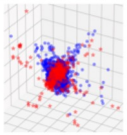
NDP Endpoint with SciDx Streaming

aws

Historical data storage

OSDF/Pelican origin

AI Analytics

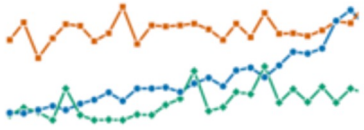


Trigger HTC workflows

AI model retraining

OSG

Real-time visualization of (derived) streams



Register metadata, derived streams

NATIONAL DATA PLATFORM HUB

- Data search & discovery
- Workspace access

- **NDP Endpoint** with SciDX Streaming Service and JupyterHub deployed on AWS close to EarthScope data facility
- Streams are visualized and shown through custom endpoint interfaces (Jupyter notebook or web)
- Derived streams are registered to **NDP Catalog** with metadata and made discoverable
  - Textual, conceptual, and map-based spatiotemporal search capabilities
- The streams are also uploaded to OSDF as back historical data so **AI workflows** can be enabled on OSG

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

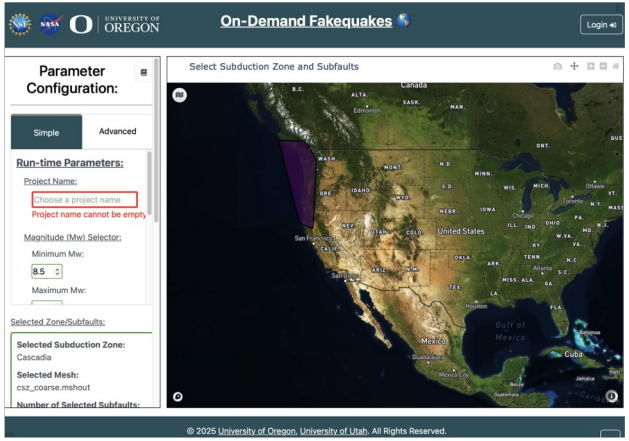
University of Colorado Boulder

EarthScope Consortium

# On Demand FakeQuakes (ODF)

**Contact:** Diego Melgar, University of Oregon

NATIONAL DATA PLATFORM

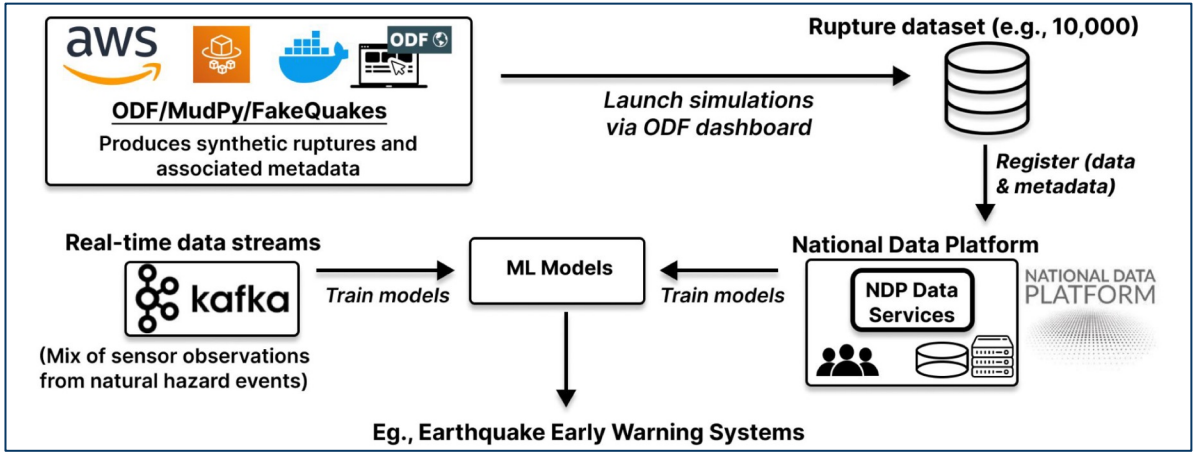



https://ondemandfakequakes.org

CRESCENT CASCADIA REGION EARTHQUAKE SCIENCE CENTER

NSF#2225286

- CRESCENT simulation outputs are ingested into **NDP Catalog**
  - Limits duplicated simulation runs
  - Search and discovery
- **NDP Endpoints** enable leveraging near-data services on AWS to stitch data together




NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

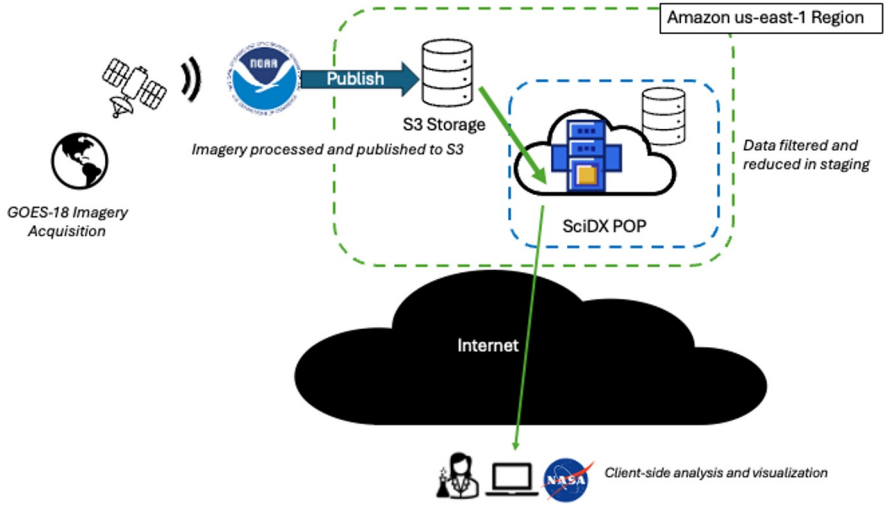
University of Colorado Boulder

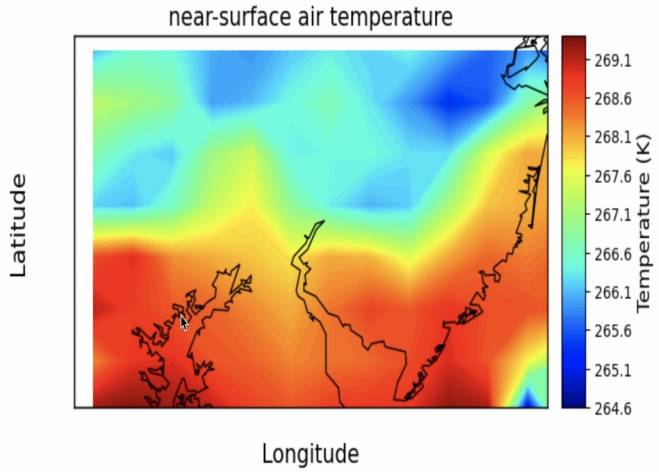
EarthScope Consortium

# NASA/NOAA Satellite Data

**Contact:** Andrew Michaelis, NASA Ames Research Center

NATIONAL DATA PLATFORM




- Multimodal high resolution satellite and modeling data
- Ingestion and discovery via **NDP Catalog**
- Management through **NDP Endpoint** on AWS with SciDX services

SciDx Science Data Exchange

SciDx Data Staging:

- near data processing
- subsetting and resampling
- transferring data outputs only

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# NDP Flow of Capabilities

NATIONAL DATA PLATFORM

**Discover**

Find data, services and models in catalog

**Catalog**

Add additional workflow inputs to catalog

**Endpoint Deployment**

Find data and models in catalog



**Collect**

Create a workspace with to begin analysis

**Collaborate**

Invite collaborators to shared workspace & storage

**Compute**

Execute workflow on the cloud

**Store**

Export final data product

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

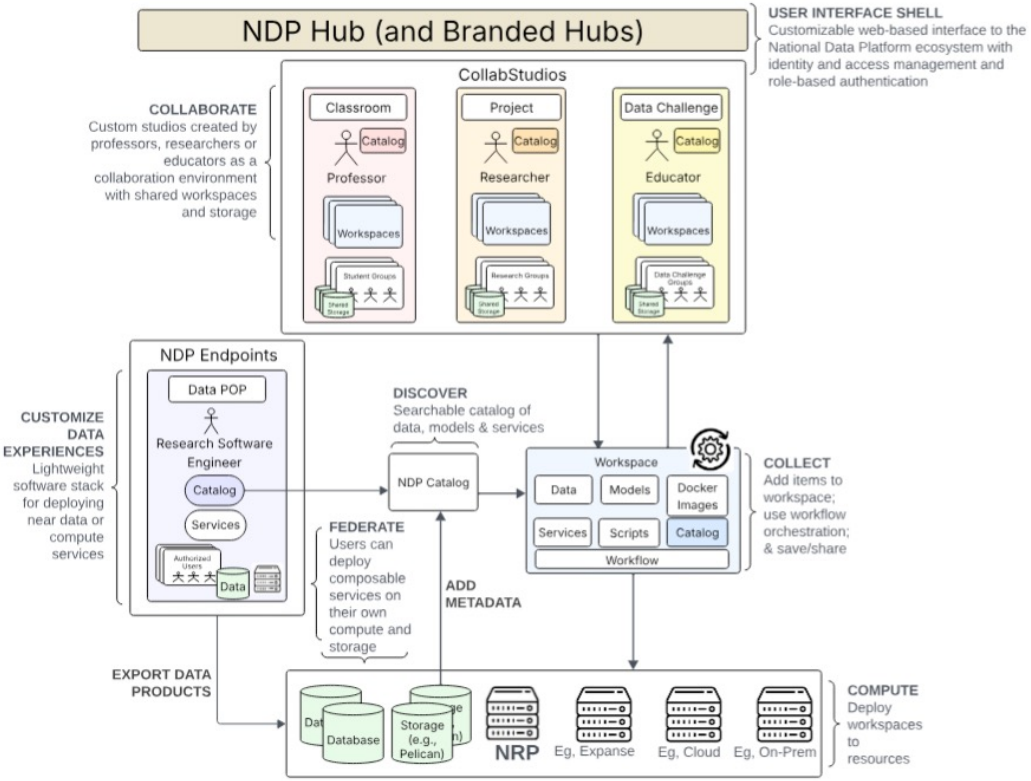
THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

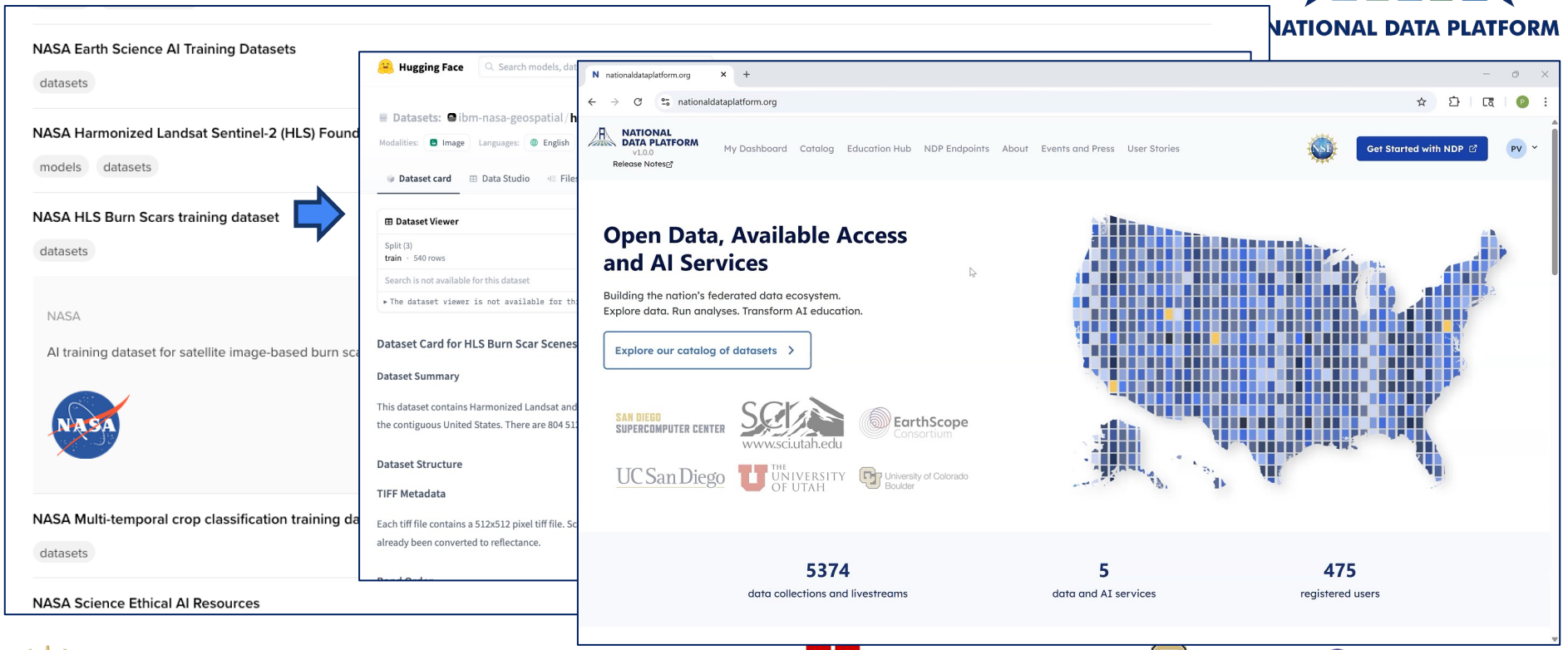
# How does NDP do it?



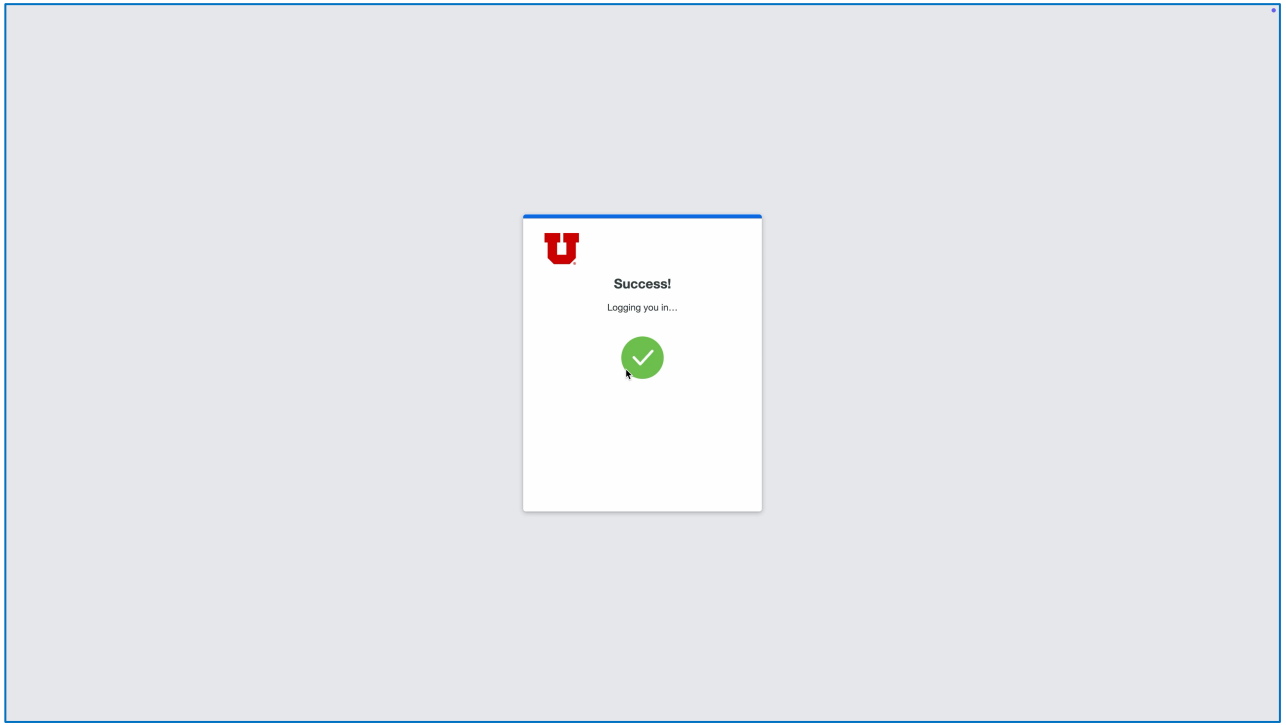

NATIONAL DATA PLATFORM

- Browser-based UI
- Keycloak authentication and authorization
- FAIR data catalog with multifaceted search
- Standard Jupyter widget service linking workspaces to infrastructure
- Customizations through the platform:
  - -- Robust data ingestion pipeline
  - -- Deployable customizable near data endpoints with tracking
  - -- Personal workspaces
  - -- Collaborative Studios for classroom, project and data challenges
  - -- Bring your own compute resources
  - -- OSDF/Pelican namespace integration
  - -- Hub shells for branded external hubs

# Hugging Face Data and Model Integration

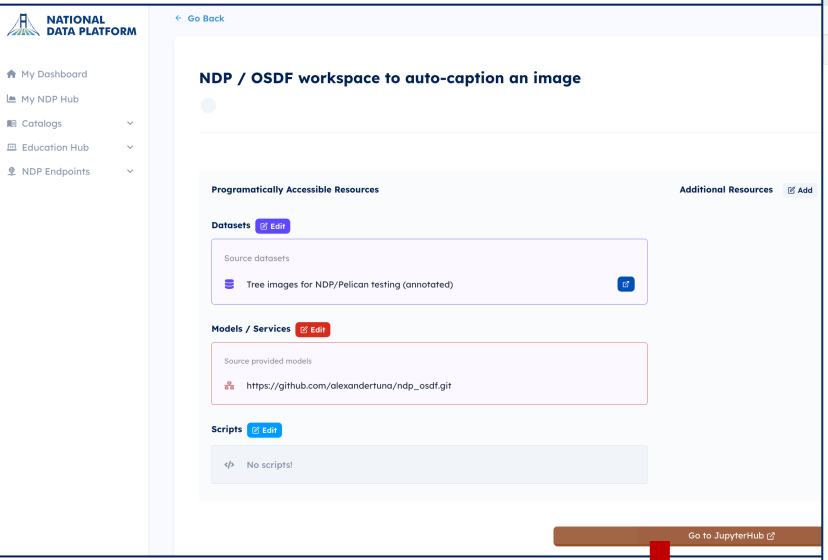


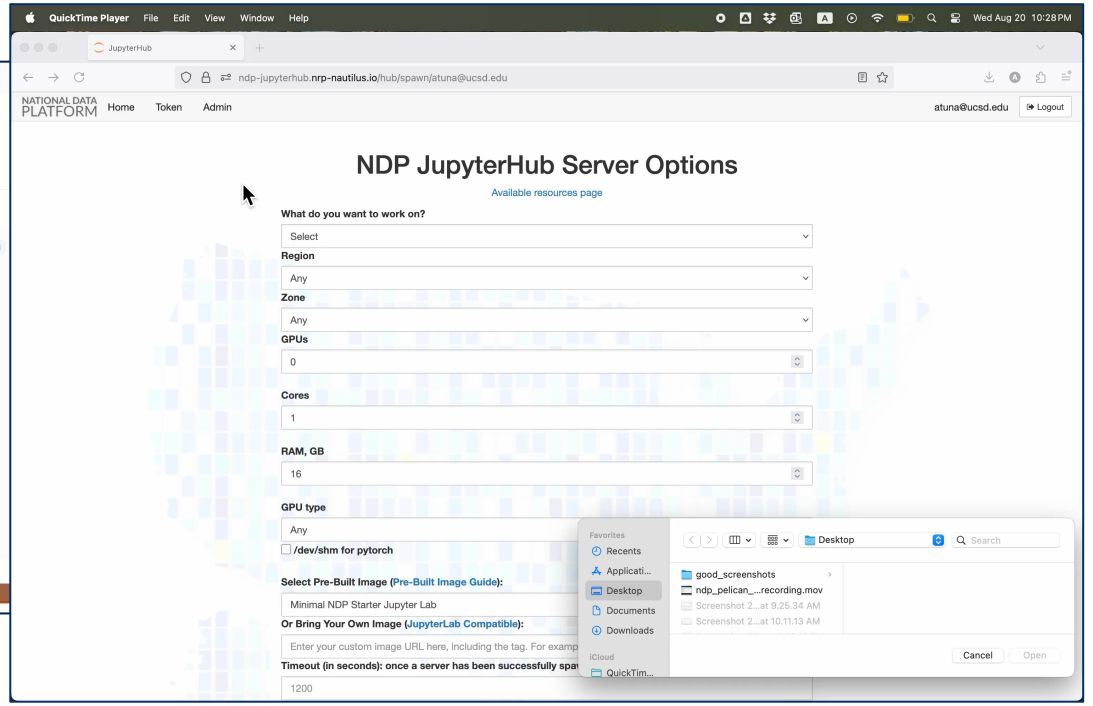
# Users can select compute endpoints…




NATIONAL DATA PLATFORM

# Workspace OSDF Pelican Integration




#2333609

# Graduate Data Science Class Assignments for Knowledge Graphs

**Professor:** Amarnath Gupta, UC San Diego

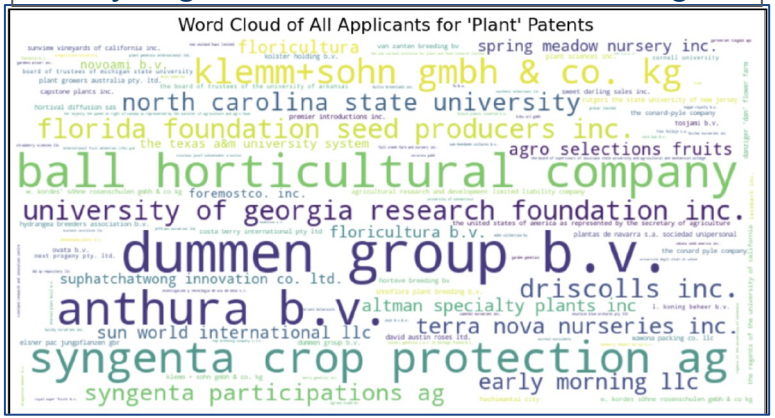
NATIONAL DATA PLATFORM

- In Fall 2024, 6 student groups taking DSE 203 accessed multiple data sets available through the **NDP Catalog**
  - Each data set belonged to a different data model using integration and interoperability of data through NDP
  - Total data set size: 40 GB
- **NDP Education Hub** workspaces was used to construct knowledge graphs based on instructions given in class
- Several groups used the **National Research Platform** as a default resource via NDP for tasks like matching textual data to ontologies
  - Data resources private to UC San Diego were provided by extending the **NDP JupyterHub** to use NRP resources at UCSD
- The integrated knowledge graph was stored in an **NDP-Administered Neo4J Graph Database** for each group
- Lessons learned from this first time classroom support includes providing NRP resources sustainably for long-running jobs

uspto UNITED STATES PATENT AND TRADEMARK OFFICE

**Example Student Project with NAIRR USPTO Dataset**

Analyzing USPTO Patents with Ontologies



# Training Students on AI using Big Landscape Data

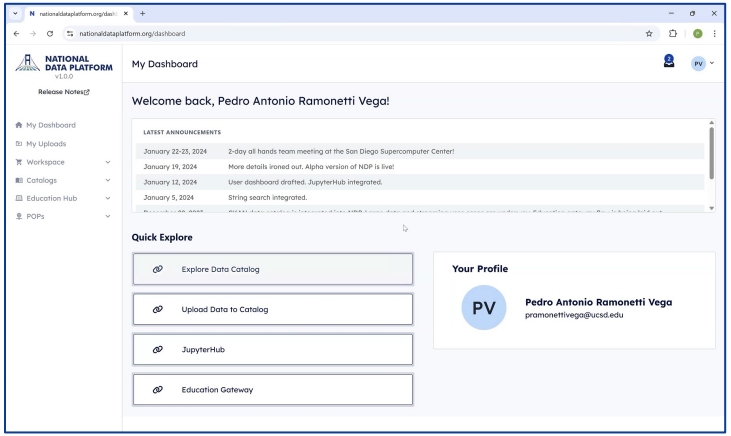
**Contacts:** Melissa Floca, UC San Diego; Russ Parsons, US Forest Service; Scott Pokswinski, New Mexico Consortium

NSF#2341120

NATIONAL DATA PLATFORM

Four sprints in the Education Hub, progressively training students in:

- Using large vegetation and fuel datasets (FIA, TLS, ALS, FastFuels) and labels from the **NDP Catalog**
  - Each team analyzed ~85GB of data using NDP JupyterHub on NRP
- Setting up scalable computing environments (Dask, Docker) on NRP
- Generating tree inventories from ALS
- Applying machine learning for predicting tree attributes
- Conducting full pipeline development and reporting
  - Challenge required building **AI Workflows using labeled data**





https://prowesscenter.org/datachallenge

Although this open challenge was advertised only to UC San Diego students, it attracted 48 students from 8 universities across the United States, working collaboratively on NDP in teams of 2–5 members. The students **self onboarded** through the **NDP Education Hub** training modules.

**List of Universities:** (1) University of California, San Diego (2) San Diego State University (3) University of California, Los Angeles (4) University of California, Berkeley (5) University of Southern California (6) Northern Arizona University (7) Oregon Institute of Technology (8) Colorado State University

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

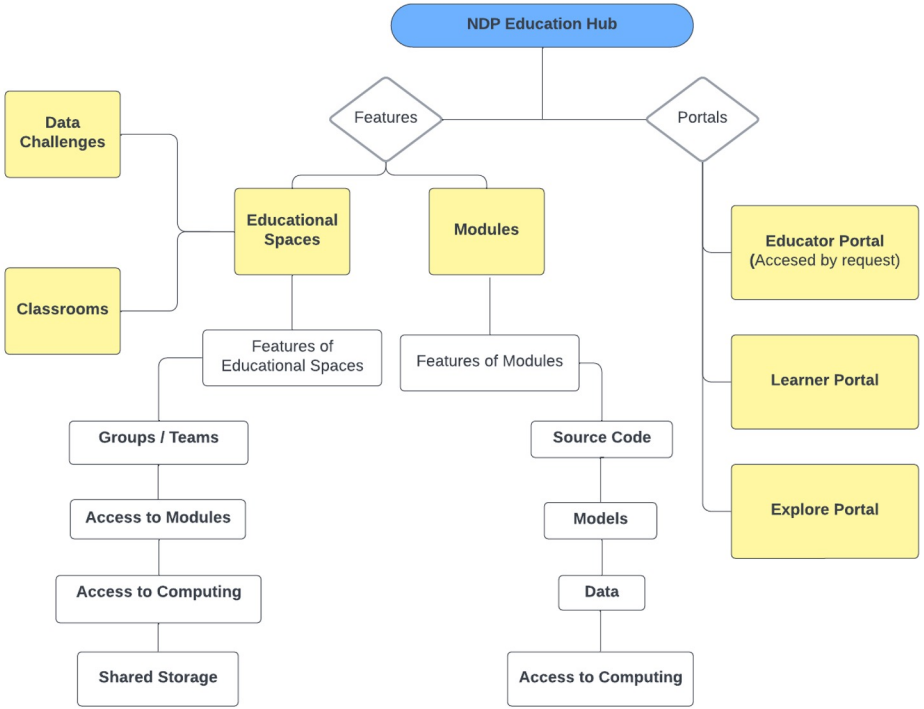
University of Colorado Boulder

EarthScope Consortium

# Compute-Friendly Education Workflows on NDP

**Contact:** Pedro Ramonetti, UC San Diego

NATIONAL DATA PLATFORM




"*Toward an Education Hub Linking Research Data and Compute to Learning Workflows in the National Data Platform*," Melissa Floca, Kate O'Laughlin, Pedro Ramonetti, Amarnath Gupta, Ilkay Altintas, Manish Parashar, at ACM PEARC 2025

## Toward an Education Hub Linking Research Data and Compute to Learning Workflows in the National Data Platform

MELISSA FLOCA*, KATE O'LAUGHLIN*, PEDRO RAMONETTI*, AMARNATH GUPTA*, and ILKAY ALTINTAS*, University of California San Diego, USA
MANISH PARASHAR*, Utah University, USA

As demand for AI literacy and data science education grows, there is a critical need for infrastructure that bridges the gap between research data, computational resources, and educational experiences. Although national-scale research platforms increasingly provide access to data and compute, integration with educational use cases remains limited. To address this gap, we developed a first-of-its-kind Education Hub within the National Data Platform. This hub enables seamless connections between collaborative research workspaces, classroom environments, and data challenge settings. By leveraging shared infrastructure resources and lowering technical barriers, the Education Hub supports hands-on, data-driven learning at scale. This paper presents the design and implementation of the Education Hub, along with lessons learned from early adopters and case studies that highlight its use in university classrooms and national-scale data challenges. Our findings underscore the value of embedding education directly into the research data ecosystem and point toward future directions for building inclusive, scalable AI and data science education platforms.

NSF #2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

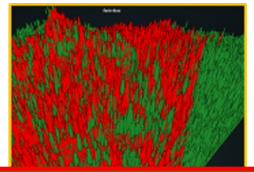
SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# BurnPro3D High-Resolution Vegetation and Fire Simulation Data

**Contact:** Leticia Lee, UC San Diego

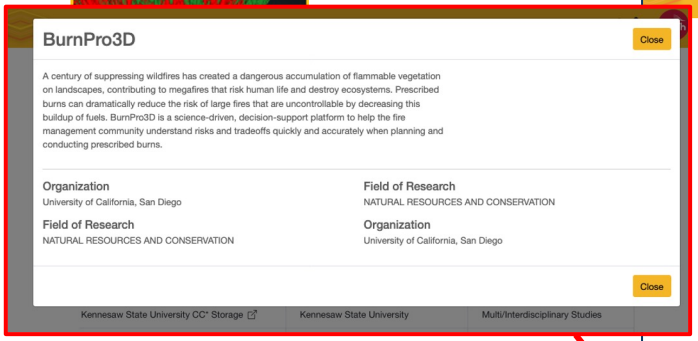


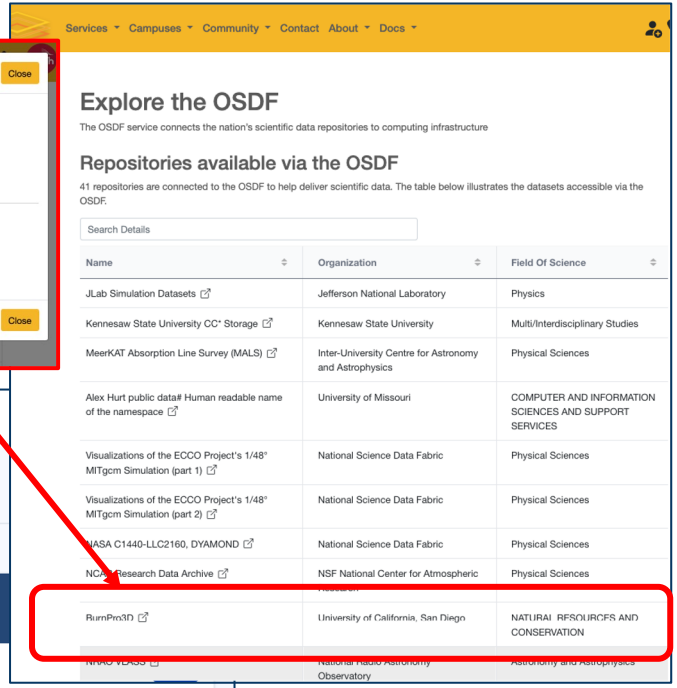
BurnPro3D NSF#2134892

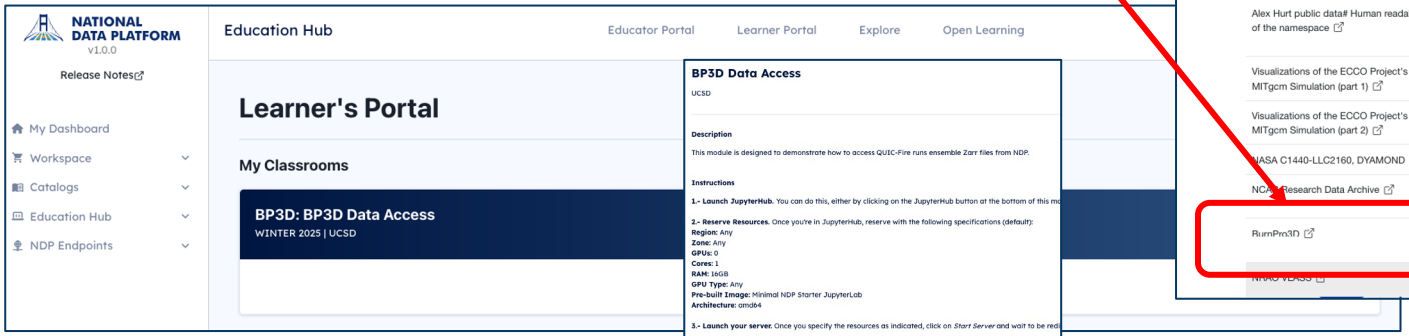
https://burnpro3d.sdsc.edu

NATIONAL DATA PLATFORM

- **NDP Catalogs** for BurnPro3D data on ODSF
- **NDP Education Hub Learner's Portal** provides training notebooks to access and analyze the data on NRP
- Data will be used in Agentic AI workflows within BurnPro3D







#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

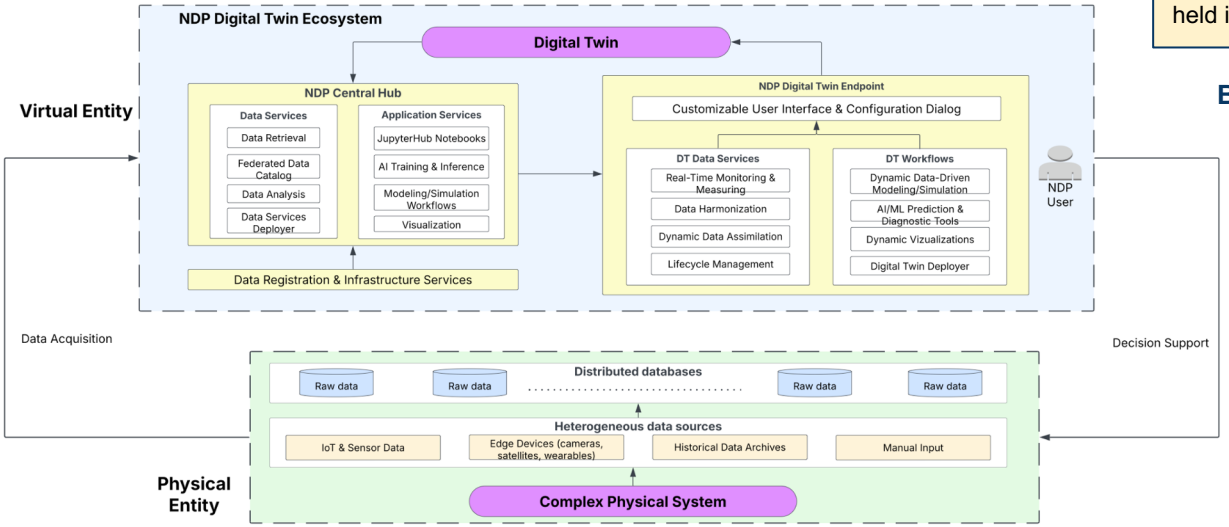
EarthScope Consortium

# Using NDP for Digital Twin Applications

**Contact:** Hena Ahmed, PhD Student, UC San Diego

NATIONAL DATA PLATFORM

## NDP Digital Twin Factory Architecture using Composable Services




"*Towards a Federated Approach to Complex Digital Twins*," Hena Ahmed, Daniel Crawl, Ilkay Altintas. FlexScience 2025 held in conjunction with HPDC 2025.

### Being tested for three digital twin applications

- Quantified Human based on multimodal data of Larry Smarr, by UC San Diego PhD student Conan Minihan.
  - Data is being uploaded to OSDF.
- WIFIRE Firemap based on the closed loop workflows for fire response, by Daniel Crawl and Hena Ahmed
  - AI workflows for data assimilation running on NRP.
- Immersive Forest based on multimodal fire environment data, by UC San Diego PhD student Isaac Nealey.
  - Spatiotemporal daya discovery, integration and AI-workflows using OSDF and NRP.

#2333609

UC San Diego

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

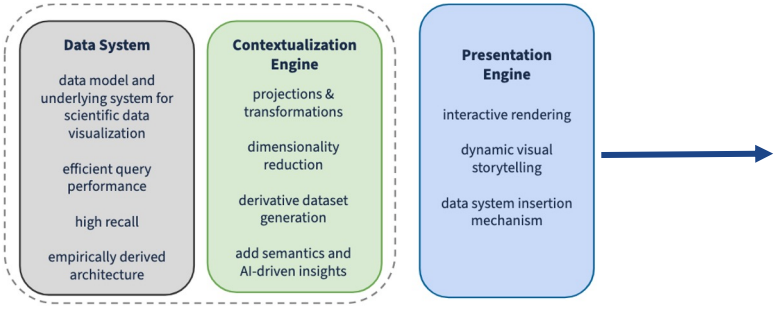
# Immersive Forest

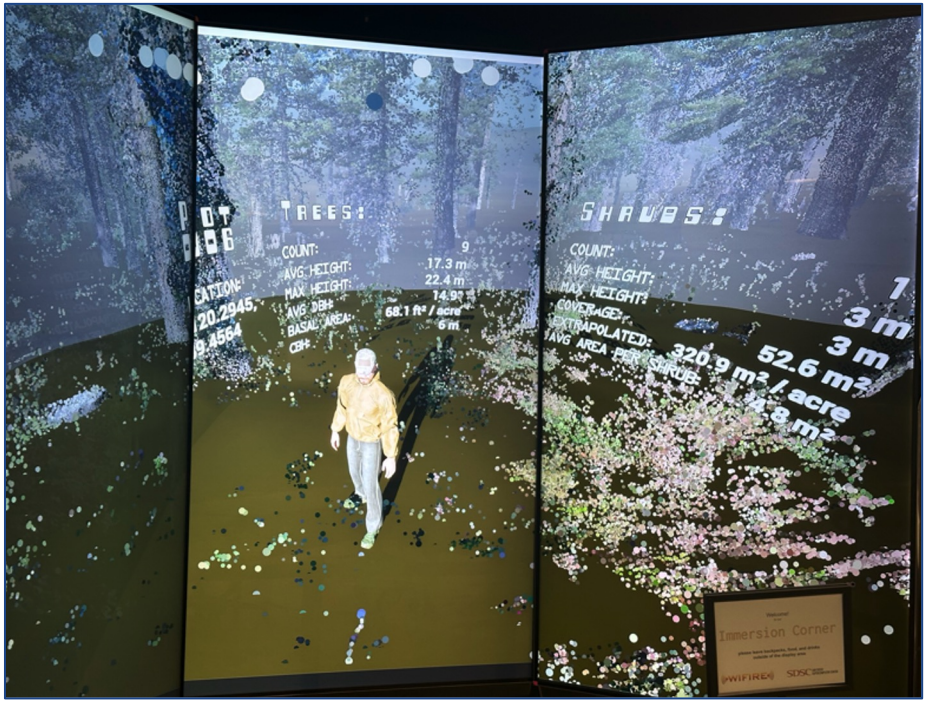
**Contact:** Isaac Nealey, PhD Student, UC San Diego

NSF#2341120

UNIVERSITY OF CALIFORNIA Office of the President

NATIONAL DATA PLATFORM




- Current calculations and AI workflows are deployed on NRP to integrate into the Immersive Forest offline

Future State:

- Fire environment data (satellite, LIDAR, weather and fire models) uploaded to OSDF will ingested into **NDP Catalog** to be discoverable.
- **NDP Endpoint** services will power the Immersive Forest, linking with **NDP Workspaces** to bring in new data and AI products into the visualization.
- **AI Workflows** integrated via **NDP Workspaces** running on NRP or other compute will generate new vegetation metrics and fire fuel characteristics to be integrated into the Immersive Forest.
- NDP will be used to link the data and contextualization to the presentation engine.

NATIONAL DATA PLATFORM

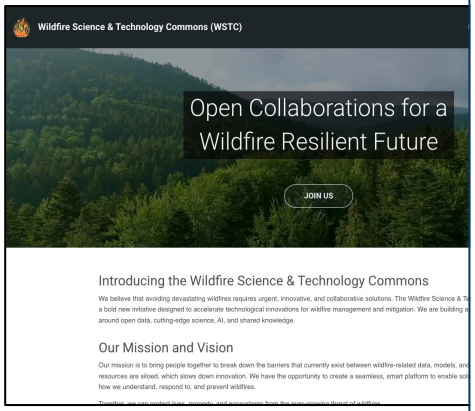
# NDP Shells

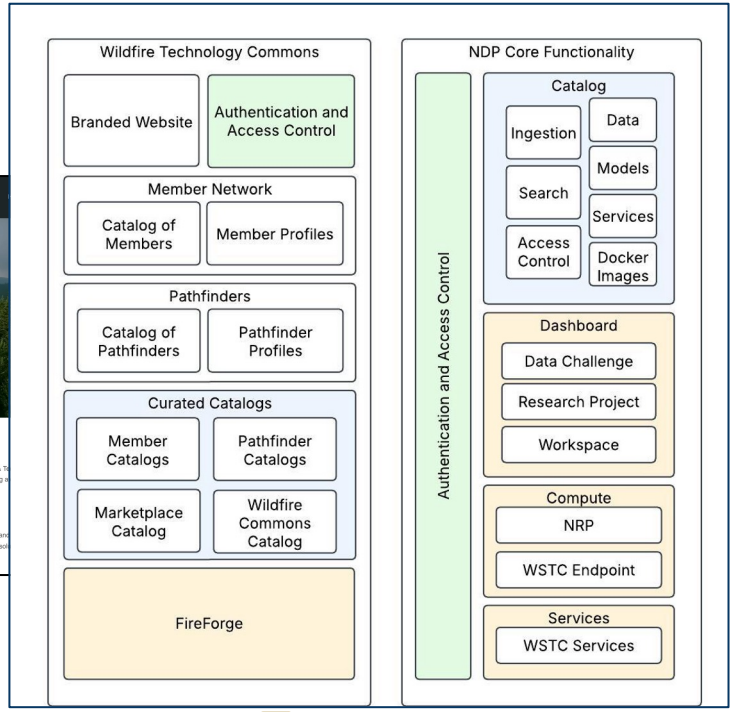
# Wildfire Science and Technology Commons

**Contact:** Claire Stirm, UC San Diego

https://www.wildfirecommons.org/

- **NDP Catalog** is being extended with catalogs for members, marketplace and pathfinders to serve a large network of wildfire experts
- FireForge platform is being built on top of an **NDP Endpoint** using **core NDP services** on Google cloud
  - Authentication, Catalogs, Project Spaces, Workspaces
- A customized look-n-feel expanding NDP endpoint backend with a specialized user experience
- Platform planned to launch July 1st, 2025
  - Version 1 with NDP compute backend and catalogs integrated into the marketplace with expert network services
  - Version 2 plans to use Democratizing Data as an NDP standard service for usage analysis




#2333609

# Wildfire Science & Technology Commons

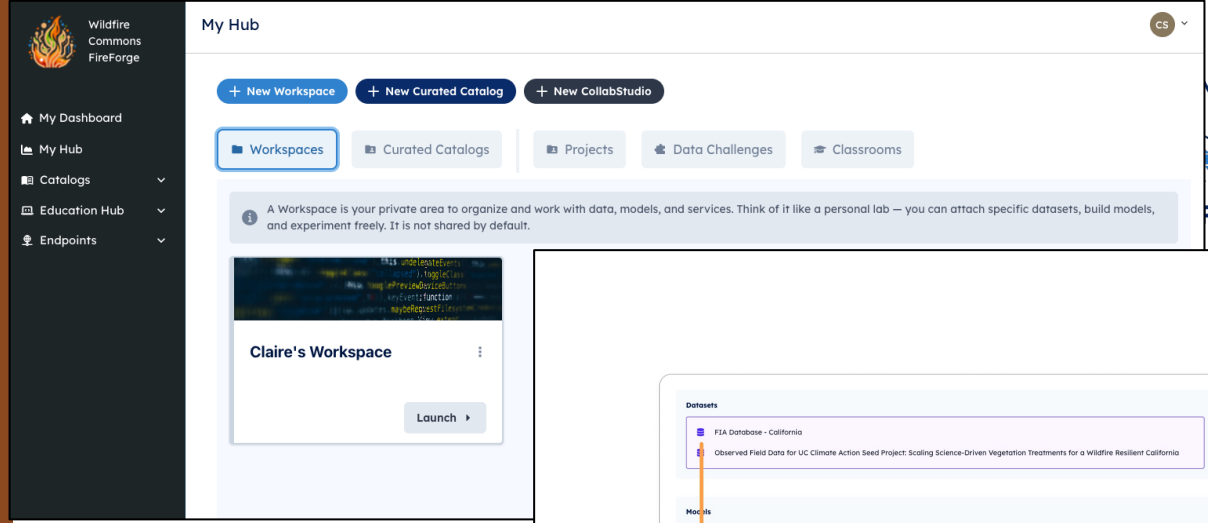
Expert Network Tool

FireForge Platform

Community Marketplace

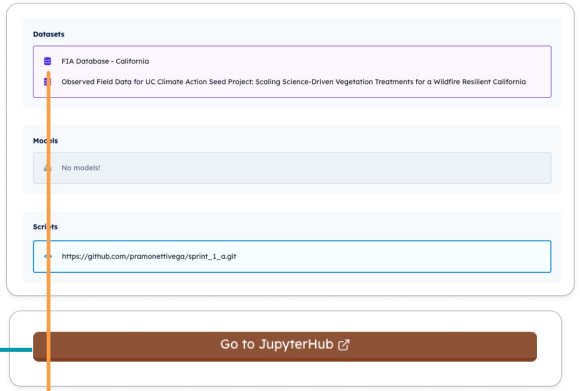
https://www.wildfirecommons.org

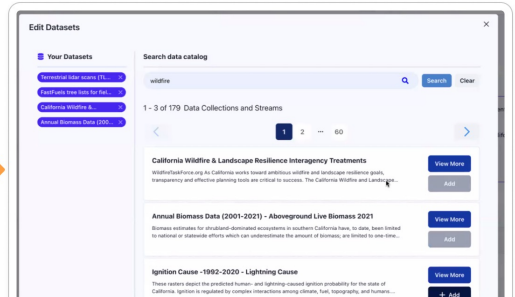
# FireForge Platform Interface









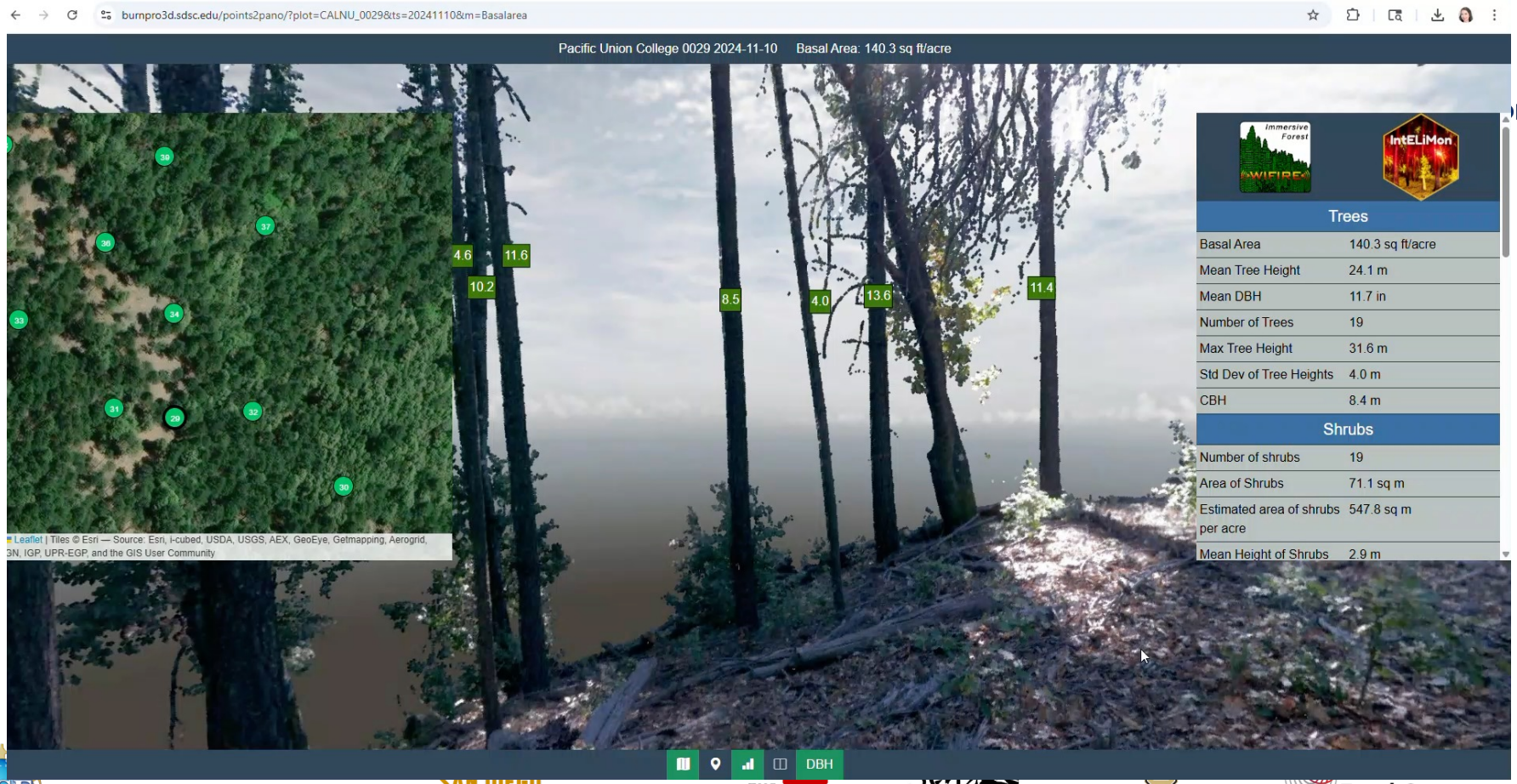

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

https://www.wildfirecommons.org

# Points2Pano app in WIFIRE's Immersive Forest

https://burnpro3d.sdsc.edu/points2pano





#2333609

# Example:

## TLS Processing Workflow Collaboration

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM

**CREATE WORKSPACE**

Create a workspace to collect data and scripts to accomplish goals

PREPARE DATA → ADD PROCESSING SCRIPTS

**CREATE CollabStudio PROJECT**

Share workspace with team members to collaborate on the project

ADD GROUP MEMBERS → LAUNCH WORKSPACE TO COMPUTE

Isaac Nealey

https://burnpro3d.sdsc.edu/points2pano

https://www.wildfirecommons.org

SAN DIEGO SUPERCOMPUTER CENTER

THE UNIVERSITY OF UTAH

SCI www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

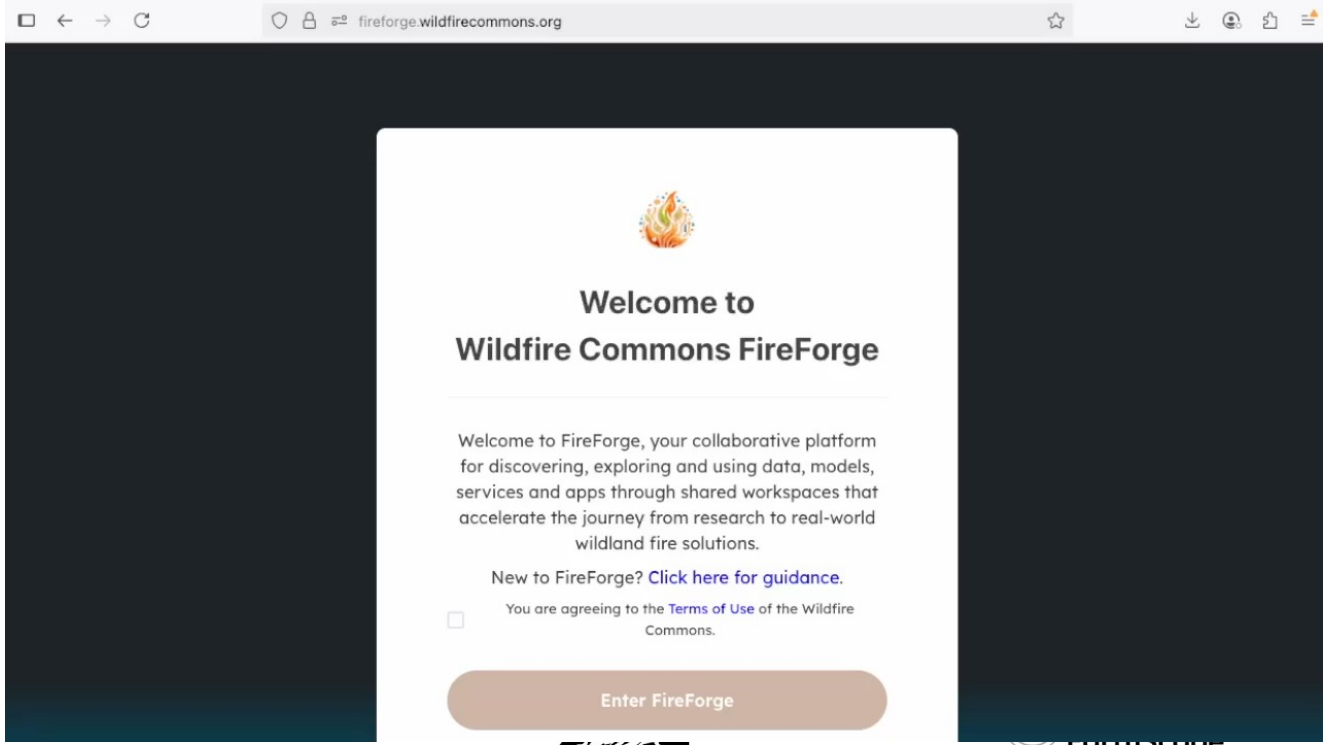
# Step 1:

## Prepare data

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM




https://www.wildfirecommons.org

SUPERCOMPUTER CENTER · UNIVERSITY OF UTAH · www.sci.utah.edu · University of Colorado Boulder · EarthScope Consortium

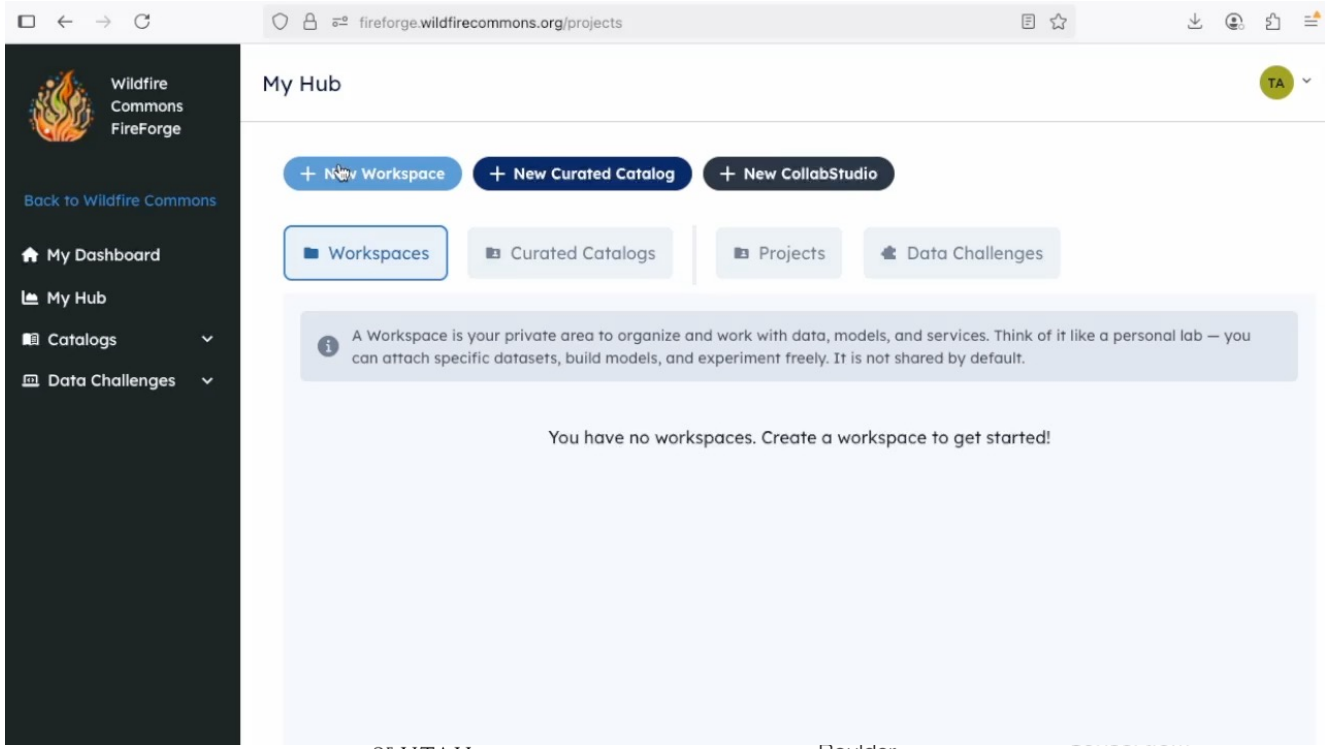
# Step 2:

## Create a workspace & project

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM




https://www.wildfirecommons.org

OF UTAH Boulder

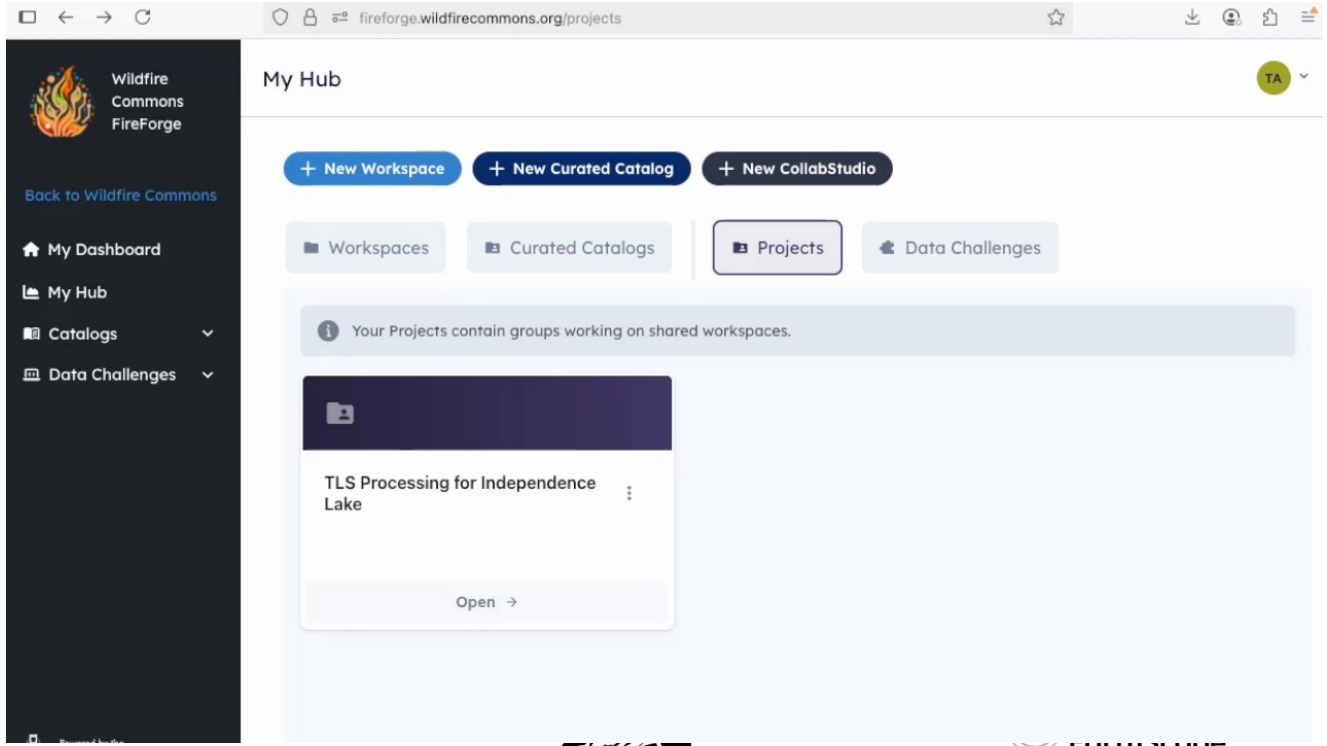
# Step 3:

## Jupyterhub

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM




Immersive Forest WIFIRE

IntELiMon

https://www.wildfirecommons.org

SUPERCOMPUTER CENTER

UNIVERSITY OF UTAH

www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

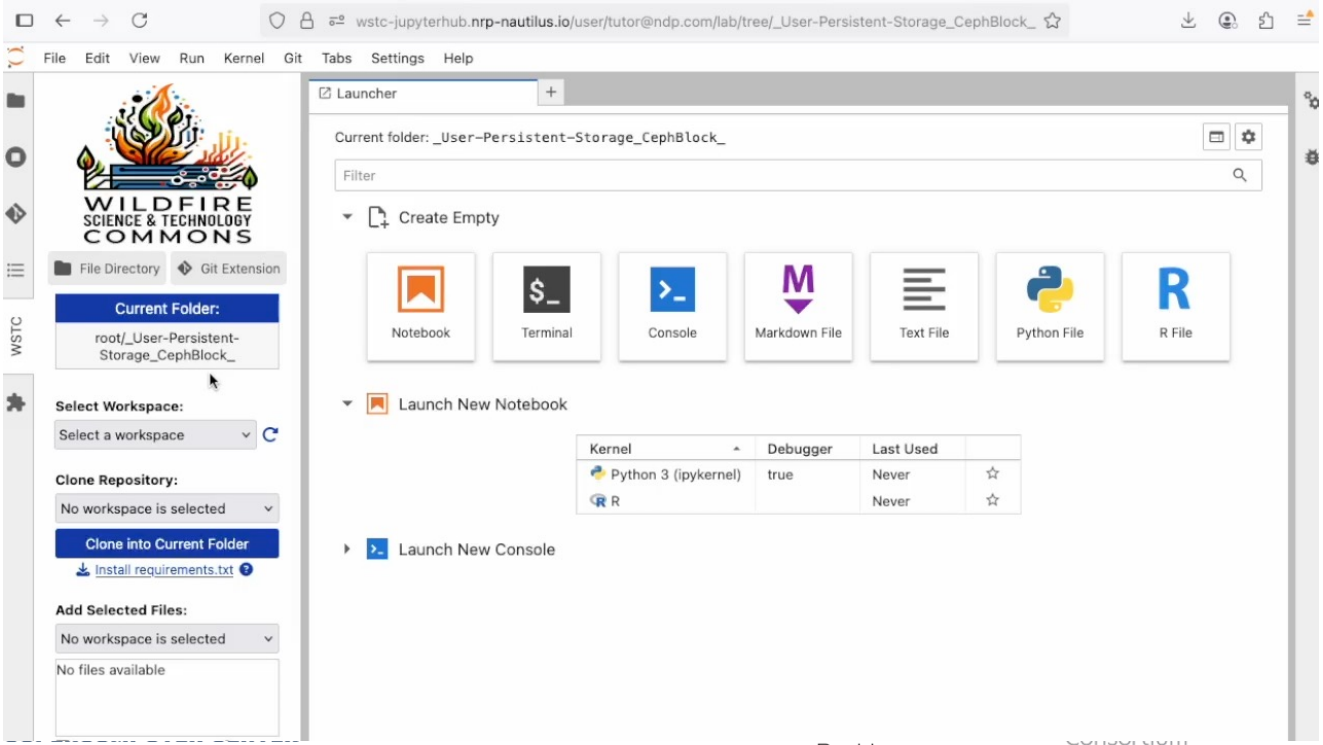
# Step 4:

## Run code on Jupyterhub

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM




Immersive Forest WIFIRE

IntELiMon

https://www.wildfirecommons.org

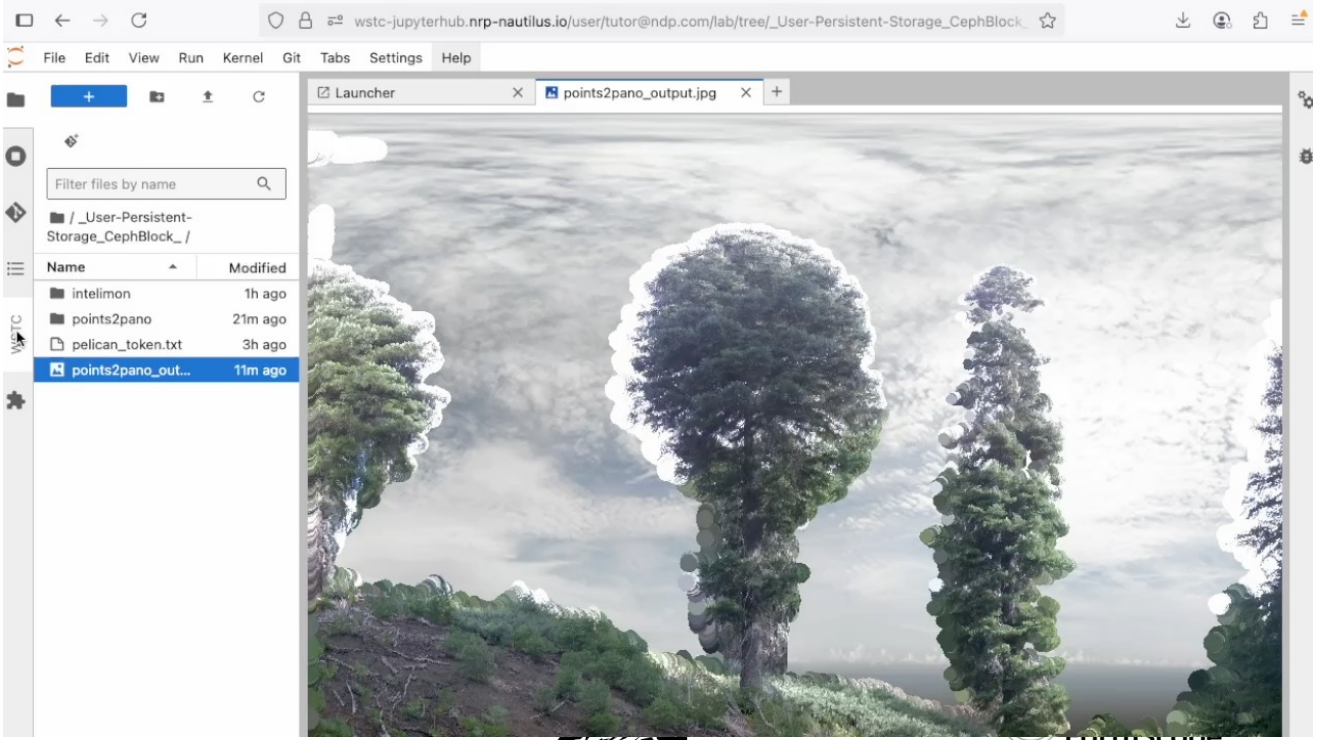
OF UTAH Boulder

# Step 5:
## Save outputs to cloud-compatible storage

Leticia Lee, PhD

**OBJECTIVE:** Process terrestrial LIDAR scans (TLS) from the USGS IntELiMon program to create visual representations of the scans to push them to data systems that integrate them into the Points2Pano app in WIFIRE's Immersive Forest.

NATIONAL DATA PLATFORM




Immersive Forest WIFIRE

IntELiMon

https://www.wildfirecommons.org

SUPERCOMPUTER CENTER

UNIVERSITY OF UTAH

www.sci.utah.edu

University of Colorado Boulder

EarthScope Consortium

# NDP and CICI Projects: What can you do on NDP?

NATIONAL DATA PLATFORM

DATA CATALOG INTEGRATION

AI AND SCIENTIFIC WORKFLOWS

CREATE CUSTOM SHELLS

NDP ENDPOINT DEPLOYMENT NEAR YOUR DATA

OFFER MATERIALS FOR CLASSROOM CAPABILITIES and CREATE DATA CHALLENGES


**İlkay ALTINTAŞ**
ialtintas@ucsd.edu

**Manish PARASHAR**
manish.parashar@utah.edu



NSF #2333609


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

NATIONAL DATA PLATFORM

# DEMO