

National Data Platform

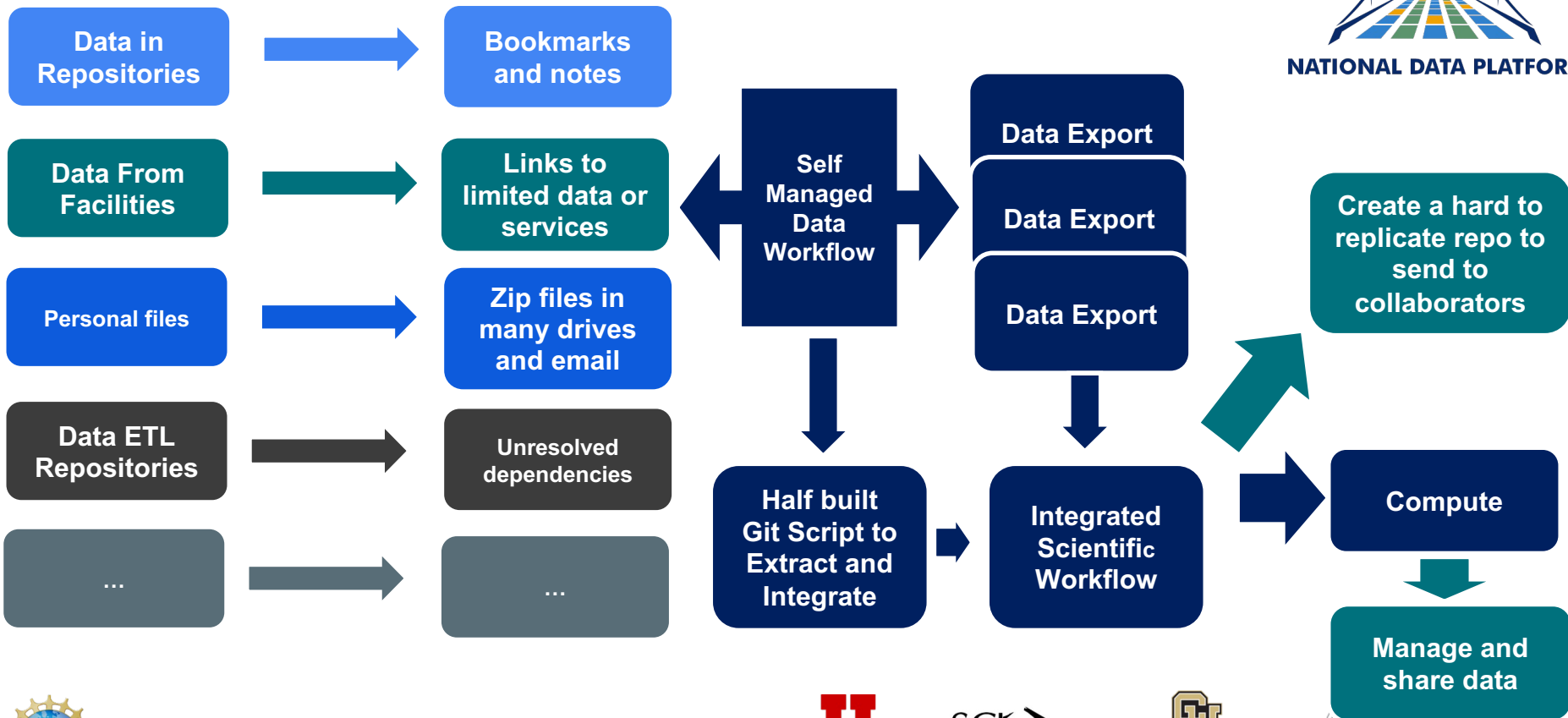
CICI National Scale Data Fabric Presentation

August 21, 2025

Working with data can be frustrating!



NATIONAL DATA PLATFORM



Why is data so challenging to use?

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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER

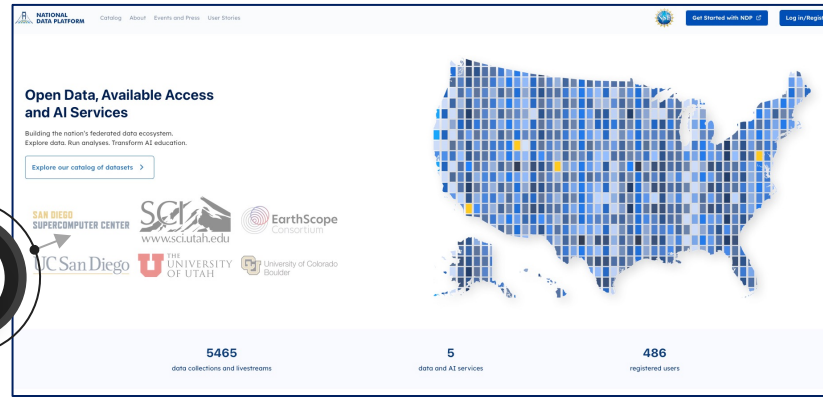
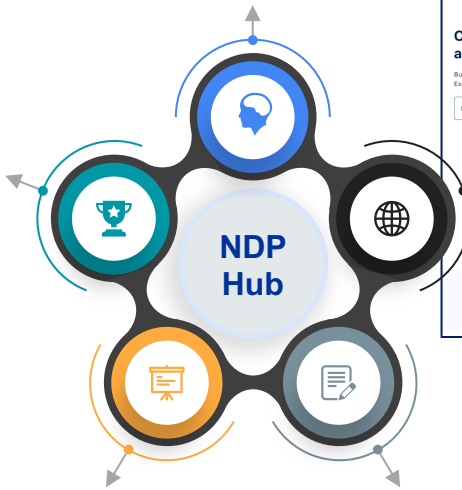


National Data Platform tackles the challenges around using and sharing data!

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.nationaldatapatform.org/>



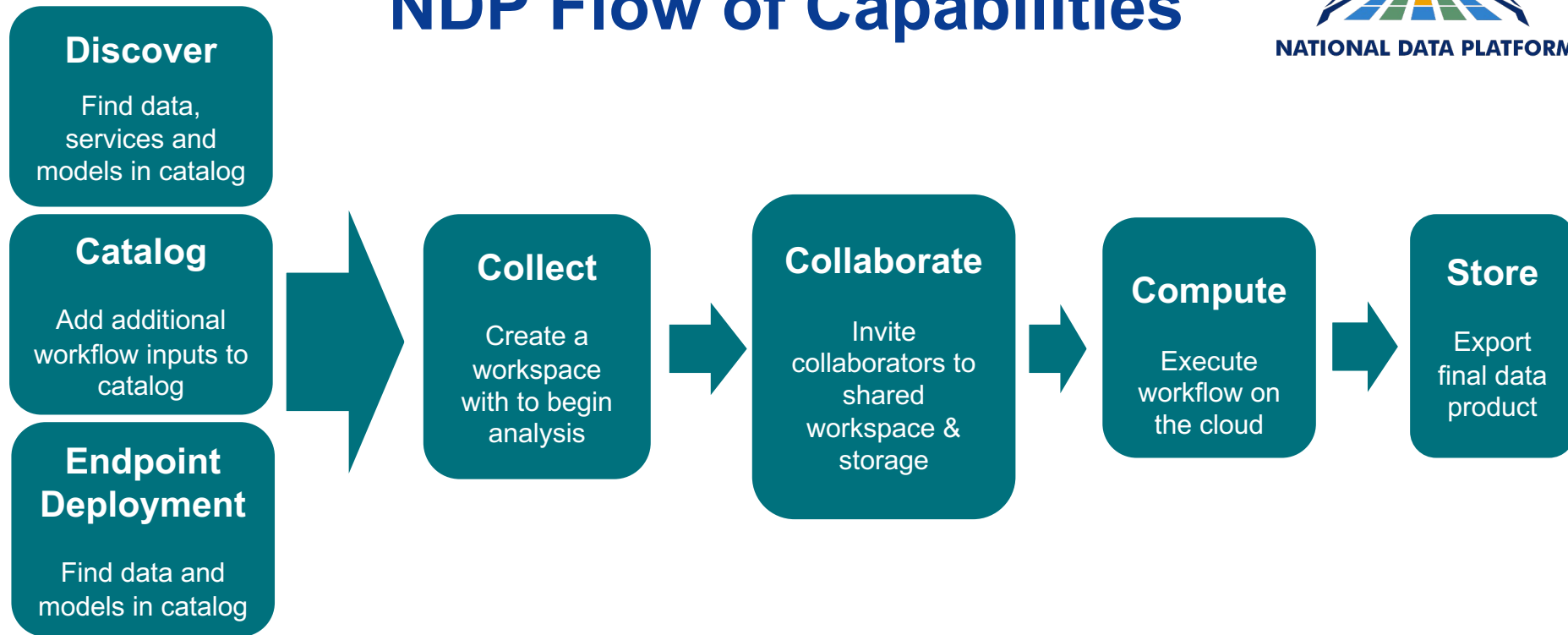
Centralized portal for discovery, access and use workspaces for research and education

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



NATIONAL DATA PLATFORM

NDP Flow of Capabilities



#2333609

UC San Diego

**SAN DIEGO
SUPERCOMPUTER CENTER**



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.



UC San Diego

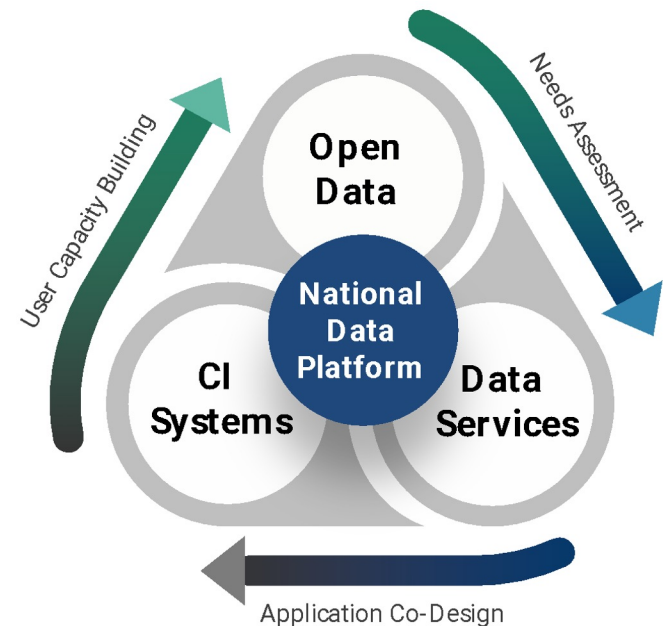
**SAN DIEGO
SUPERCOMPUTER CENTER**



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



#2333609

UC San Diego

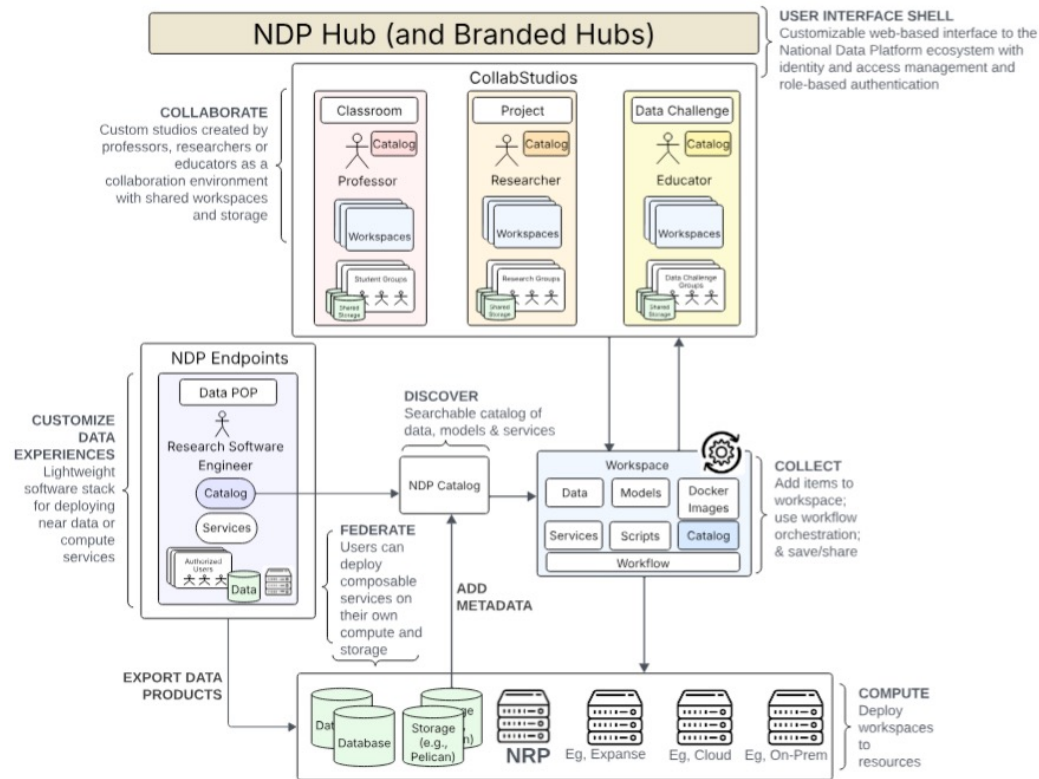
SAN DIEGO
SUPERCOMPUTER CENTER



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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



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



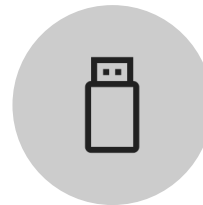
DATA CATALOG
INTEGRATION



AI AND SCIENTIFIC
WORKFLOWS



CREATE CUSTOM
SHELLS



NDP ENDPOINT
DEPLOYMENT NEAR
YOUR DATA



OFFER MATERIALS
FOR CLASSROOM
CAPABILITIES and
CREATE DATA
CHALLENGES

Ilkay ALTINTAŞ

ialtintas@ucsd.edu

Manish PARASHAR

manish.parashar@utah.edu



NATIONAL DATA PLATFORM

NDP Flow of Capabilities

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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



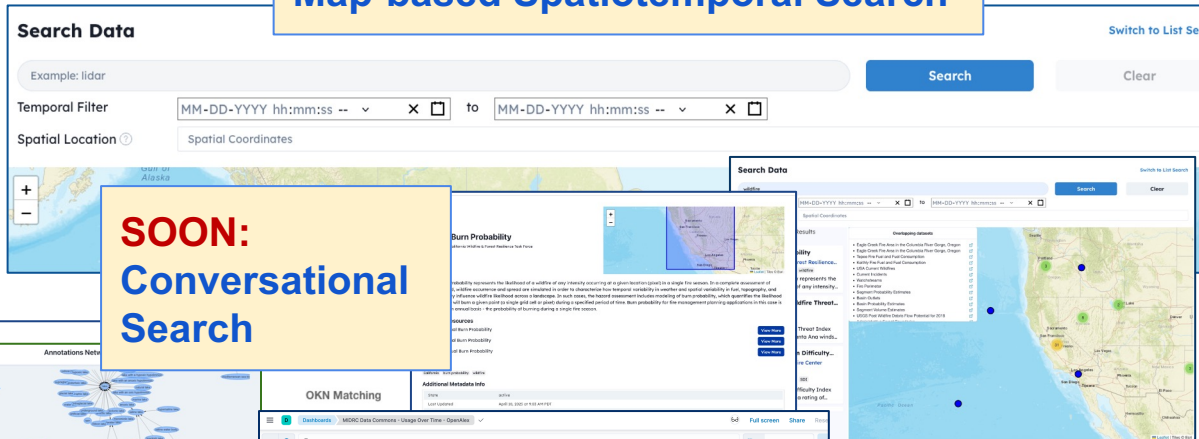
Data Discovery and Usage Exploration

Multimodal Search and Exploration Interfaces



NATIONAL DATA PLATFORM

Map-based Spatiotemporal Search



SOON: Conversational Search

Text Metadata Search

Conceptual Search

Search Data

Substring search Select Org...

Search catalog...

Search Clear

View Contributing Organizations

Search

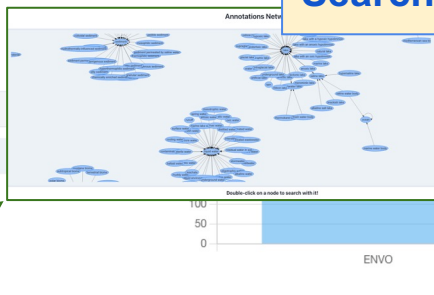
Conceptual search

lake

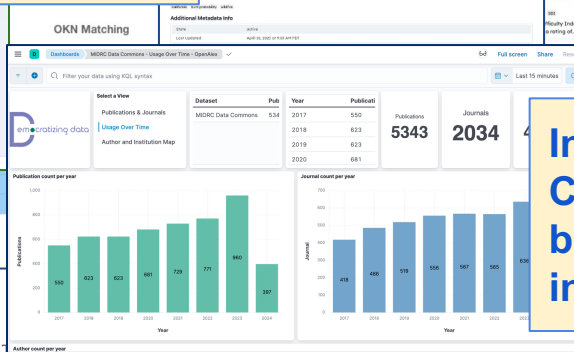
Search Clear

View Contributing Organizations

View Network Graph



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



NATIONAL DATA PLATFORM

NDP Flow of Capabilities

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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



Data Ingestion and Cataloging

Formal Data Ingestion into the NDP Catalog

My Uploads

+ Register dataset

You have not submitted any requests to add datasets yet!

General

General Overview

Contributors and Contact Info

Access and Usage

Metadata

General Metadata

Spatial Information

Temporal Information

Additional

Data

Spatial Information

Resolution of Spatial Data

Spatial Coverage

Bounding box of data

Temporal Information

Time Zone

Start Date and Time

End Date and Time

Data

Spatial Coverage Format

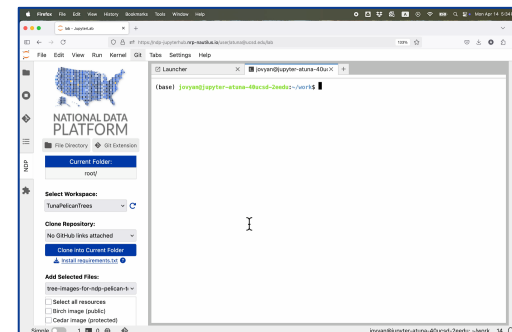
Bounding box Format

Resolution of Temporal Data

You have not added any data.



OSDF Integration



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



NSF HydroFrame and Hydrogen Digital Twin NAIRR Pilot

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



NATIONAL DATA PLATFORM

- 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

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

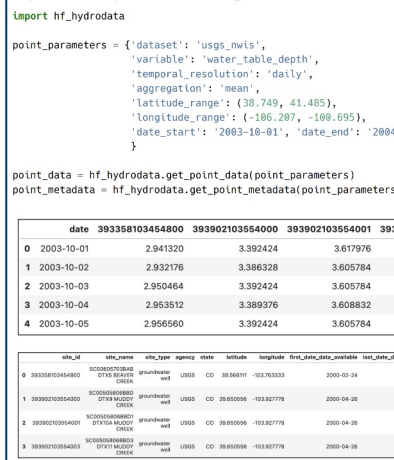


Figure 1: Image of example site-level point observations DataFrame and select site returned by the provided example function calls.

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

HydroFrame

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.

HydroGen
APPLICATIONS
Explore current and future watershed conditions across the US with HydroGen

HydroData
DATA AND SIMULATIONS
Access hydrologic datasets and model results

ParFlow CONUS1.0
DATA AND SIMULATIONS
Learn more about the first generation national ParFlow model

ParFlow CONUS2.0
DATA AND SIMULATIONS
Learn more about the second generation national ParFlow model

ParFlow Resources
MODELING TOOLS
Find training resources and tools for working with ParFlow models

SubsetTools
MODELING TOOLS
Build your own watershed model from the national ParFlow platform



#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



NATIONAL DATA PLATFORM

SPIn4D, NSF#2008344



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."

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/>

- 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



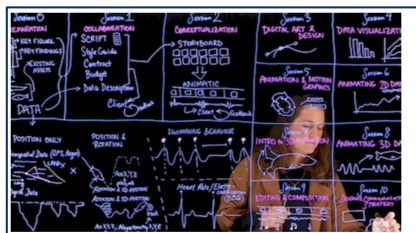
Biologging for Marine Animal Behavior and Physiology

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

- 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://lifeinthedeepest.nrp-nautilus.io/>

pyologger
a Python package for analyzing physiological, behavioral, & spatiotemporal biologging data

Schmidt Sciences

DiveDB
a database for high-resolution multimodal biologging data (multi-logger multi-sensor)

LIFE IN THE DEEP Web Explorer
a web tool for visualizing physiology and behavior alongside oceanographic variables

marinesomniac
a Python package for flexibly automating timeseries segmentation

Parquet data lake: Deployed on:

PostgreSQL, DuckDB, Nautilus (NATIONAL RESEARCH PLATFORM)

<https://lifeinthedeepest.nrp-nautilus.io/>

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



#2333609

UC San Diego

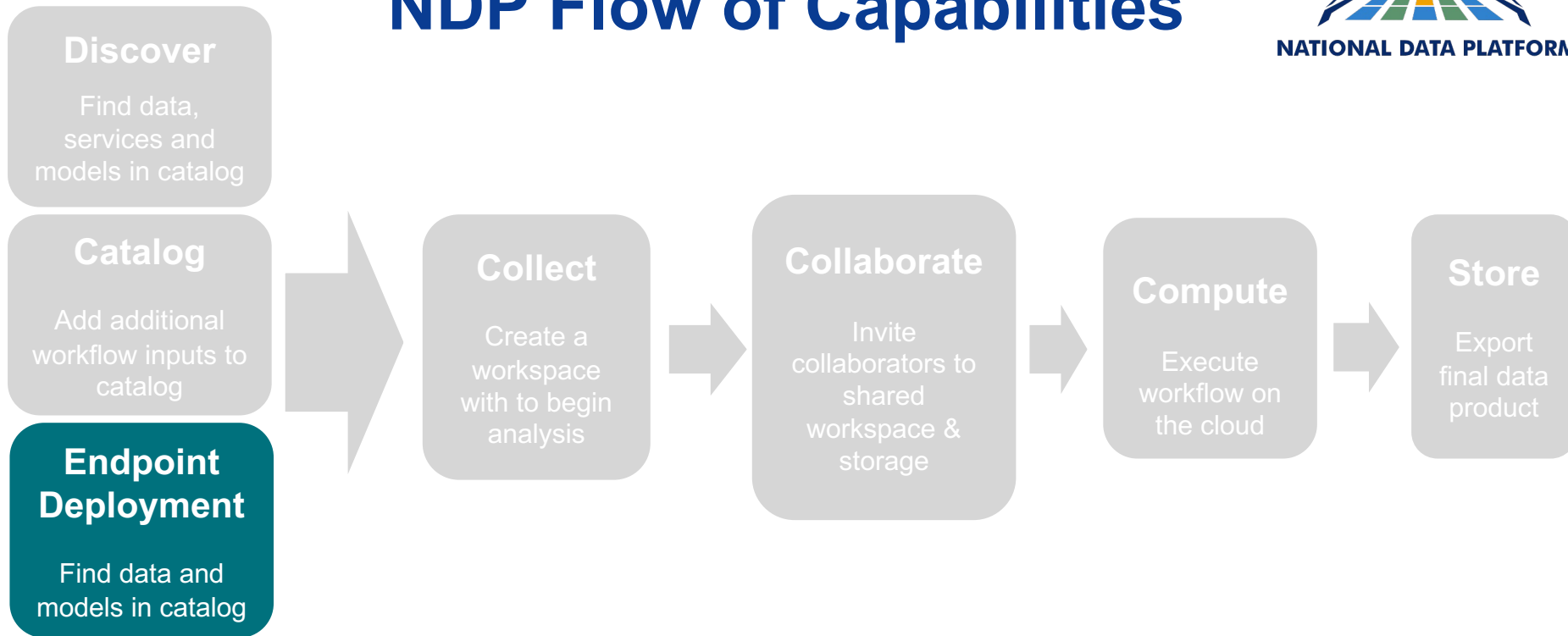
SAN DIEGO
SUPERCOMPUTER CENTER





NATIONAL DATA PLATFORM

NDP Flow of Capabilities



#2333609

UC San Diego

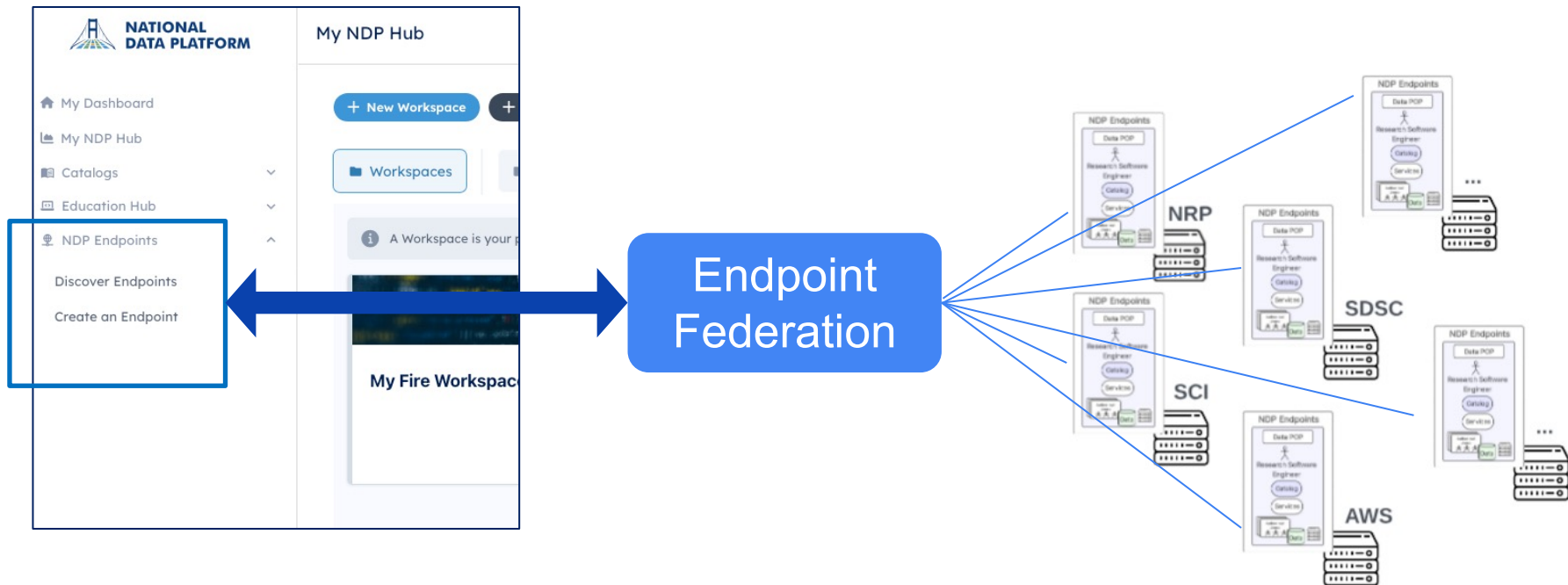
SAN DIEGO
SUPERCOMPUTER CENTER



Endpoint Deployment on Composable Resources and Federation



NATIONAL DATA PLATFORM



#2333609

UC San Diego

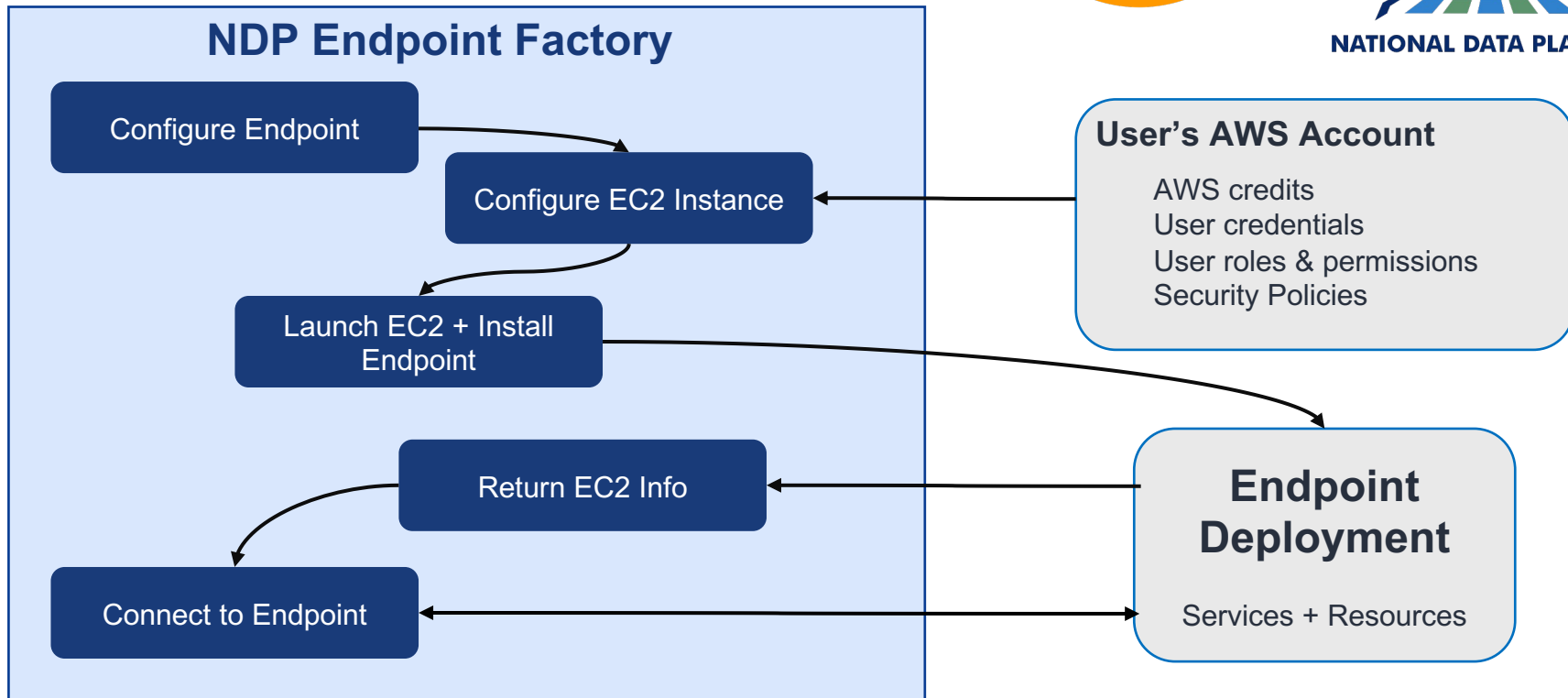
SAN DIEGO
SUPERCOMPUTER CENTER



Launching Endpoints on aws



NATIONAL DATA PLATFORM



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



EarthScope Data Streaming Workflows

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



NATIONAL DATA PLATFORM



Subscription

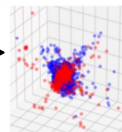


Historical data storage



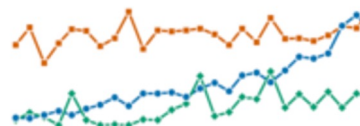
Real-time visualization of (derived) streams

AI Analytics



Trigger HTC workflows

AI model retraining



Register metadata, derived streams



- **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



UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER

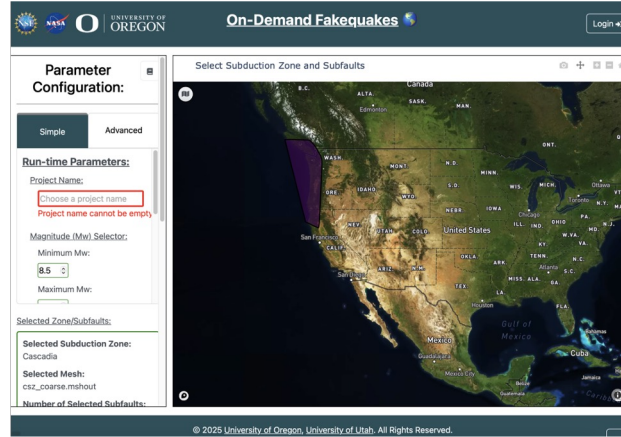


On Demand FakeQuakes (ODF)

Contact: Diego Melgar, University of Oregon



NATIONAL DATA PLATFORM



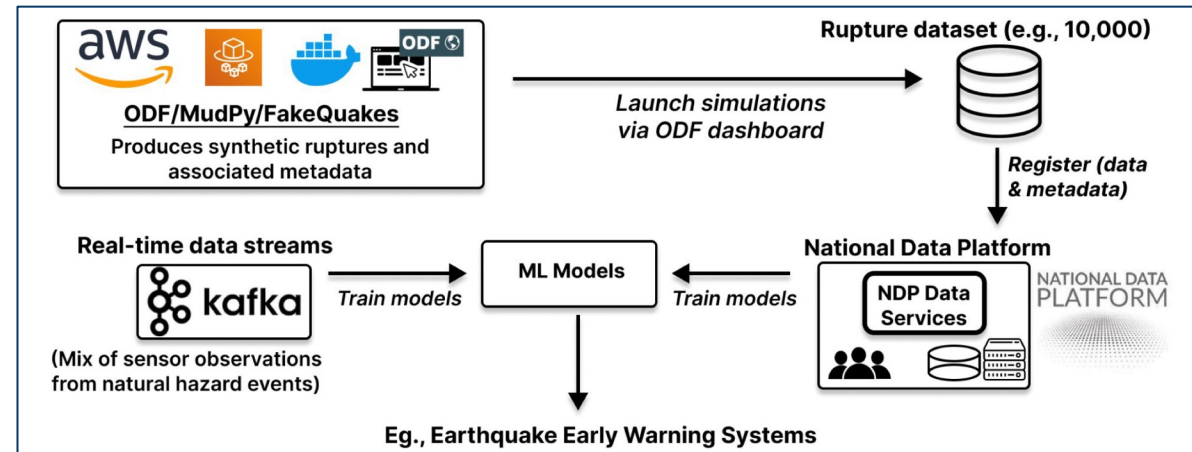
<https://ondemandfakequakes.org>



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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER

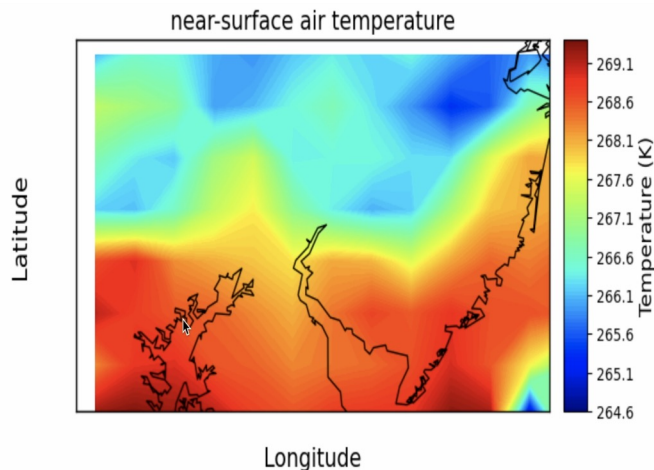
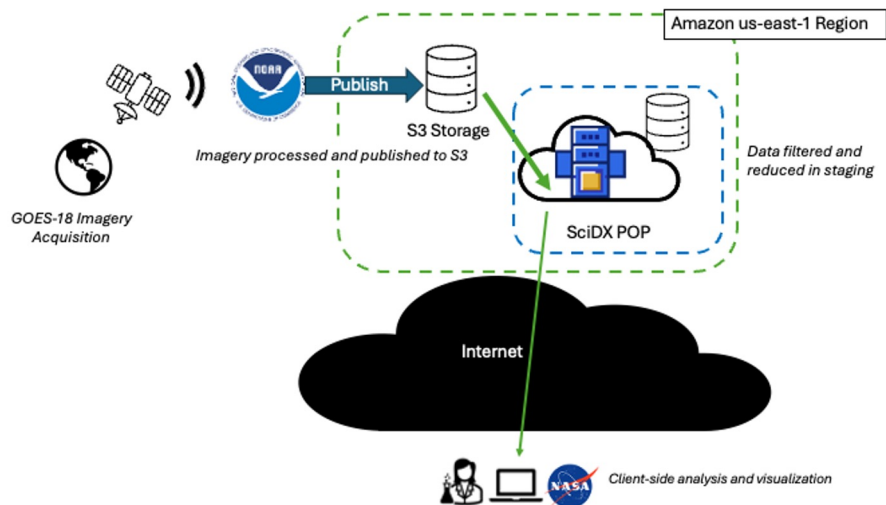


NASA/NOAA Satellite Data

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 Data Staging:

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



#2333609

UC San Diego

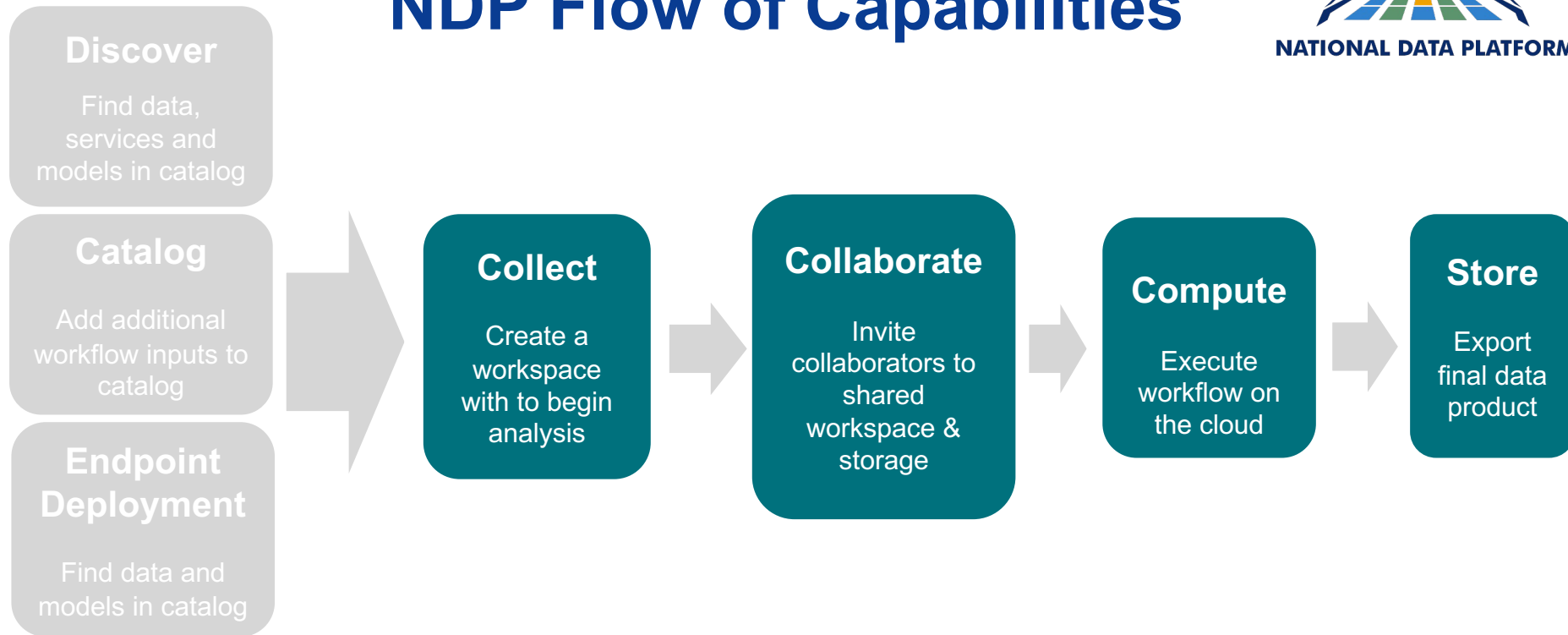
SAN DIEGO
SUPERCOMPUTER CENTER





NATIONAL DATA PLATFORM

NDP Flow of Capabilities



#2333609

UC San Diego

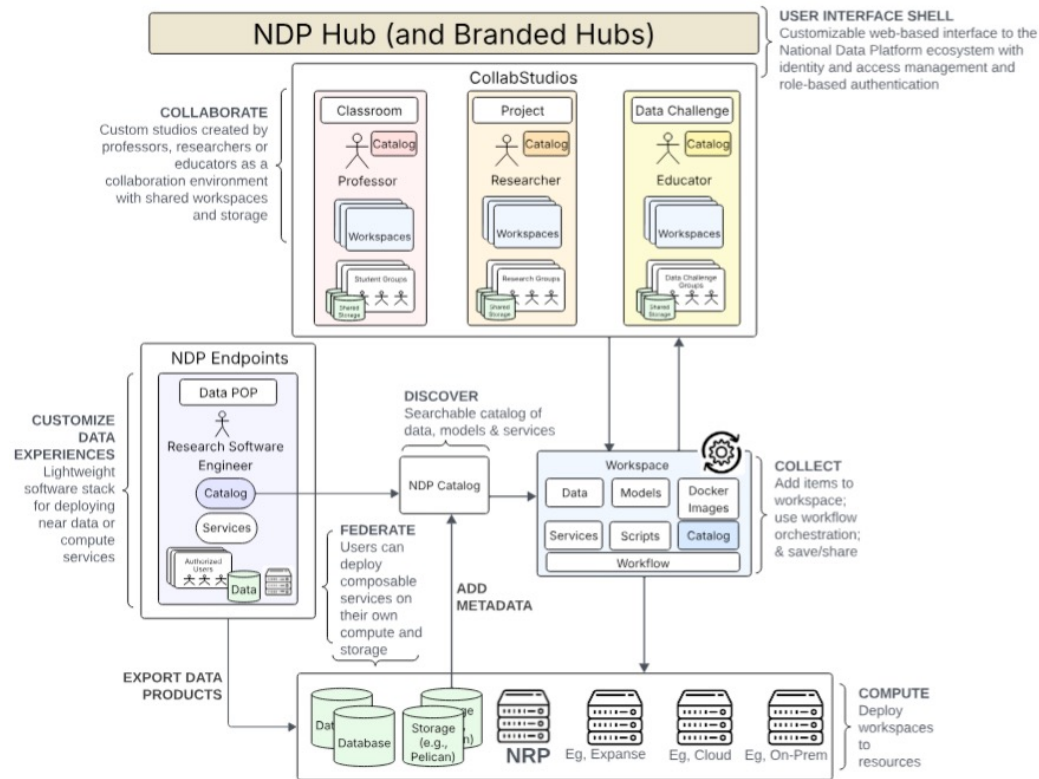
SAN DIEGO
SUPERCOMPUTER CENTER



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



#2333609

UC San Diego

**SAN DIEGO
SUPERCOMPUTER CENTER**



24

Hugging Face Data and Model Integration



NATIONAL DATA PLATFORM

NASA Earth Science AI Training Datasets

datasets

NASA Harmonized Landsat Sentinel-2 (HLS) Found

models

datasets

NASA HLS Burn Scars training dataset

datasets

NASA

AI training dataset for satellite image-based burn scar



NASA Multi-temporal crop classification training da

datasets

NASA Science Ethical AI Resources

Hugging Face Search models, datasets, and spaces

Datasets: **ibm-nasa-geospatial/hls-burn-scars**

Modalities: **Image** Languages: **English**

Dataset card Data Studio Files

Dataset Viewer

Split (3)
train - 540 rows

Search is not available for this dataset

The dataset viewer is not available for this dataset

Dataset Card for HLS Burn Scar Scenes

Dataset Summary

This dataset contains Harmonized Landsat and Sentinel-2 data for the contiguous United States. There are 804,512 rows in the dataset.

Dataset Structure

TIFF Metadata

Each tiff file contains a 512x512 pixel tiff file. Some files have already been converted to reflectance.

NATIONAL DATA PLATFORM v1.0.0

Release Notes

My Dashboard Catalog Education Hub NDP Endpoints About Events and Press User Stories

Open Data, Available Access and AI Services

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

[Explore our catalog of datasets >](#)

5374
data collections and livestreams

5
data and AI services

475
registered users



#2333609

UC San Diego

**SAN DIEGO
SUPERCOMPUTER CENTER**

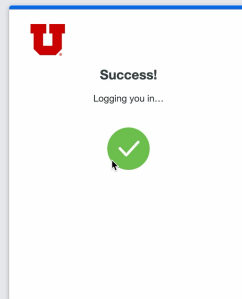


25

Users can select compute endpoints...



NATIONAL DATA PLATFORM



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



Workspace OSDF Pelican Integration



NATIONAL DATA PLATFORM

NATIONAL DATA PLATFORM

My Dashboard
My NDP Hub
Catalogs
Education Hub
NDP Endpoints

NDP / OSDF workspace to auto-caption an image

Programatically Accessible Resources

Datasets [Edit](#)

Source datasets

Tree images for NDP/Pelican testing (annotated)

Models / Services [Edit](#)

Source provided models

https://github.com/alexandertuna/ndp_osdf.git

Scripts [Edit](#)

No scripts!

[Go to JupyterHub](#)

QuickTime Player File Edit View Window Help

JupyterHub

ndp-jupyterhub.nrp-nautilus.io/hub/spawn/atuna@ucsd.edu

NATIONAL DATA PLATFORM Home Token Admin

atuna@ucsd.edu Logout

NDP JupyterHub Server Options

[Available resources page](#)

What do you want to work on?

Select

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:

Timeout (in seconds): once a server has been successfully spawned

1200

Favorites

Recents

Applications

Desktop

Documents

Downloads

good_screenshots

ndp_pelican_recording.mov

Screenshot 2...at 9:25:34 AM

Screenshot 2...at 10:11:13 AM

Cancel Open



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



27

Graduate Data Science Class Assignments for Knowledge Graphs

Professor: Amarnath Gupta, UC San Diego



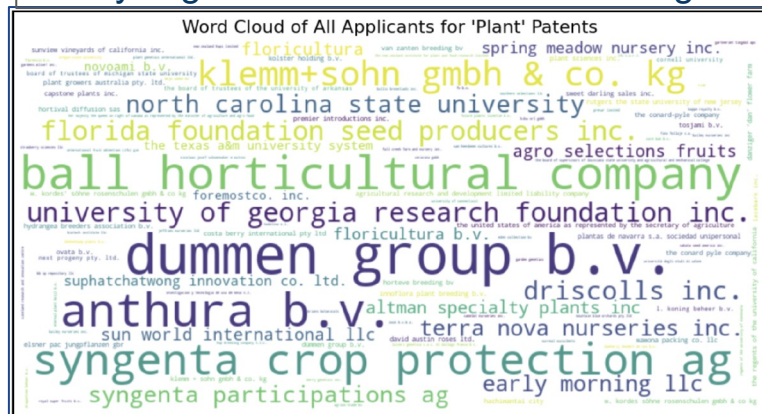
NATIONAL DATA PLATFORM

- 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



Example Student Project with NAIRR USPTO Dataset

Analyzing USPTO Patents with Ontologies



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



EarthScope
Consortium

28

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

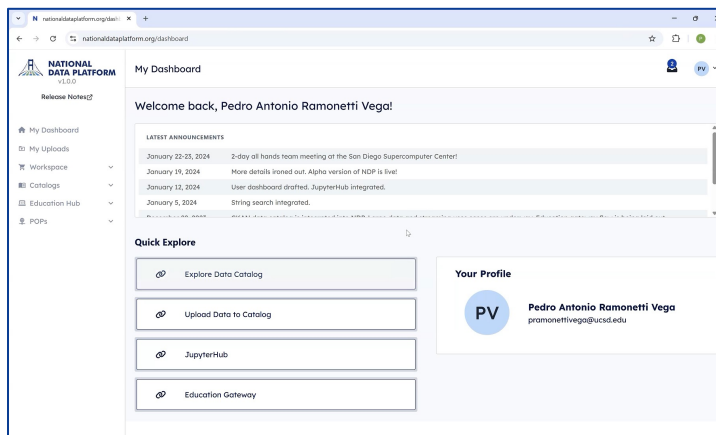


NATIONAL DATA PLATFORM

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**



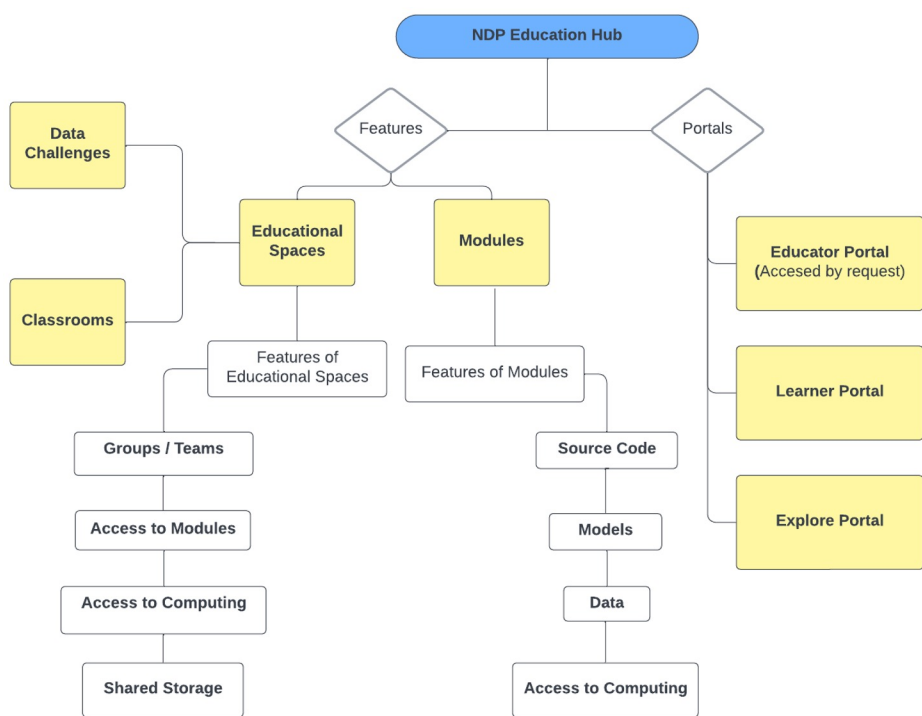
29

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.



#2333609

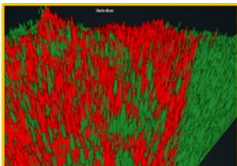
UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



BurnPro3D High-Resolution Vegetation and Fire Simulation Data

Contact: Leticia Lee, UC San Diego



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

**NATIONAL DATA PLATFORM**
v1.0.0

Release Notes

My Dashboard

Workspace

Catalogs

Education Hub

NDP Endpoints

Education Hub

Educator Portal

Learner Portal

Explore

Open Learning

Learner's Portal

My Classrooms

BP3D: BP3D Data Access
WINTER 2025 | UCSD

BP3D Data Access
UCSD

Description
This module is designed to demonstrate how to access QUIC-Fire runs ensemble Zarr files from NDP.

Instructions
1- Launch JupyterHub. You can do this, either by clicking on the JupyterHub button at the bottom of this module or by clicking on the JupyterHub button at the bottom of the module.
2- Reserve Resources. Once you're in JupyterHub, reserve with the following specifications (default):
Region: Any
Zone: Any
GPUs: 0
Cores: 1
RAM: 16GB
GPU Type: Any
Pre-built Image: Minimal NDP Starter JupyterLab
Architecture: amd64
3- Launch your server. Once you specify the resources as indicated, click on Start Server and wait to be redirected to the JupyterLab interface.

Explore the ODSF

The ODSF service connects the nation's scientific data repositories to computing infrastructure

Repositories available via the ODSF

41 repositories are connected to the ODSF to help deliver scientific data. The table below illustrates the datasets accessible via the ODSF.

Name	Organization	Field Of Science
JLab Simulation Datasets	Jefferson National Laboratory	Physics
Kennesaw State University CC* Storage	Kennesaw State University	Multi/Interdisciplinary Studies
MeerKAT Absorption Line Survey (MALS)	Inter-University Centre for Astronomy and Astrophysics	Physical Sciences
Alex Hurt public data# Human readable name of the namespace	University of Missouri	COMPUTER AND INFORMATION SERVICES AND SUPPORT SERVICES
Visualizations of the ECCO Project's 1/48° MITgcm Simulation (part 1)	National Science Data Fabric	Physical Sciences
Visualizations of the ECCO Project's 1/48° MITgcm Simulation (part 2)	National Science Data Fabric	Physical Sciences
IASA C1440-LLC2160, DYAMOND	National Science Data Fabric	Physical Sciences
NCAR Research Data Archive	NSF National Center for Atmospheric Research	Physical Sciences
BurnPro3D	University of California, San Diego	NATURAL RESOURCES AND CONSERVATION
WPAO VRS3	National Radio Astronomy Observatory	Astronomy and Astrophysics



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



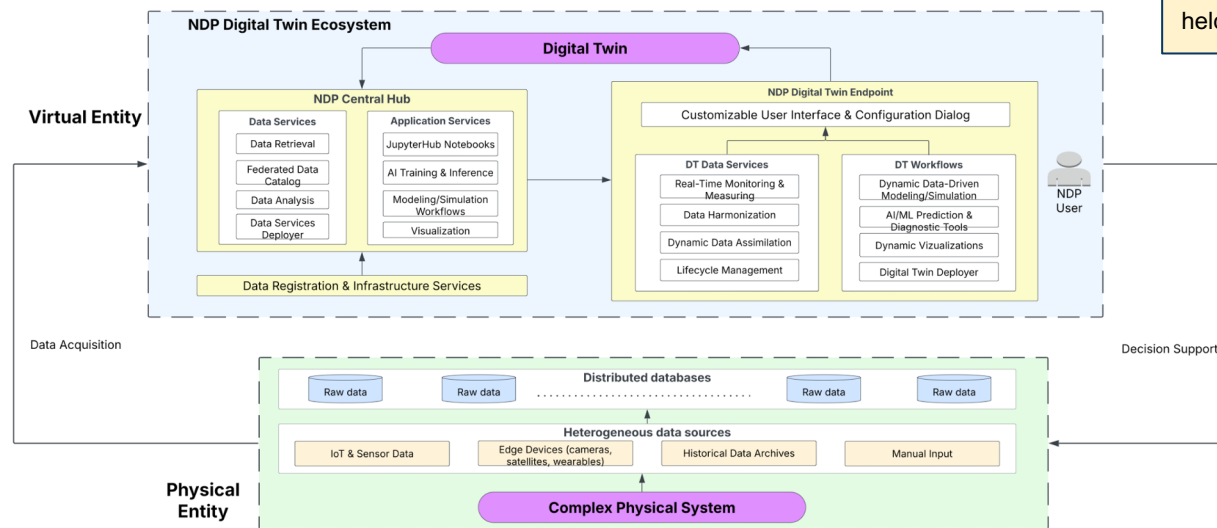
Using NDP for Digital Twin Applications

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 data discovery, integration and AI-workflows using OSDF and NRP.



#2333609

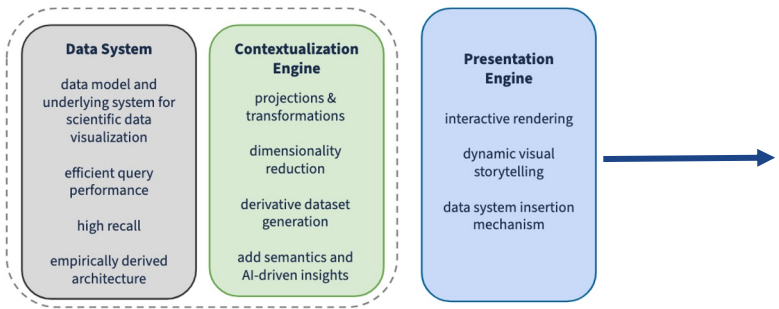
UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



Immersive Forest

Contact: Isaac Nealey, PhD Student, UC San Diego



- 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



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER



Wildfire Science and Technology Commons

Contact: Claire Stirm, UC San Diego



NIST

UC San Diego

SCIL

Societal Computing
and Innovation Lab

PROWESS

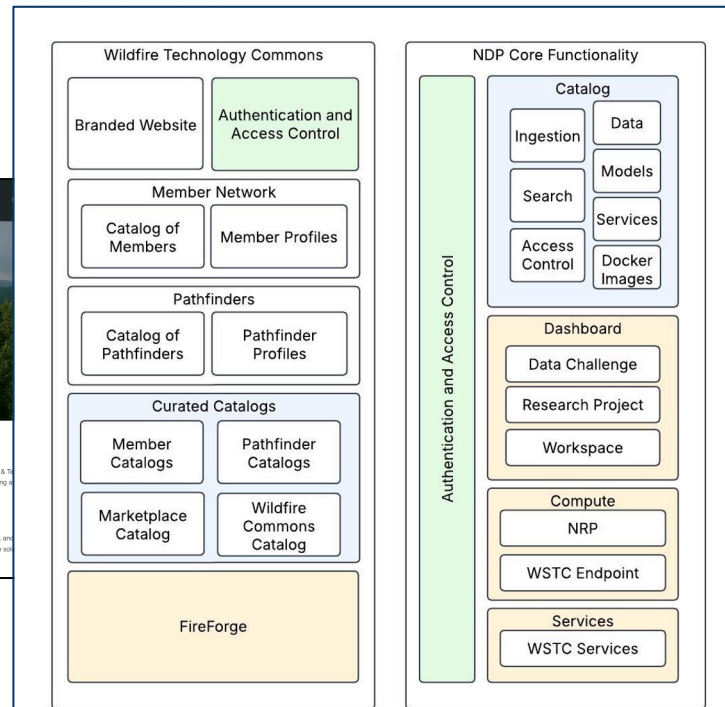
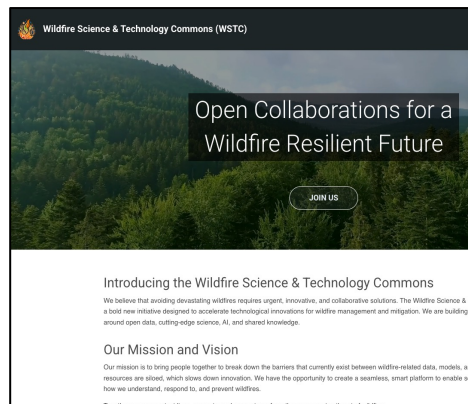
Proactive Wildfire & Environmental Sustainability Solutions



NATIONAL DATA PLATFORM

<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

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER

THE
UNIVERSITY
OF UTAH

SCIL
www.sci.utah.edu

CU
University of Colorado
Boulder

EarthScope
Consortium

35

Wildfire Science & Technology Commons



Expert
Network
Tool



FireForge
Platform



Community
Marketplace



FireForge Platform Interface

<https://www.wildfirecommons.org>

My Dashboard

My Hub

Catalogs

Education Hub

Endpoints

My Hub

+ New Workspace

+ New Curated Catalog

+ New CollabStudio

Workspaces

Curated Catalogs

Projects

Data Challenges

Classrooms

A Workspace is your private area to organize and work with data, models, and services. Think of it like a personal lab — you can attach specific datasets, build models, and experiment freely. It is not shared by default.

Claire's Workspace

Launch

Datasets

FIA Database - California

Observed Field Data for UC Climate Action Seed Project: Scaling Science-Driven Vegetation Treatments for a Wildfire Resilient California

Models

No models

Scripts

https://github.com/pramodnetivega/print_1_a.git

Go to JupyterHub

WSTC JupyterHub Server Options

Region

Zone

GPUs

Cores

RAM, GB

GPU type

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

Or Bring Your Own Image (JupyterLab Compatible)

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

Edit Datasets

Your Datasets

Search data catalog

1 - 3 of 179 Data Collections and Streams

California Wildfire & Landscape Resilience Interagency Treatments

Annual Biomass Data (2001-2021) - Aboveground Live Biomass 2021

Ignition Cause - 1992-2020 - Lightning Cause

SAN DIEGO
SUPERCOMPUTER CENTER

THE
UNIVERSITY
OF UTAH

Points2Pano app in WIFIRE's Immersive Forest

<https://burnpro3d.sdsc.edu/points2pano>

burnpro3d.sdsc.edu/points2pano/?plot=CALNU_0029&ts=20241110&m=Basalarea

☆ ↻ ↗ ⬇ 👤 ⋮

Pacific Union College 0029 2024-11-10 Basal Area: 140.3 sq ft/acre



Leaflet | Tiles © Esri — Source: Esri, I-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UPR-EGP, and the GIS User Community

Trees	
Basal Area	140.3 sq ft/acre
Mean Tree Height	24.1 m
Mean DBH	11.7 in
Number of Trees	19
Max Tree Height	31.6 m
Std Dev of Tree Heights	4.0 m
CBH	8.4 m
Shrubs	
Number of shrubs	19
Area of Shrubs	71.1 sq m
Estimated area of shrubs per acre	547.8 sq m
Mean Height of Shrubs	2.9 m

DBH



#2333609

UC San Diego

SAN DIEGO
SUPERCOMPUTER CENTER

THE
UNIVERSITY
OF UTAH

www.sci.utah.edu

University of Colorado
Boulder

EarthScope
Consortium

Example:

TLS Processing Workflow Collaboration



<https://burnpro3d.sdsc.edu/points2pano>

<https://www.wildfirecommons.org>

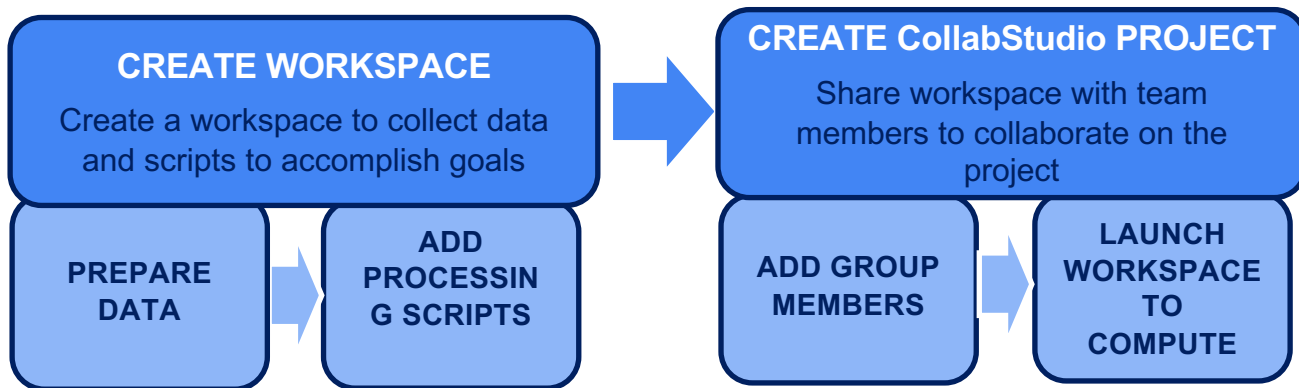


Leticia Lee, PhD

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



Isaac Nealey

**SAN DIEGO
SUPERCOMPUTER CENTER**

**THE
UNIVERSITY
OF UTAH**

SCI
www.sci.utah.edu

CU
University of Colorado
Boulder

**EarthScope
Consortium**

Step 1:

Prepare data



<https://www.wildfirecommons.org>

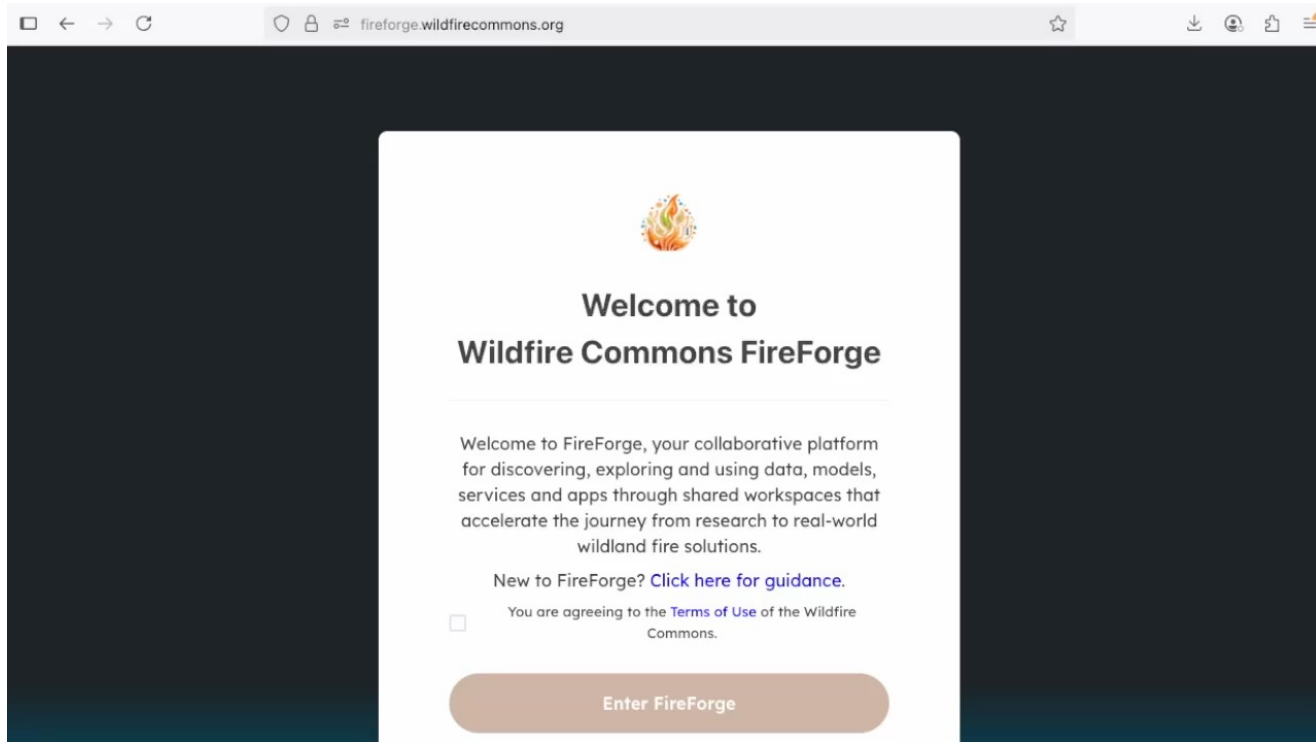


Leticia Lee, PhD

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



SUPERCOMPUTER CENTER

UNIVERSITY
OF UTAH

www.sci.utah.edu

University of Colorado
Boulder

Earthscope
Consortium

Step 2:

Create a workspace & project



<https://www.wildfirecommons.org>

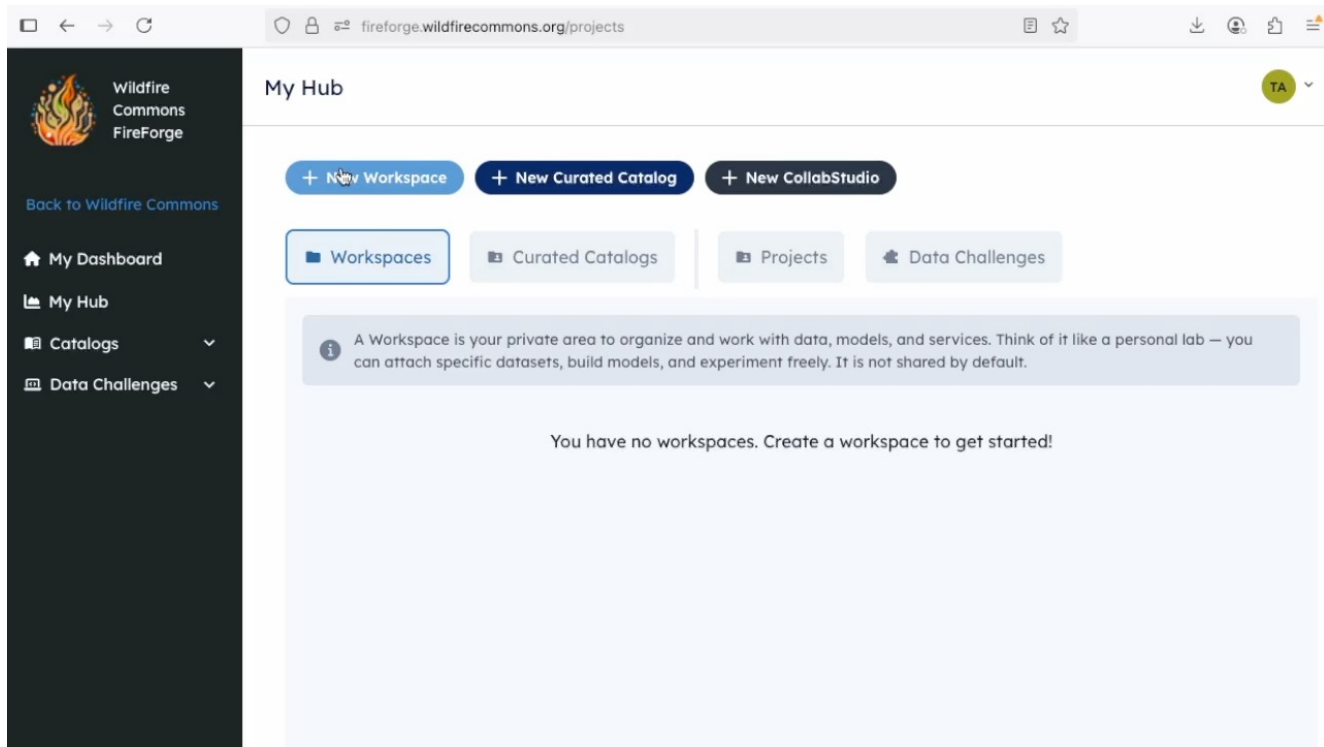


Leticia Lee, PhD

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



OF UTAH

Boulder

Step 3:

Jupyterhub



<https://www.wildfirecommons.org>



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

A screenshot of the Wildfire Commons FireForge web application. The browser address bar shows 'fireforge.wildfirecommons.org/projects'. The page has a dark sidebar on the left with navigation links: 'My Dashboard', 'My Hub', 'Catalogs', and 'Data Challenges'. The main content area is titled 'My Hub' and features three buttons: '+ New Workspace', '+ New Curated Catalog', and '+ New CollabStudio'. Below these are four tabs: 'Workspaces', 'Curated Catalogs', 'Projects' (which is active), and 'Data Challenges'. A notification banner states 'Your Projects contain groups working on shared workspaces.' Below this, a project card is visible for 'TLS Processing for Independence Lake' with an 'Open' button and a right arrow.

SUPERCOMPUTER CENTER

UNIVERSITY
OF UTAH

www.sci.utah.edu

University of Colorado
Boulder

Earthscope
Consortium

Step 4:

Run code on Jupyterhub



<https://www.wildfirecommons.org>



Leticia Lee, PhD

OBJECTIVE: Process terrestrial LIDAR scans (TLS) from the USCS 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

A screenshot of the Jupyterhub web interface. The browser address bar shows the URL: wstc-jupyterhub.nrp-nautilus.io/user/tutor@ndp.com/lab/tree/_User-Persistent-Storage_CephBlock_. The interface includes a sidebar with the Wildfire Science & Technology Commons logo and navigation options like File Directory and Git Extension. The main area is titled "Launcher" and shows the current folder as "_User-Persistent-Storage_CephBlock_". It offers options to "Create Empty" (Notebook, Terminal, Console, Markdown File, Text File, Python File, R File) or "Launch New Notebook". A table lists available kernels: Python 3 (ipykernel) and R. The "Launch New Console" option is also visible.

Kernel	Debugger	Last Used	
Python 3 (ipykernel)	true	Never	☆
R		Never	☆

OF UTAH

Boulder

© 2023 WIFIRE

Step 5: Save outputs to cloud- compatible storage



<https://www.wildfirecommons.org>

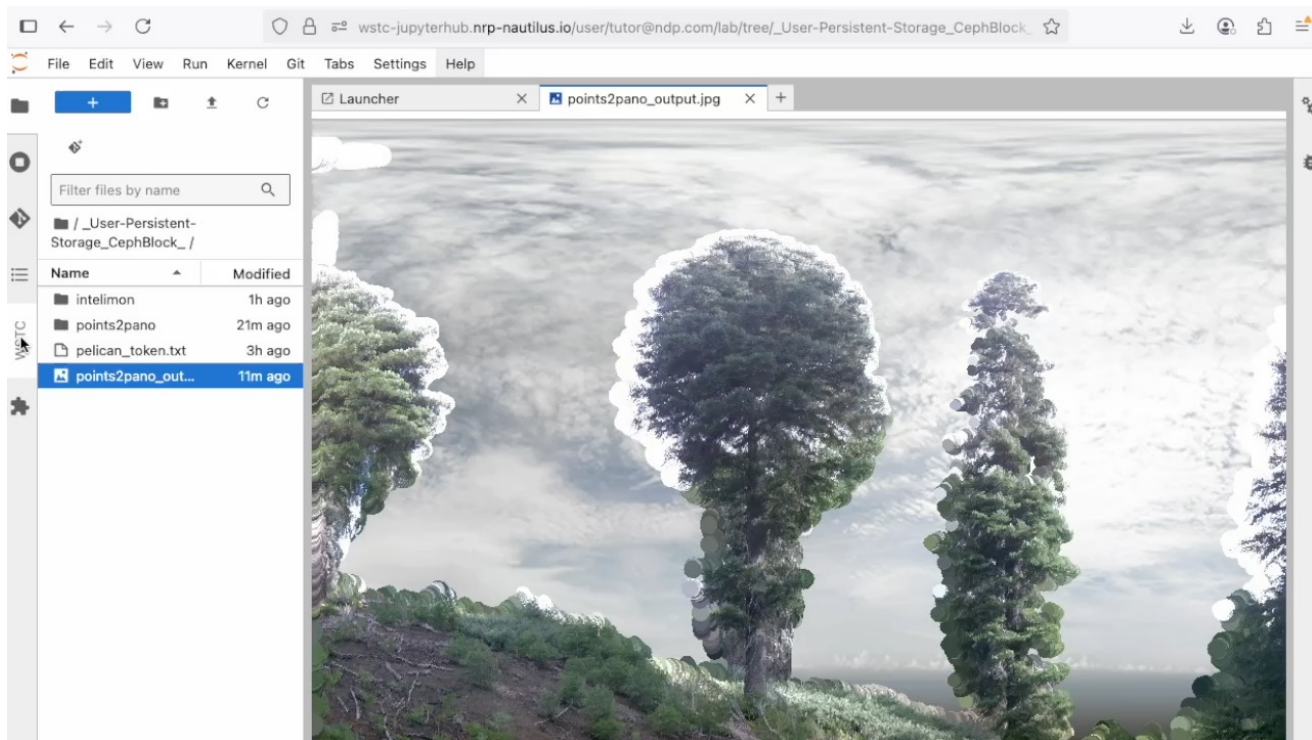


Leticia Lee, PhD

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



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?



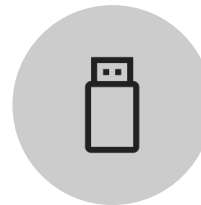
DATA CATALOG
INTEGRATION



AI AND SCIENTIFIC
WORKFLOWS



CREATE CUSTOM
SHELLS



NDP ENDPOINT
DEPLOYMENT NEAR
YOUR DATA



OFFER MATERIALS
FOR CLASSROOM
CAPABILITIES and
CREATE DATA
CHALLENGES

Ilkay ALTINTAŞ

ialtintas@ucsd.edu

Manish PARASHAR

manish.parashar@utah.edu

DEMO