

Example SDSC AI Services And Applications

Ilkay Altintas*

Chief Data Science Officer
San Diego Supercomputer Center

Frank Wuerthwein

Director
San Diego Supercomputer Center

*Presenter: ialtintas@ucsd.edu

Mental Picture of 3 types of Services

- Compute, Storage, Jupyter, Matrix chat, ticketing service, ...
 - Standard set of HPC services
- LLM as a Service via LiteLLM and LibreChat UX => **National Research Platform**
 - We offer any LLM that is available via Hugging Face
 - At any time we instantiate 6-10 different AI models as services
 - We can integrate ChatGPT access via LiteLLM to facilitate comparisons between different LLMs
 - Scalability Goal: Support Million++ students, researchers & educators on LLM services
- AI workflows as a service => **National Data Platform**
 - Imagine curated data, curated AI agents, curated RAG workflows, ... all available via a simple intuitive interface to compose custom AI services that fundamentally require workflow execution to instantiate the service.

Ready to learn more about the NRP?

Join our [National Research Platform Fall Training](#) between **September 2nd to September 30th**.

For more information [Click Here](#)

The National Research Platform (NRP): Accelerating Research and Education

The NRP is a community-owned research and education platform connecting researchers and educators to foster collaboration, accelerate innovation, and share resources. Supported by over 50 institutions, including leadership from UC San Diego, the University of Nebraska-Lincoln, and the Massachusetts Green High Performance Computing Center, the NRP provides access to cutting-edge technologies in AI, high-performance computing, data storage, and networking. Open to all nonprofit higher education institutions, from community colleges to top research universities, the NRP advances learning and scientific breakthroughs with support from the U.S. National Science Foundation, Department of Energy, and Department of Defense among others.

[Learn more](#)

- Start Here
- User Guide
 - Start
 - Tutorials
 - Running
 - Jupyter
 - Coder
 - AI/LLMs
 - NRP Managed LLM**
 - LLM as a Service
 - LLM in JupyterHub
 - Cloud AI 100 Cards
 - Networks
 - FPGA
 - Storage
 - Development
- Admin guide

NRP-Managed LLMs

The NRP provides several hosted open-weights LLM for either API access, or use with our hosted chat interfaces.

Chat with NRP LLMs →

Use the LibreChat interface to chat with NRP hosted LLMs

Get an API token for NRP LLMs →

Get an API token to programatically interact with the LLMs or use the LLMs in other apps

Chat Interfaces

If you are looking to chat with an LLM model similar to the interface provided by [ChatGPT](#), we provide [LibreChat](#), based on the [LibreChat](#) project. This is a simple chat interface for all of the NRP hosted models. You can use it to chat with the models, or to test out the models.

[Visit the LibreChat interface](#)

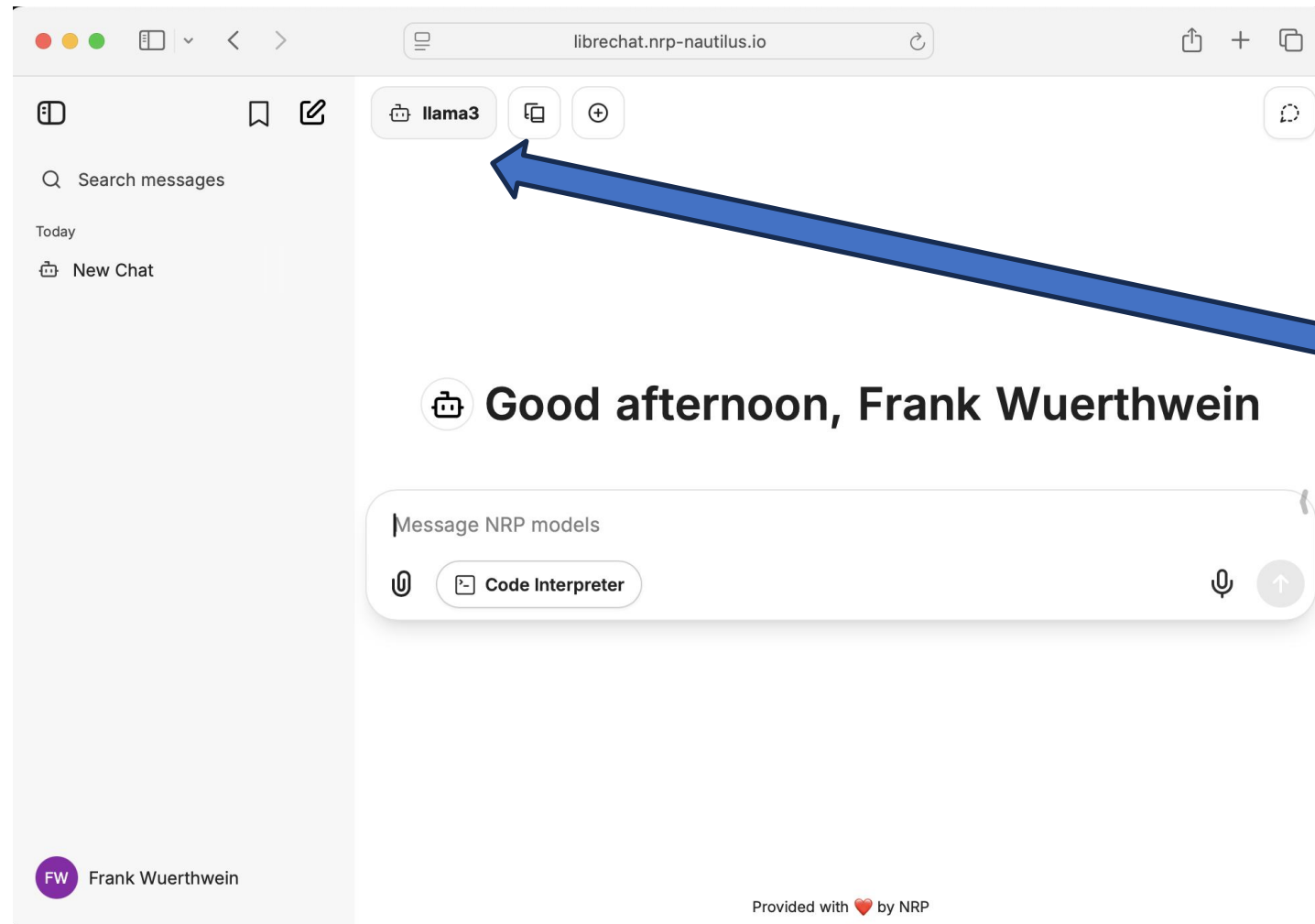


On this page

- [Overview](#)
- [Chat Interfaces](#)
- [API Access to LLMs](#)
- [Example Python Code](#)
- [Available Models](#)

Click here

Then login via your institutional login



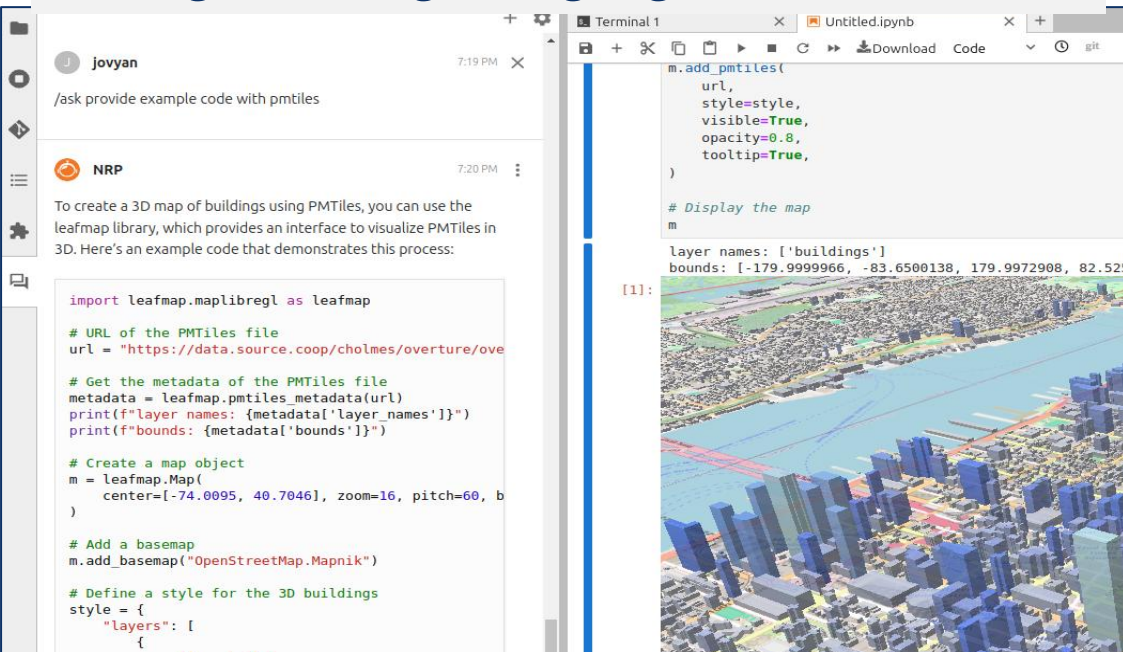
You can pick different models, and explore.

E.g. a fun thing to do is ask deepSeek about Tiananmen Square, and compare what you get with llama3.

Example for complex classroom use

Students learn how to visualize a 3D map of buildings in NYC. They start with a vanilla LLM, that has no clue how to do this. Then teach that LLM about PMTiles by ingesting its documentation.

Coding with Large Language Models (LLM)



The screenshot shows a Jupyter Notebook interface with a chat window on the left and a code editor on the right. The chat window shows a user asking for example code with PMTiles, and the LLM (NRP) responding with a detailed explanation and code. The code editor shows the code being generated, which includes importing the leafmap library, setting the URL of the PMTiles file, getting the metadata, creating a map object, adding a basemap, and defining a style for the 3D buildings. The output of the code is a 3D map of a city area, showing buildings and a river.

```
import leafmap.maplibregl as leafmap

# URL of the PMTiles file
url = "https://data.source.coop/choimes/overture/ove

# Get the metadata of the PMTiles file
metadata = leafmap.pmtiles_metadata(url)
print(f"layer names: {metadata['layer_names']}")
print(f"bounds: {metadata['bounds']}")

# Create a map object
m = leafmap.Map(
    center=[-74.0095, 40.7046], zoom=16, pitch=60, b

# Add a basemap
m.add_basemap("OpenStreetMap.Mapnik")

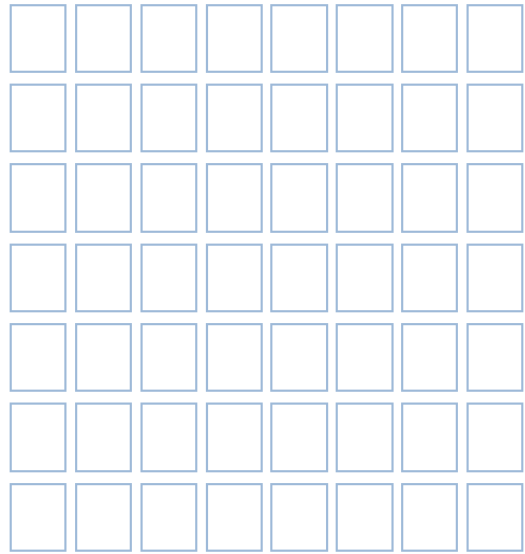
# Define a style for the 3D buildings
style = {
    "layers": [
        {
```



Classroom of **Carl Boettiger**
University of California, Berkeley

Voila, the LLM has learned how to provide working code for Jupyter@NRP

- 122 students
- Active learning classroom



NATIONAL DATA PLATFORM



What is NDP?

AI needs data to learn and make decisions, and building AI requires massive computing power.

The National Data Platform (NDP) lowers those entry barriers by bridging the gaps that arise when researchers, educators, and students attempt to find and use AI-ready data to solve problems. To achieve this, NDP fosters an online data ecosystem that promotes collaboration and innovation atop existing investments in research computing. NDP broadens access to and understanding of essential data and compute building blocks--enhancing science and AI in the process.

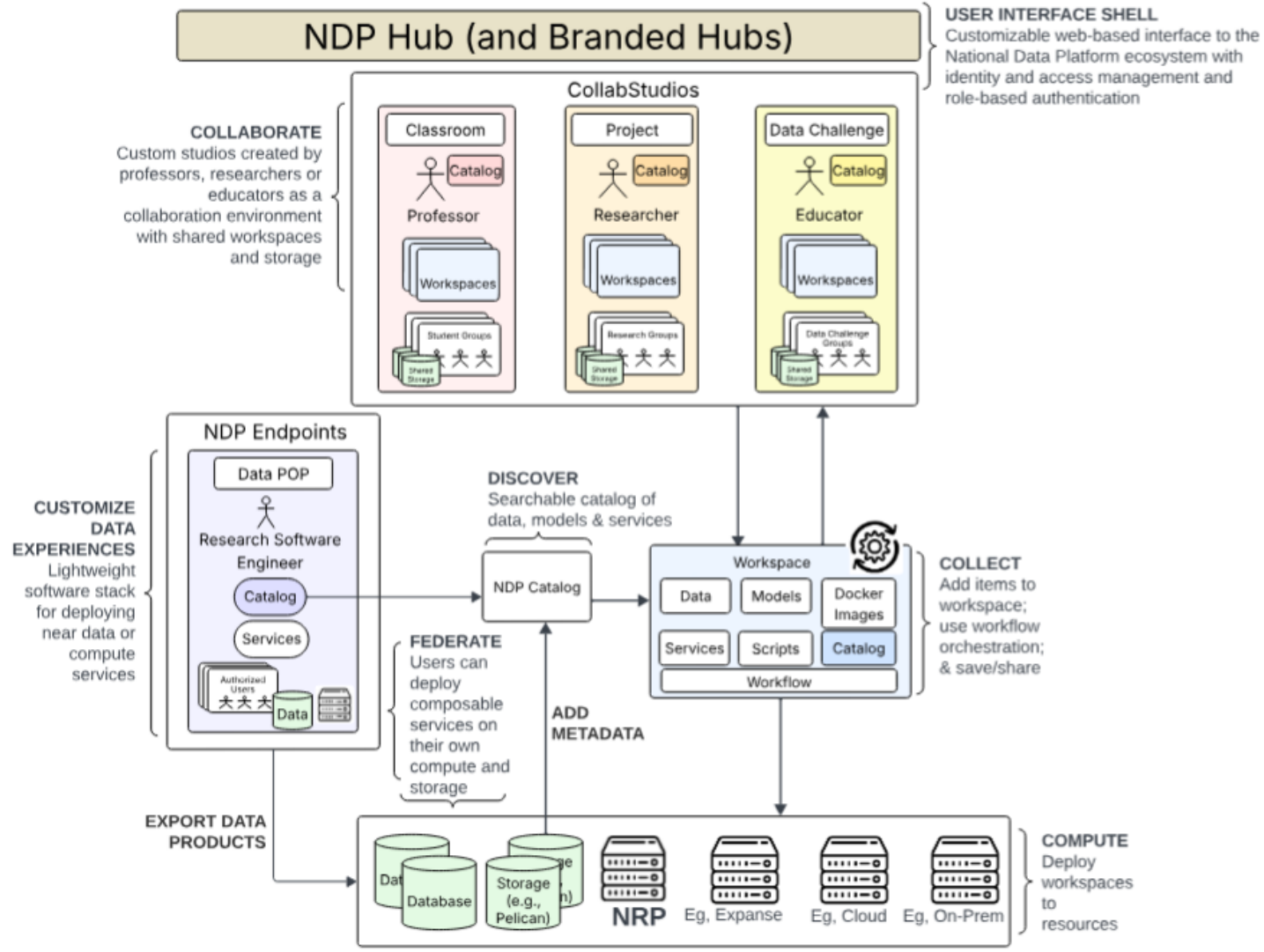
The CENIC community can use NDP in the following ways:

- Data providers can make their data discoverable and accessible
- Researchers can seamlessly create and integrate data analysis and AI into their work
- Educators can find and create educational resources for AI
- Students can learn AI skills through engaging data projects

NDP offers data coordination for diverse, distributed datasets and services for users to engage in research, learning, and discovery.



Ilkay Altintas (PI)
NSF Award #2333609



NATIONAL DATA PLATFORM



National Data Platform: An AI workflow engine for education

- NDP offers a simple and user-friendly format for CollaborationStudios where:
 - Educators set up Classrooms and/or Data Challenges
- CollaborationStudios can be stocked with Workspaces that include:
 - Relevant data, models, services, docker images, source code
 - Baseline solutions to orient students to AI workflows
- Students form groups where they can collaborate seamlessly using JupyterHub or Rstudio, or work on more complex AI workflows, all leveraging NRP's CPUs, GPUs, and FPGAs
- Each student group receives their own Collaborative Workspaces where they can customize workflows by adding additional data, models (e.g. any LLM that is available via Hugging Face), and repositories as needed
- Each group is provided with shared storage on NRP
- Each student is allocated personal persistent storage



Example Data Challenge with Solutions Available Today on NDP



Next Generation Science for Wildfire Resilience

- Next-generation 3D vegetation and fire models can enable the new tools fire managers need for real fuel and fire management solutions.
- Machine learning methods for site-specific prediction of critical tree metrics contribute to the improvement of 3D physics-based fire models that forecast fire behavior and effects with greater accuracy.

Challenge Problem Statement

- Collecting field data as model inputs for ecosystem monitoring can be very time intensive and cost prohibitive.
- New approaches that leverage sensed data can improve the speed and scalability of gathering crucial information for modeling wildfires and wildfire resilience.



NATIONAL DATA PLATFORM

Your Task

- For a series of given sites across California, use a set of representative lidar scans and additional data (85 GB total) to determine the distribution of trees by plant functional type, genus, and species



Release Notes

- My Dashboard
- Catalogs
- Education Hub
- NDP Endpoints

My Dashboard

[Go Back](#)

Example Data Challenge and Onboarding

University of California, San Diego

This is an open data challenge focused on advancing the prediction of tree species, height, location, and other critical forest metrics. The goal is to improve the accuracy of 3D physics-based fire models in simulating fire behavior and its ecological effects.

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

Deadline: December 30, 2099

27201 DAYS LEFT

Team Details

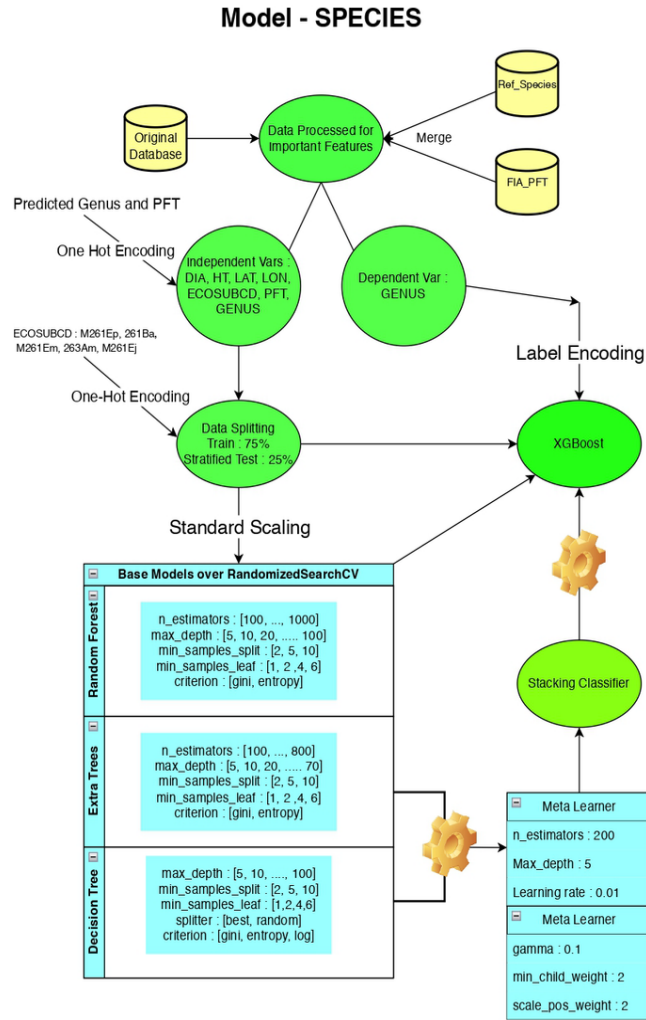
Melissa [Edit](#)

mfloca@ucsd.edu

Modules

<p>Onboarding Module</p> <p>University of California, San Diego</p> <p>In this module, you will complete a simple Machine Learning training task using daily weather data from San Diego over a three-year...</p> <p>View JupyterHub</p>	<p>Sprint 1 - Module A</p> <p>University of California, San Diego</p> <p>In this module, you will work with two datasets: the FIA Database and Observer Field Data from the UC Climate Action Seed Project. To help...</p> <p>View JupyterHub</p>	<p>Sprint 1 - Module B</p> <p>University of California, San Diego</p> <p>In this module, you will work with ALS and TLS data. To help you get started, an introductory notebook is provided: <i>lidar_data.ipynb</i>...</p> <p>View JupyterHub</p>	<p>Sprint 2 - Module A</p> <p>University of California, San Diego</p> <p>In this module, you will work with the ALS dataset again. One introductory notebook is provided: <i>tops-from-als.ipynb</i>. Additionally, a bas...</p> <p>View JupyterHub</p>	<p>Sprint 2 - Module B</p> <p>University of California, San Diego</p> <p>In this module, you will learn about FastFuels. Two introductory notebooks are provided: <i>fastfuels-api.ipynb</i> and <i>fastfuels-...</i></p> <p>View JupyterHub</p>	<p>Sprint 3</p> <p>University of California, San Diego</p> <p>In this module, you will learn about relevant metrics derived from TLS. You will also start working on an approach to predict tree metrics using ML. Two introductory noteboo...</p> <p>View JupyterHub</p>	<p>Sprint 4</p> <p>University of California, San Diego</p> <p>Throughout this challenge, you've explored a diverse set of data sources and tools that are critical for understanding forest structure to predict fire...</p> <p>View JupyterHub</p>
<p>Sample Solution 1 - ...</p> <p>UCSD</p> <p>This module contains a</p>	<p>Sample Solution 2 - ...</p> <p>UCSD</p> <p>This module contains a</p>	<p>Sample Solution 3 - ...</p> <p>UCSD</p> <p>This module contains a</p>				

Data Challenge Solutions



Winning Team (solution is executable on NDP)

Step 1: Predict Plant Functional Type

- **Features:** DIA, HT, LAT, LON, ECOSUBCD
- **Model:** Stacking Classifier Architecture with Ensemble methods
- **Meta-Learner:** XGBoost

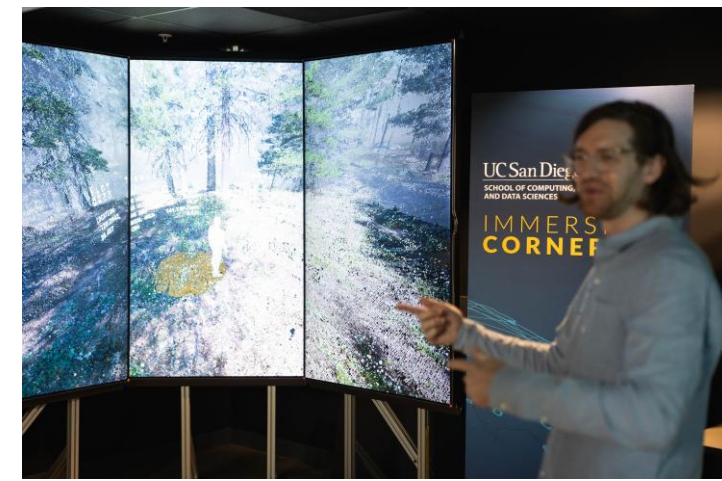
- **Accuracy:** 89.7%, F1 (weighted): 0.89

Step 2: Predict Genus

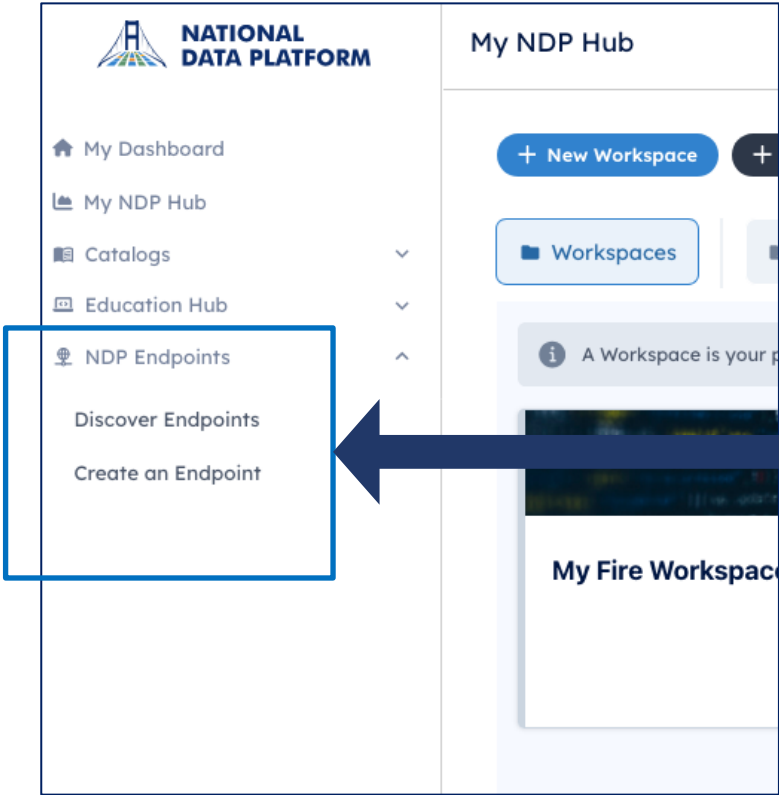
- Adds predicted PFT to features
- **Accuracy:** 83.3%, F1 (macro): 0.86

Step 3: Predict Species

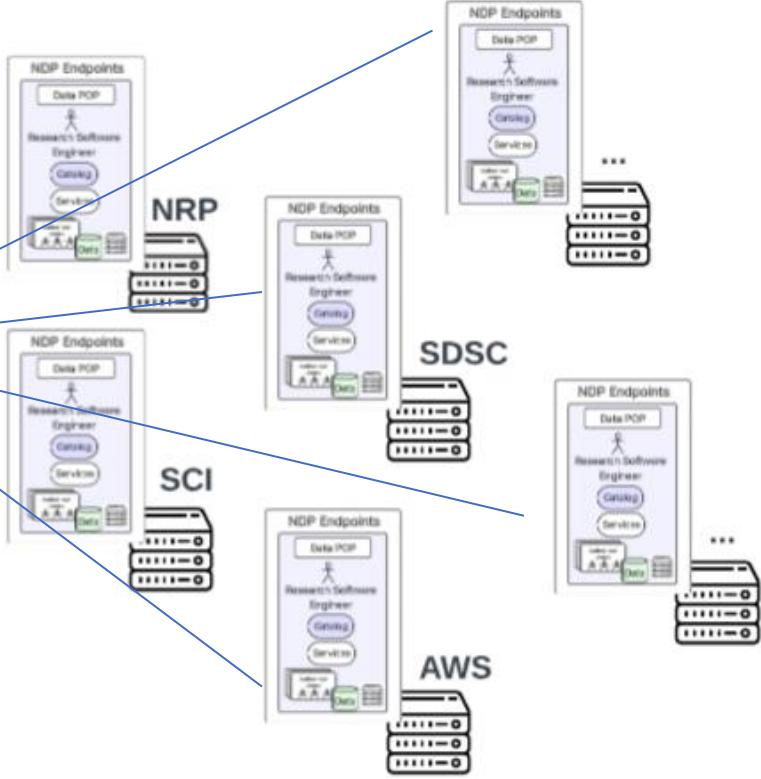
- Adds predicted Genus
- **Accuracy:** 96.9%, F1 (macro): 0.92



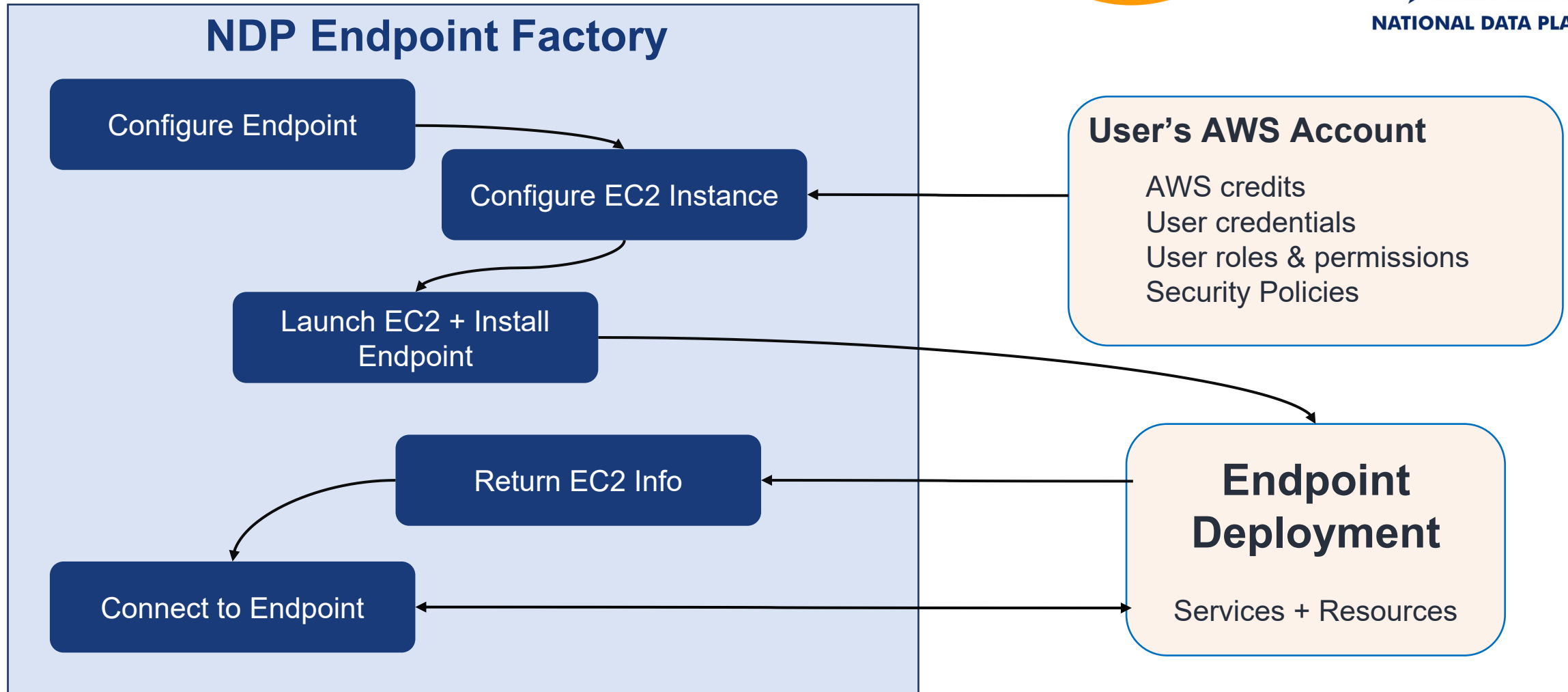
Endpoint Deployment on Composable Resources and Federation



Endpoint Federation



Launching Endpoints on

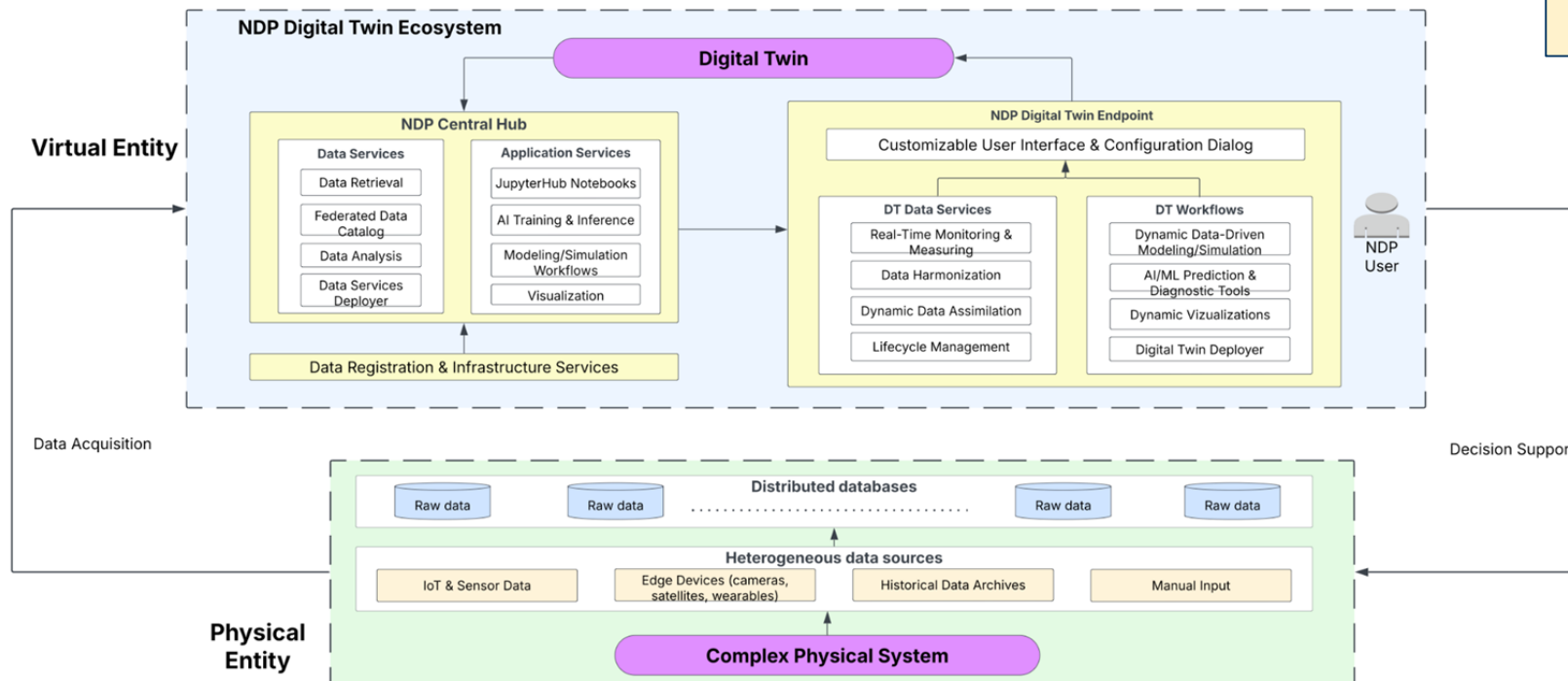


Using NDP for Digital Twin Applications

Contact: Hena Ahmed, PhD Student, UC San Diego



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
- WIFIRE Firemap based on the closed loop workflows for fire response, by Daniel Crawl and Hena Ahmed
- Immersive Forest based on multimodal fire environment data, by UC San Diego PhD student Isaac Nealey

Digital Twins for the Fire Environment

- Characterizing the dynamic fire environment : Variation of wind, smoke, moisture, fuels, fire perimeter, ...
- Detection of fire ignitions and other related fire effects
- Monitoring of active fire and smoke

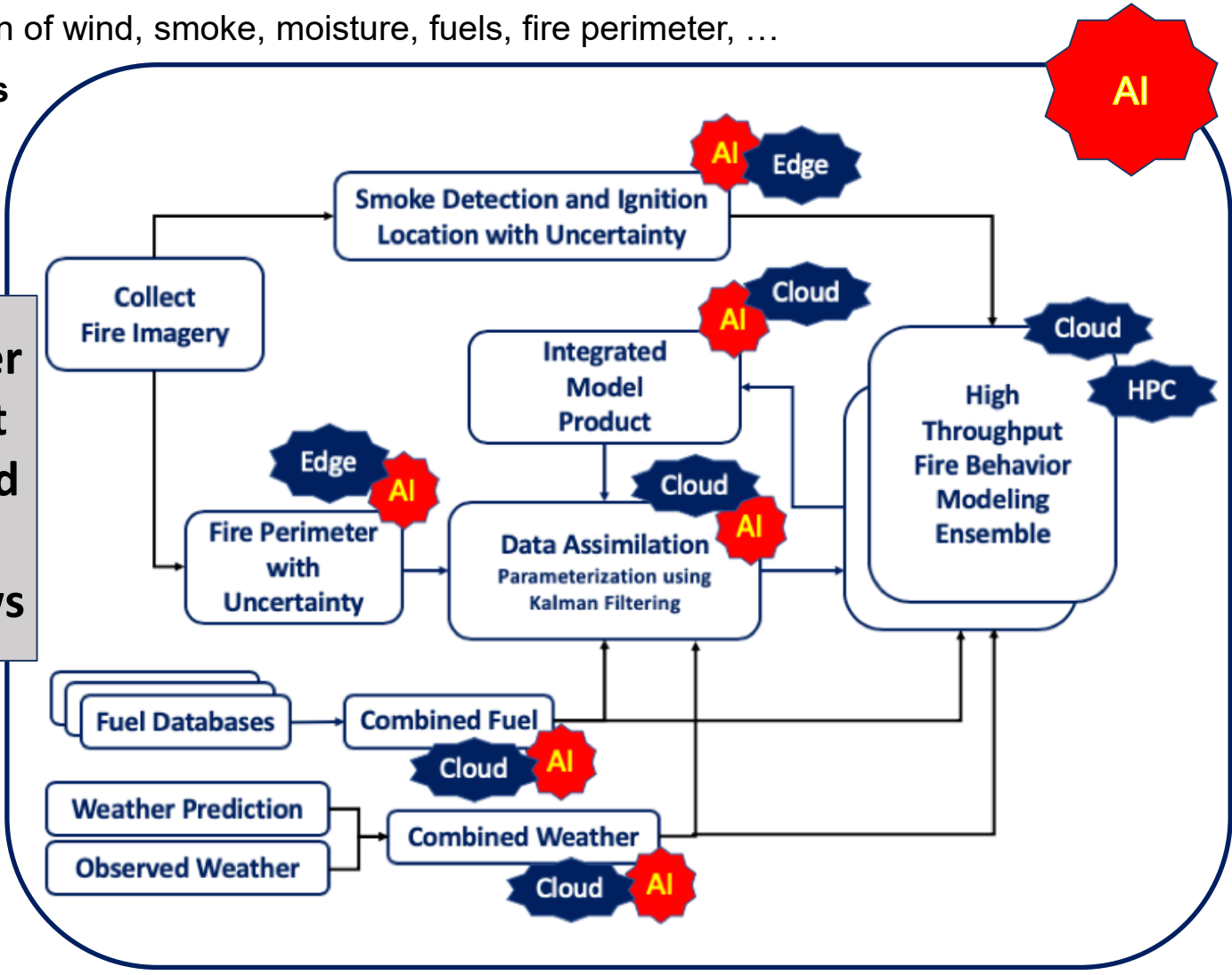
FIREMAP



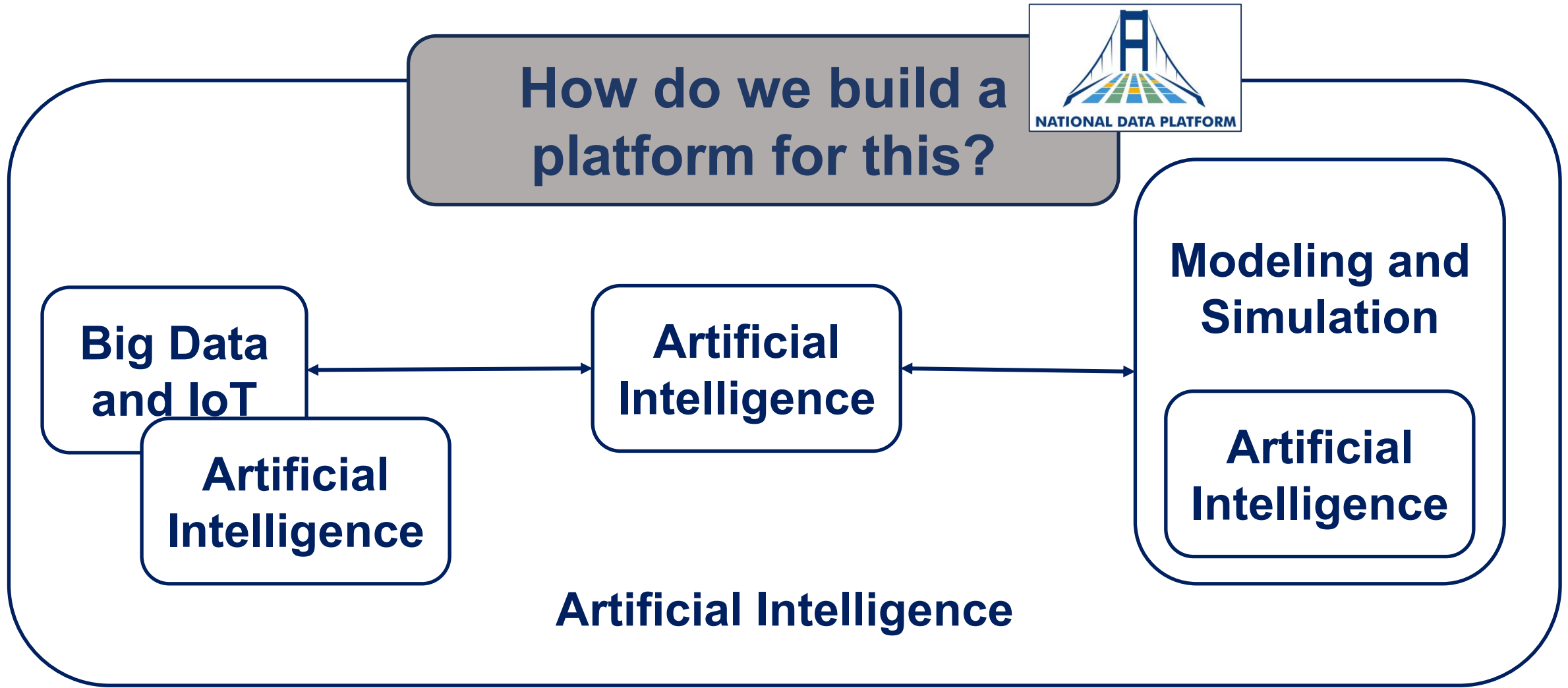
AI at the outer shell is about the integrated application and workflows

Since 2015

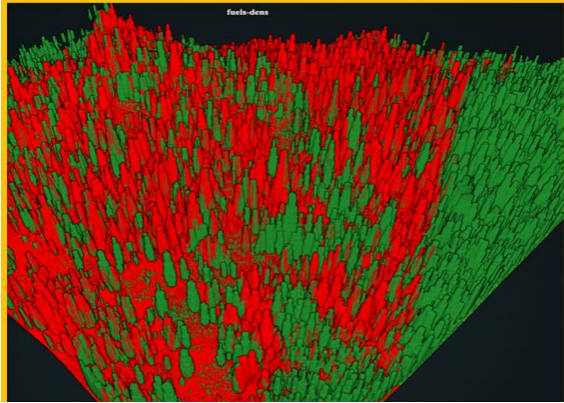
Firemap provides real-time information on weather conditions and fire ignitions to monitor and predict speed of fire spread and communities at risk.



Where are continuum workflows going?



The Future: Agentic AI in Workflows



- **Multi-Agent Collaboration** where each agent plays a role in a broader plan
- **Task Decomposition** to break complex goals into parallel subtasks
- **Memory Systems** to adopt based on feedback and context, and to optimize over time
- **Dynamic Orchestration** between role-based agents

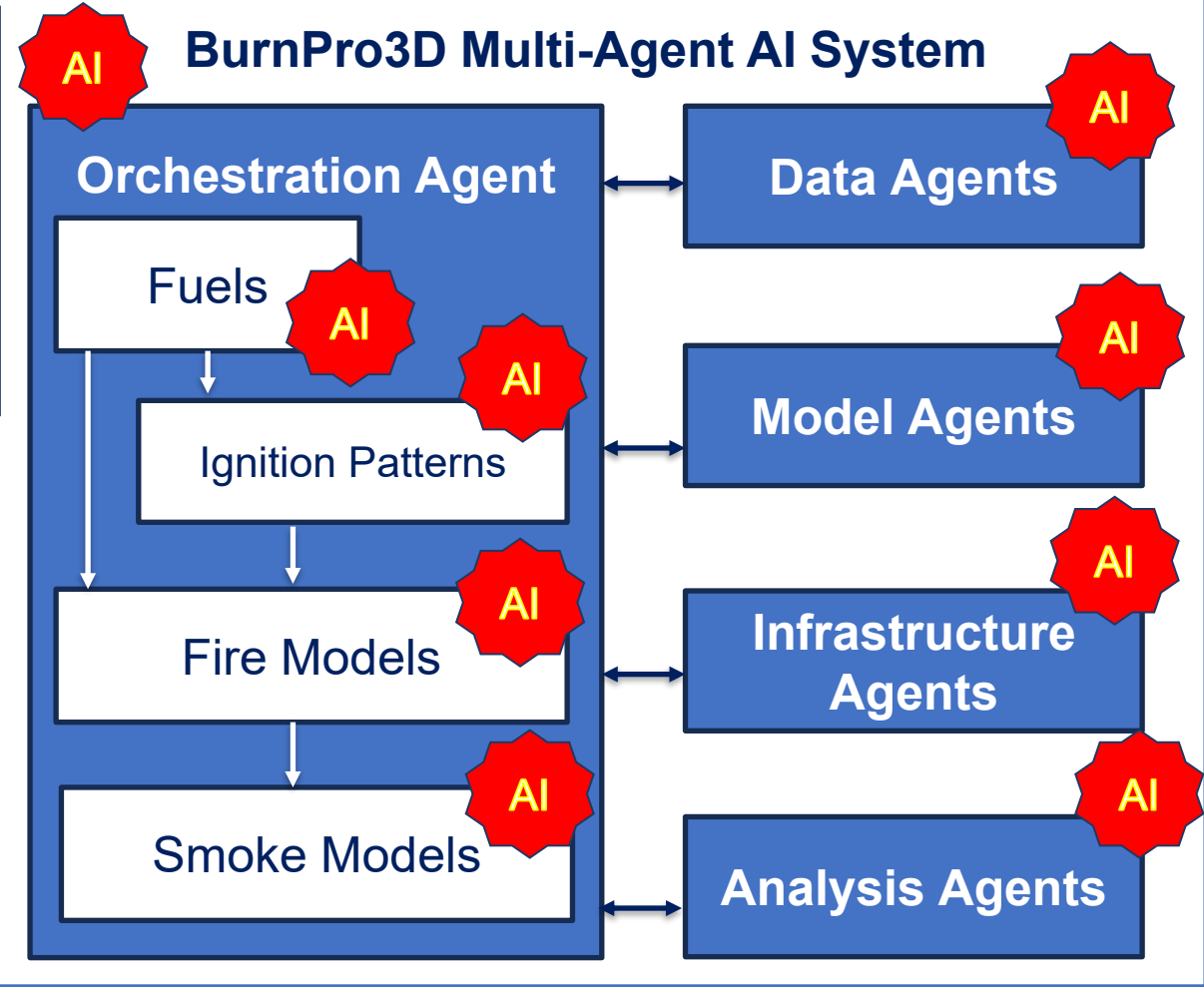
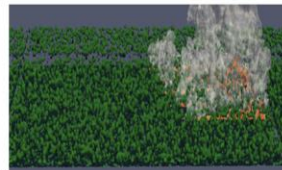
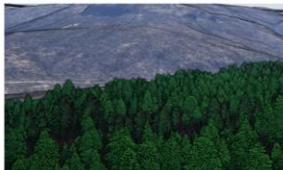
BurnPro3D allows burn bosses to create and visualize high-resolution 3D fire simulations and compare fuel consumption and risk under different weather and ignition scenarios.

FastFuels
3D Fuels for Fire Models

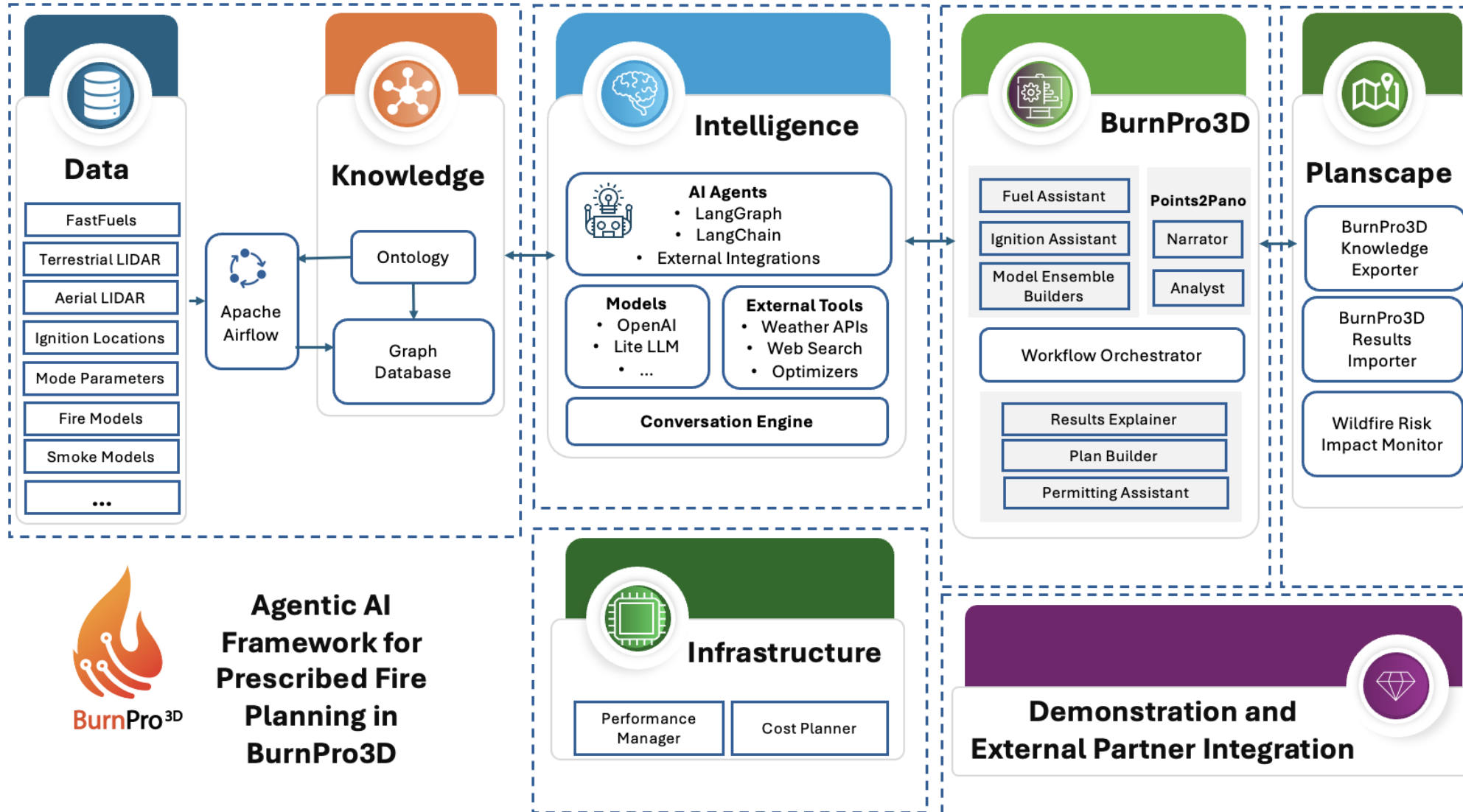
AI Models
Lidar-Derived Tree Positions, Heights and Species

DripTorch
Custom Ignition Patterns

QUIC-Fire
Prescribed Fire Simulator



Since 2020



Agentic AI Framework for Prescribed Fire Planning in BurnPro3D

Immersive Forest

Contact: Isaac Nealey, PhD Student, UC San Diego

