



# NEPA GIS

## Digitizing Historic NEPA Layers Using GeoAI

**Anastasia Bernat**

Data Scientist

Pacific Northwest National Laboratory



U.S. DEPARTMENT  
of ENERGY

**BATTELLE**

PNNL-SA-222332

PNNL is operated by Battelle for the U.S. Department of Energy

# PermitAI Mission

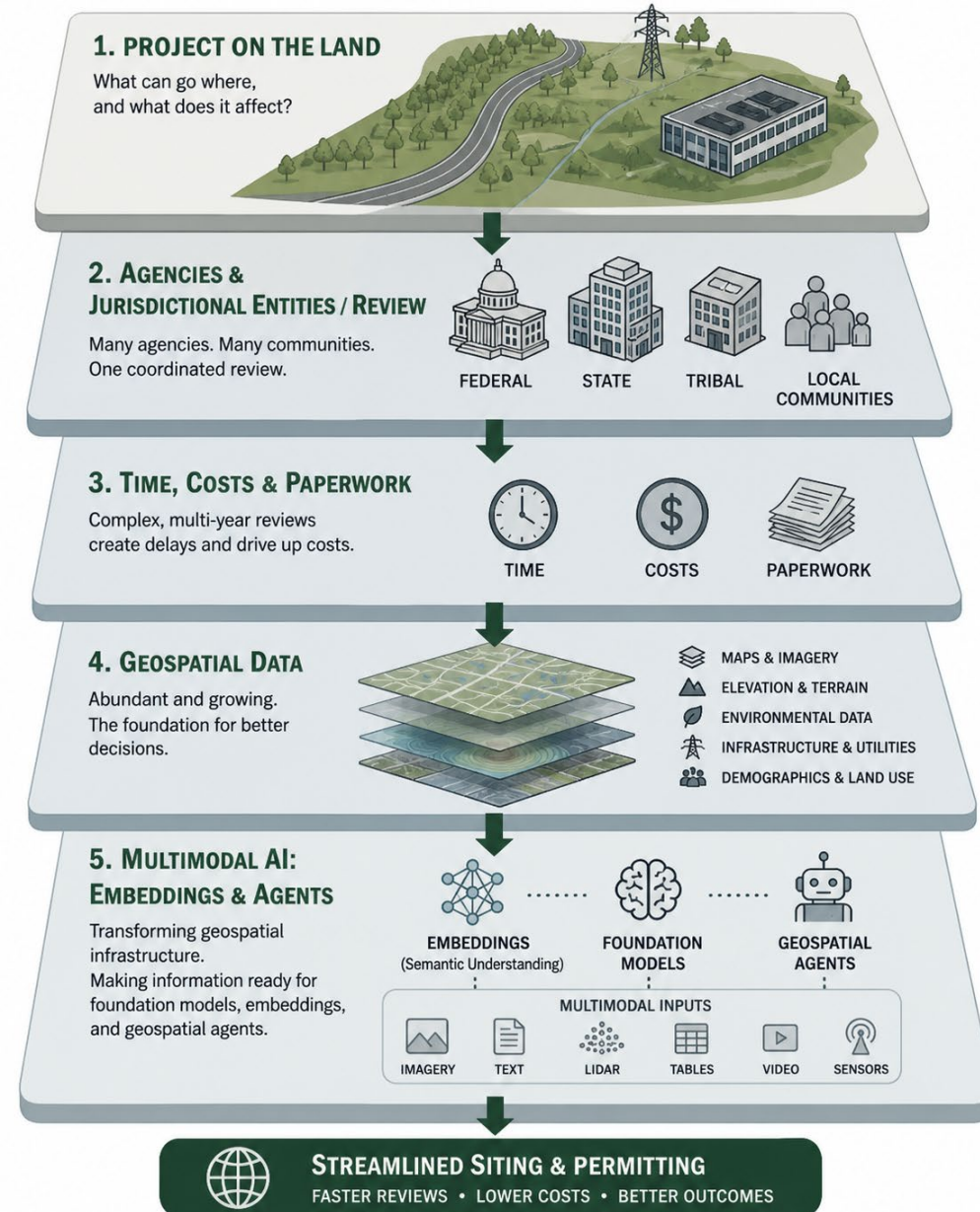
Every project on U.S. lands — whether it’s a road, a power line, or a data center — has to answer the same question: **where can it go and what does it affect?**

Environmental reviews for projects are often large, multiyear, and multicommunal public land management **undertakings**.

Agencies are under increasing pressure to complete high-quality reviews faster and cheaper

Text and geospatial information are abundant and AI is transforming knowledge infrastructure(s).

We can streamline permitting by making multimodal information **ready for a world of foundation models, embeddings, and agents**.



# GeoAI Reshapes the Pace and Precision of Permitting



## PERMIT-AI FAST FEDERAL PERMITTING



**NEPA maps are not optional illustrations**—they are the building blocks of geospatial infrastructure for government, and they act as policy tools that bound the affected area, communicate alternatives, quantify impacts, and track mitigation.

1

### ENVIRONMENTAL SCREENING

**Where is the action?  
What's the area of influence?**

Generate base maps and parse key spatial entities, like project boundaries, area of influence, and affected counties and jurisdictions.

2

### PROJECT SCOPING

**What resources may be affected?  
Who and where are stakeholders?**

Integrates map overlays (cultural resources, habitats, water, land use) with tribal consultation, land ownership, environmental review, and jurisdictional regulatory screening.

3

### PROJECT ALTERNATIVES

**What are reasonable alternatives?  
What sites best balance constraints?**

Evaluates project alternatives by digitizing spatial footprints and integrating socioeconomic, biodiversity, air quality, and flood risk data to compare impacts and identify alternatives.

4

### AFFECTED ENVIRONMENT

**What's there now (by  
resource)?**

Analyzes existing features and generates maps, documentation, and regulatory citations.

5

### IMPACT & MITIGATION

**What changes occur with  
each alternative?**

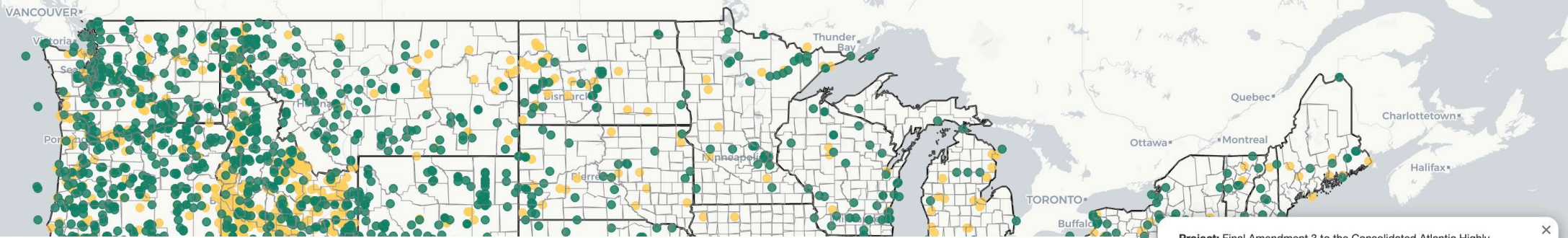
Run spatial overlay analyses to quantify and geovisualize impact areas to support public review and the record of decision.



**Pacific Northwest**  
NATIONAL LABORATORY

# Mapping Six Decades of NEPA Projects

Environmental reviews are often large, multiyear, and multicommunal public land management undertakings.



## NEPATEC Dataset



**60+**

AGENCIES



**61,881**

TOTAL PROJECTS



**142,083**

TOTAL FILES



**6,967,739**

TOTAL PAGES

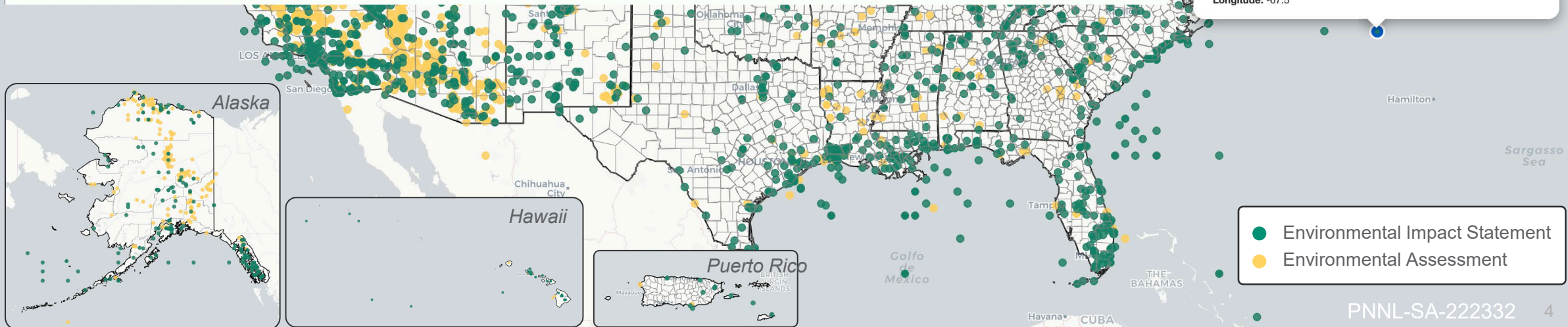


**1M+**

TOTAL MAPS



**Project:** Final Amendment 3 to the Consolidated Atlantic Highly Migratory Species Fishery Management Plan  
**Project Type:** Laws, Policies, Regulations, and Guidance; Ecosystem Management and Restoration; Threatened and Endangered Species Management  
**Project Sector:** Agriculture and Natural Resource Management; Environmental Policy and Regulation  
**Lead Agency:** Department of Commerce - National Oceanic and Atmospheric Administration  
**Process Type:** Environmental Impact Statement (EIS)  
**Location:** Marine zone located off the U.S. Atlantic coast (Lat/Lon: 35.0, -67.5)  
**Latitude:** 35.0  
**Longitude:** -67.5



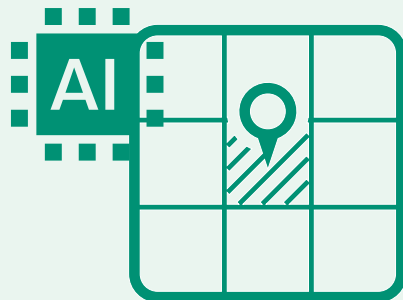
- Environmental Impact Statement
- Environmental Assessment

# Geointelligent Future of Federal Permitting

## PermitAI

### GEOAI PIPELINE

Intakes NEPA map images and documents to generate classified and georeferenced, geospatial layers.



### GEOBENCH

Design and develop expert-driven ground truths that can be used to evaluate pipeline output.



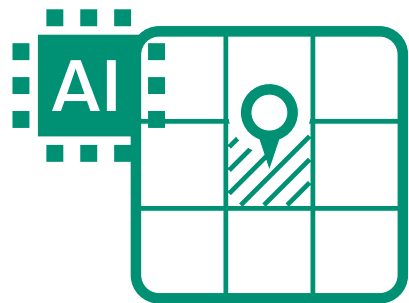
### NEPATEC

A high-quality, standardized dataset that collates and tags the NEPA corpus and its more than one million maps.



[pnnl.gov/projects/permitai](https://pnnl.gov/projects/permitai)

**PermitAI is a multimodal large language model (LLM) testbed** led by the Pacific Northwest National Laboratory and funded by the Department of Energy, Office of Policy (DOE OP) that seeks to use AI to streamline the environmental permitting and review process.



# GeoAI Pipeline



# Cartographic Challenges of NEPA Maps

## ABUNDANT AND HETEROGENEOUS

>1M maps with copious map types and map themes and a varying, open set symbology

## VARYING SPATIAL SCALES

NEPA projects vary highly in their spatial extent, breaking initial expectations of projects being merely county-level

## MIXED IMAGE RESOLUTIONS

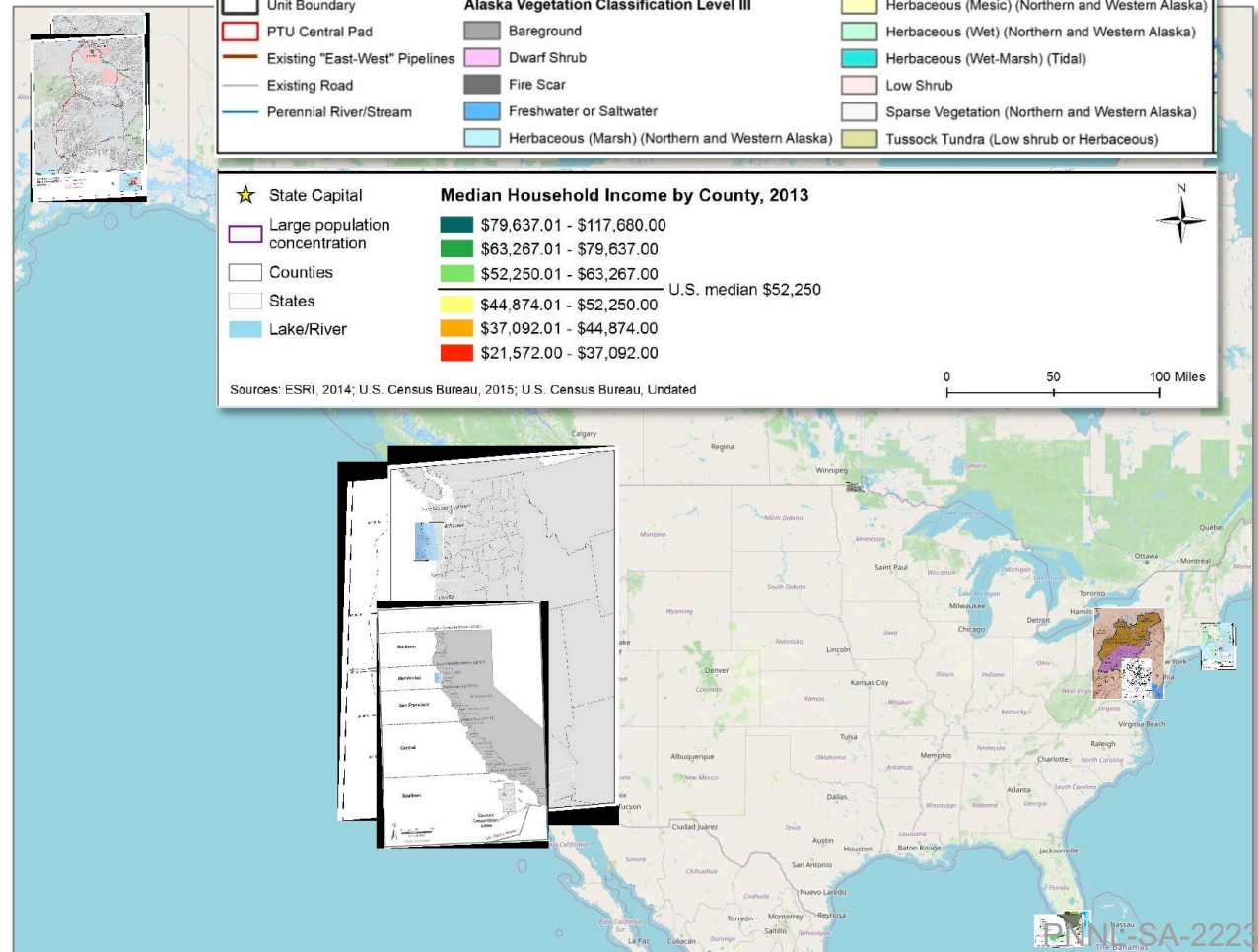
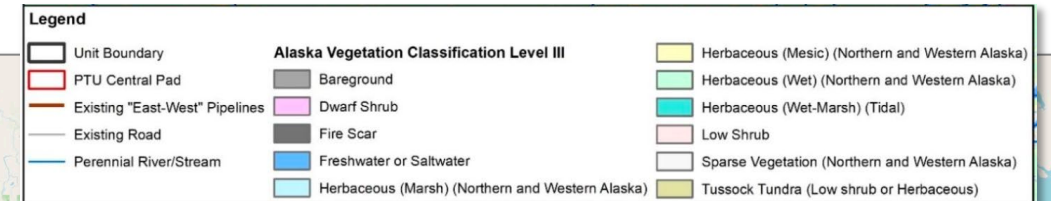
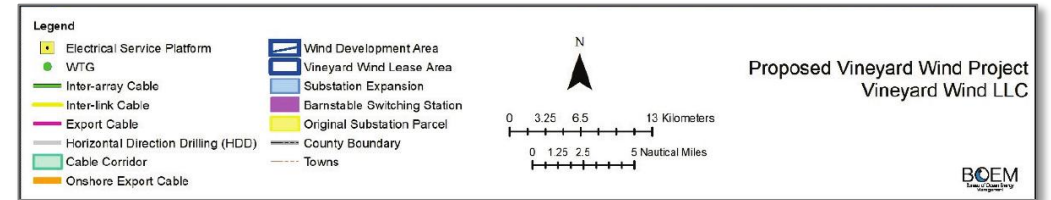
Blurry text from lower resolution can reduce the quality of text parsing

## MULTI-SITED PROJECTS

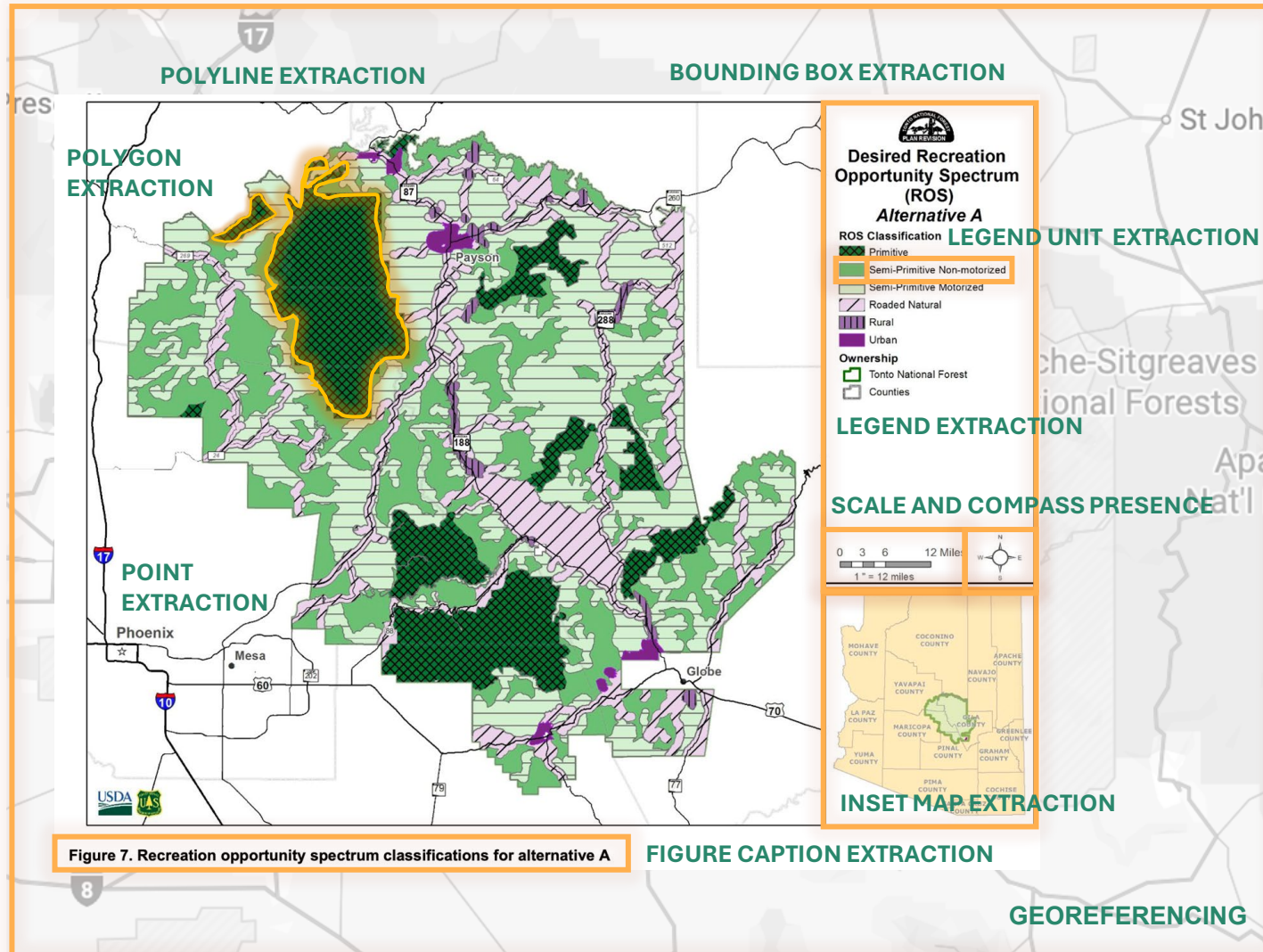
NEPA projects can be sited across multiple states and cite multiple locations, which can lead to “visual forgetting” on project area specifics

## CARTOGRAPHIC VS. IMAGE-LIKE MAPS

Maps can either exhibit detailed cartographic practices or depend heavily on visual characteristics to signal and support geospatial-specific enrichments



# Cartographic Maps



## LOCATION

Gila, Maricopa, Yavapai, Coconino, and Pinal Counties in Arizona

## MAP CLASSIFICATION

Map

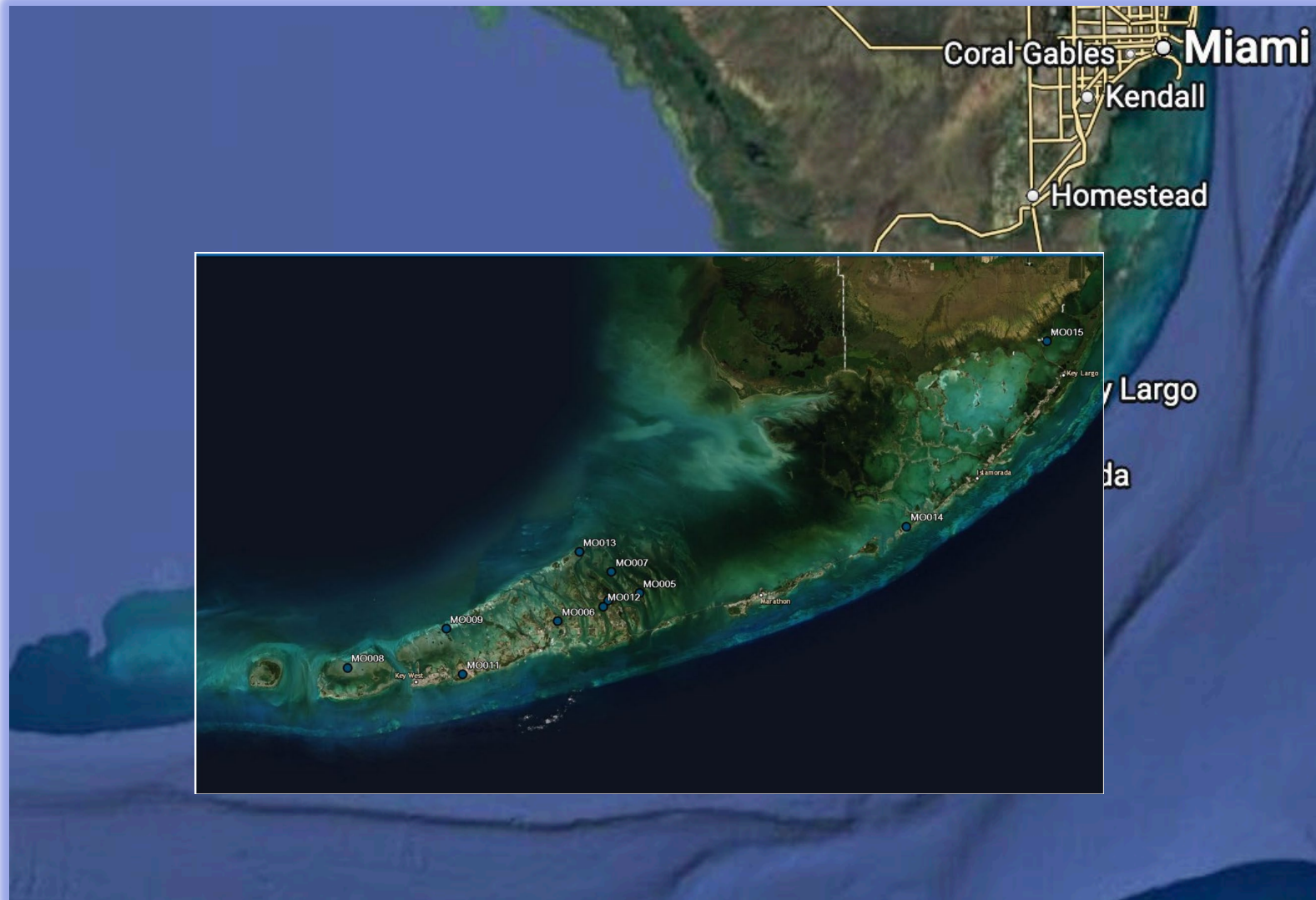
## MAP TYPE CLASSIFICATION

Political Administrative

## MAP THEME CLASSIFICATION

Proposed Action and Alternatives

# Image-Like Maps



## COLOR SEGMENTATION

Group pixels based on similarity to support feature extraction

## TEXTURE SEGMENTATION

Differentiate between e.g., land vs. water (speckled vs. smooth)

## FEATURE DESCRIPTIONS

Enables image-to-map alignment against different basemaps

## EDGE/CONTOUR ANALYSIS

Emphasize map features and reduce texture

# Choosing A Model for Multimodal Geospatial Data

## General Frontier Multimodal Vision-Language Models (VLMs)

*Broadest multimodal capability but not currently specialized to scientific discovery*

**GPT-4o**  
(OpenAI)

**Gemini 2.5 Pro**  
(Google)

**Claude 3.7 Sonnet**  
(Anthropic)

## Physics Geospatial Foundation Models (GeoFMs)

*Build for the physical world:  
simulations, geophysics, and spatiotemporal dynamics*

**Earth-2**  
(NVIDIA)

**GraphCast**  
(Google)

## Open Multimodal VLMs

*Open-weight models with growing vision-language capabilities*

**Qwen-VL**  
(Alibaba)

**LLaVA-NeXT**  
(ByteDance, HKUST, CUHK, NTU)

**GRAFT**  
(Cornell-Columbia)

## Earth Observation Foundation Models

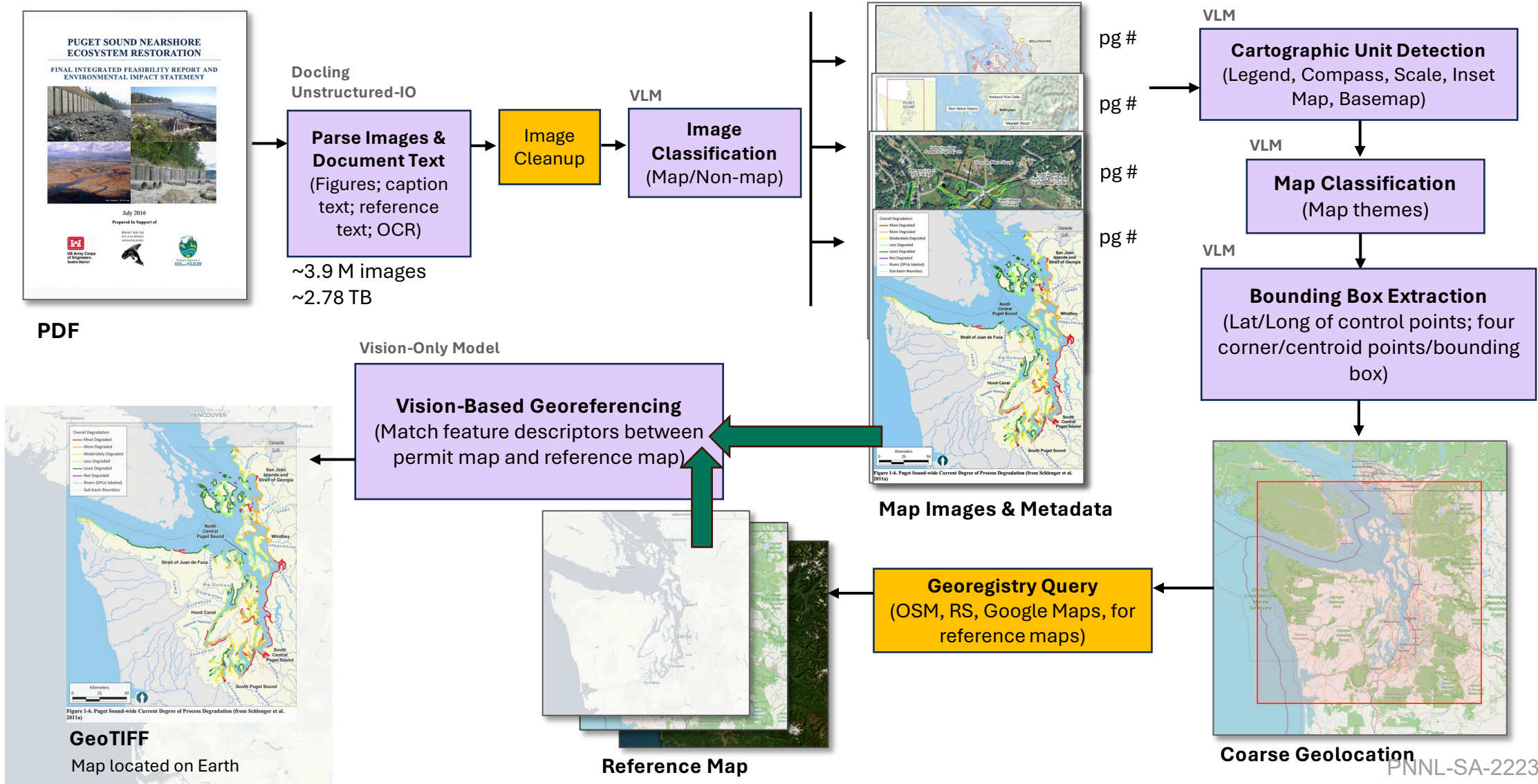
*Earth-observation-native models for spatial, spectral, and temporal understanding*

**Prithvi**  
(IBM-NASA)

**SatMAE**  
(Stanford)

**Clay Foundation Model**  
(Clay)

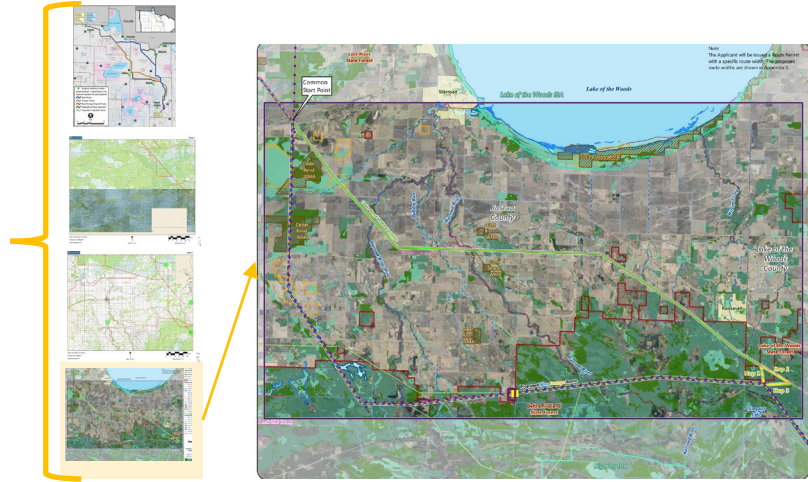
# GeoAI Pipeline to Digitize and Enrich Maps



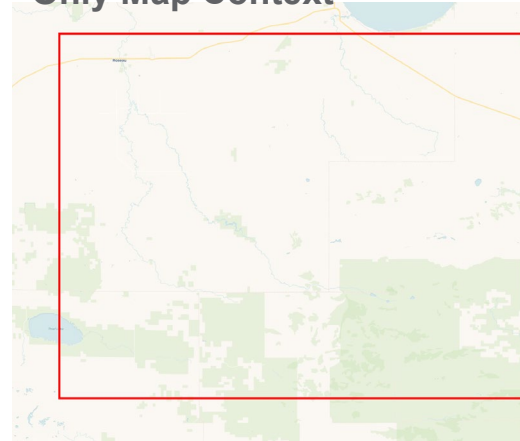
# Project Context Improves Locations

## BOUNDING BOX EXTRACTION

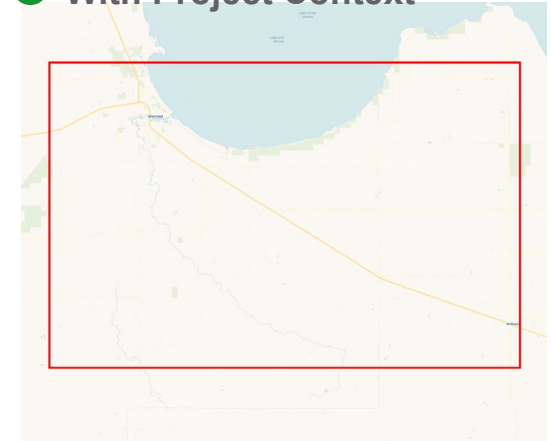
Great Northern Transmission Line Project



Only Map Context

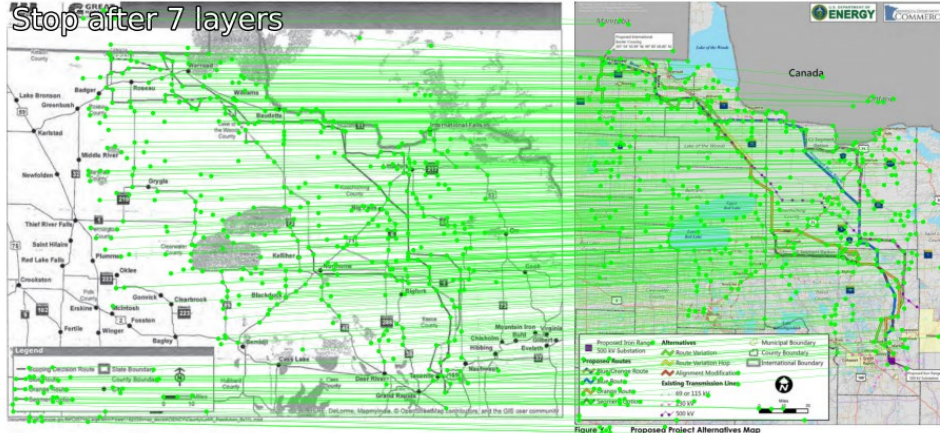


✓ With Project Context



## GEOREFERENCING

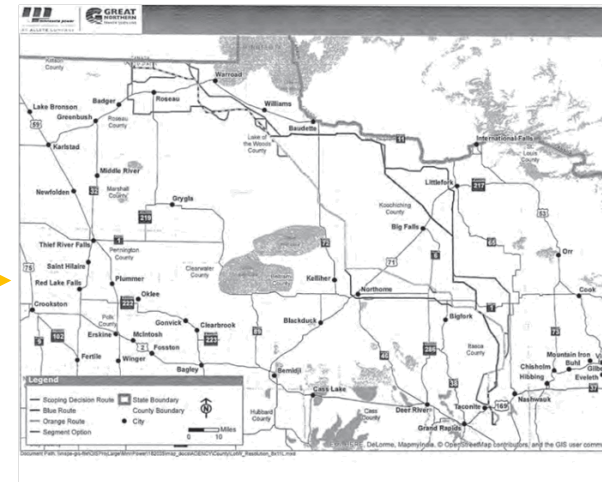
Stop after 7 layers



Successful georeference



Failed georeference

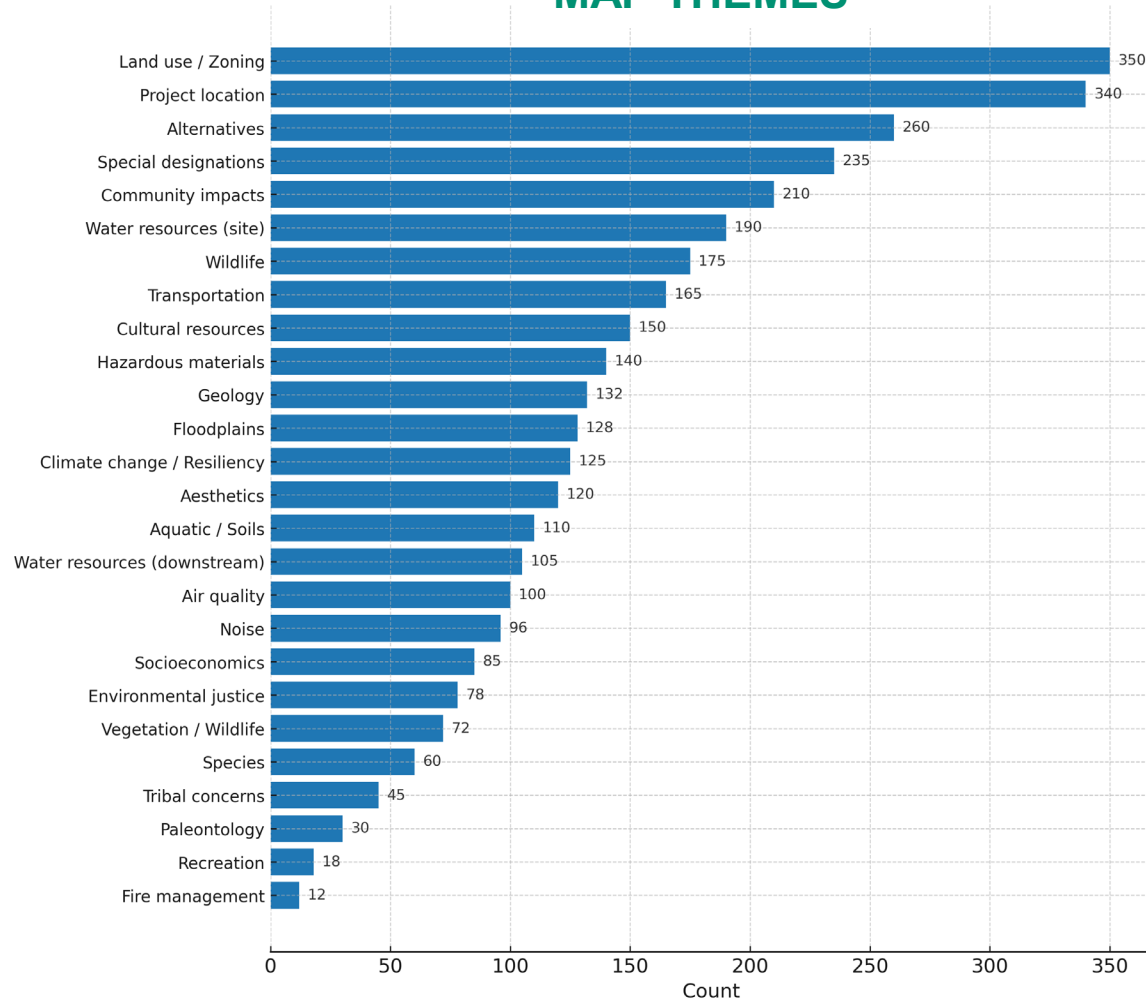


Recovered!

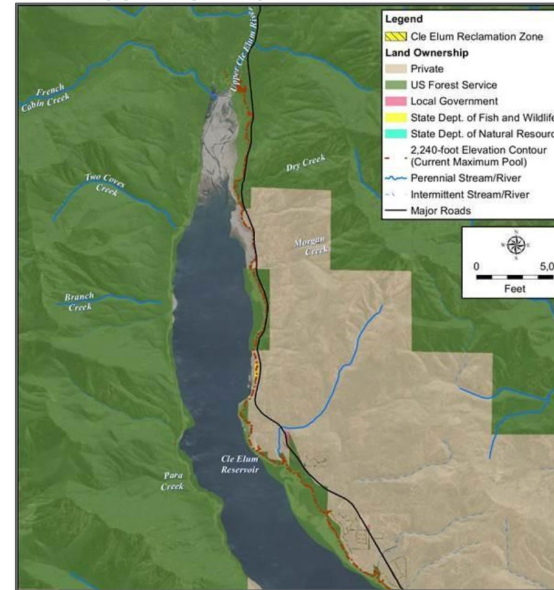
(animation)

# Map Classification

## MAP THEMES



## LAND USE / ZONING



## ALTERNATIVES

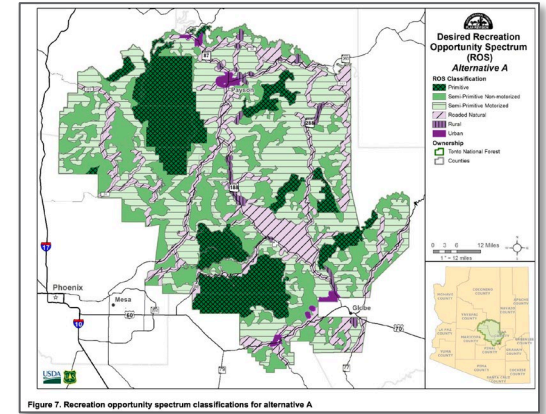
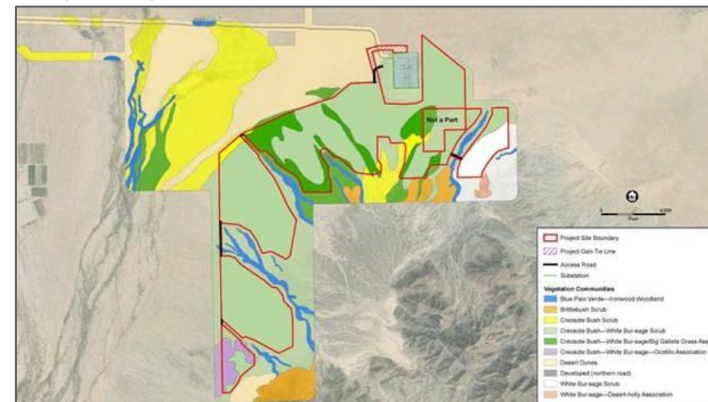


Figure 7. Recreation opportunity spectrum classifications for alternative A.

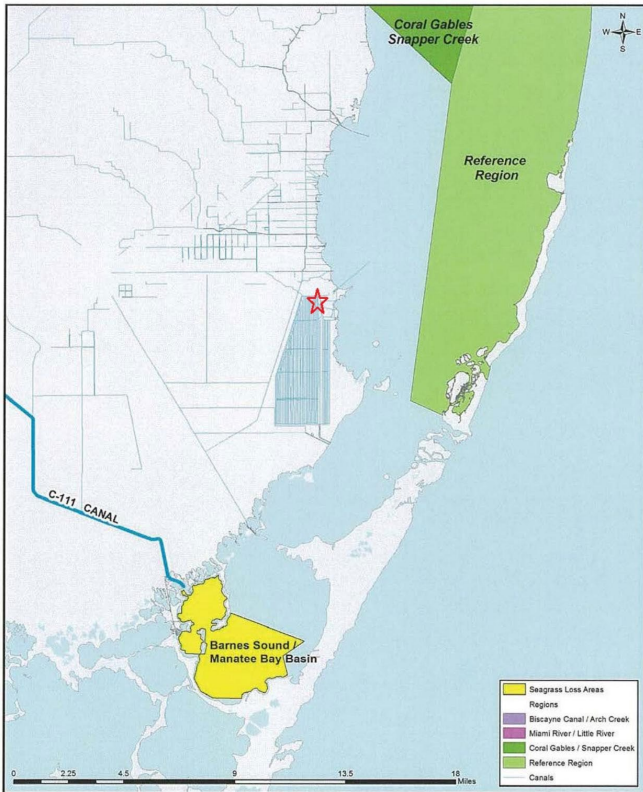
## VEGETATION





Pacific Northwest  
NATIONAL LABORATORY

# Extract Layers with VLMs



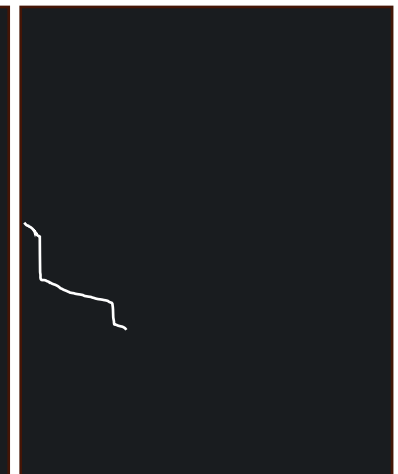
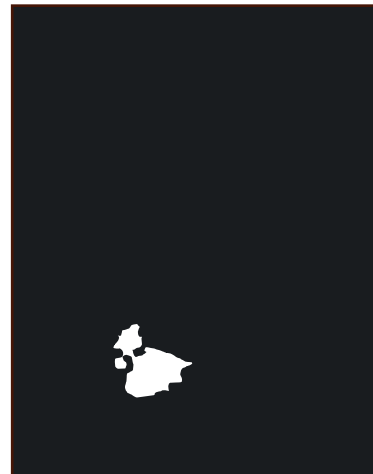
★ Turkey Point Site

Dominant Colors



#C0DFE8 #A3CF77 #ADC5C9 #EDE839 #78A253

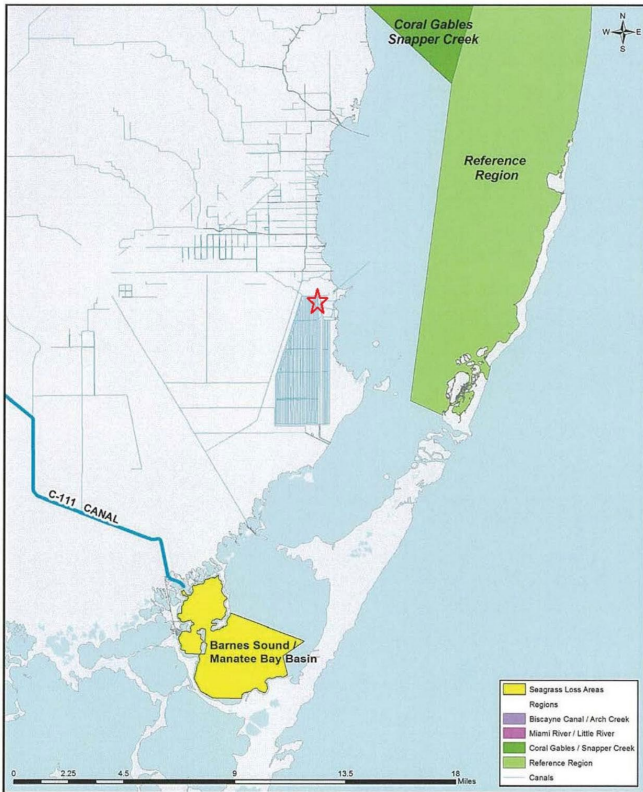
SHAPE	COLOR	MAP UNIT NAME
Rectangle	Yellow	Seagrass Loss Areas
Rectangle	Purple	Biscayne Canal / Arch Creek
Rectangle	Pink	Miami River / Little River
Rectangle	Green	Coral Gables / Snapper Creek
Rectangle	Light Green	Reference Region
Line	Light Blue	Canals





Pacific Northwest  
NATIONAL LABORATORY

# Extract Layers with VLMs



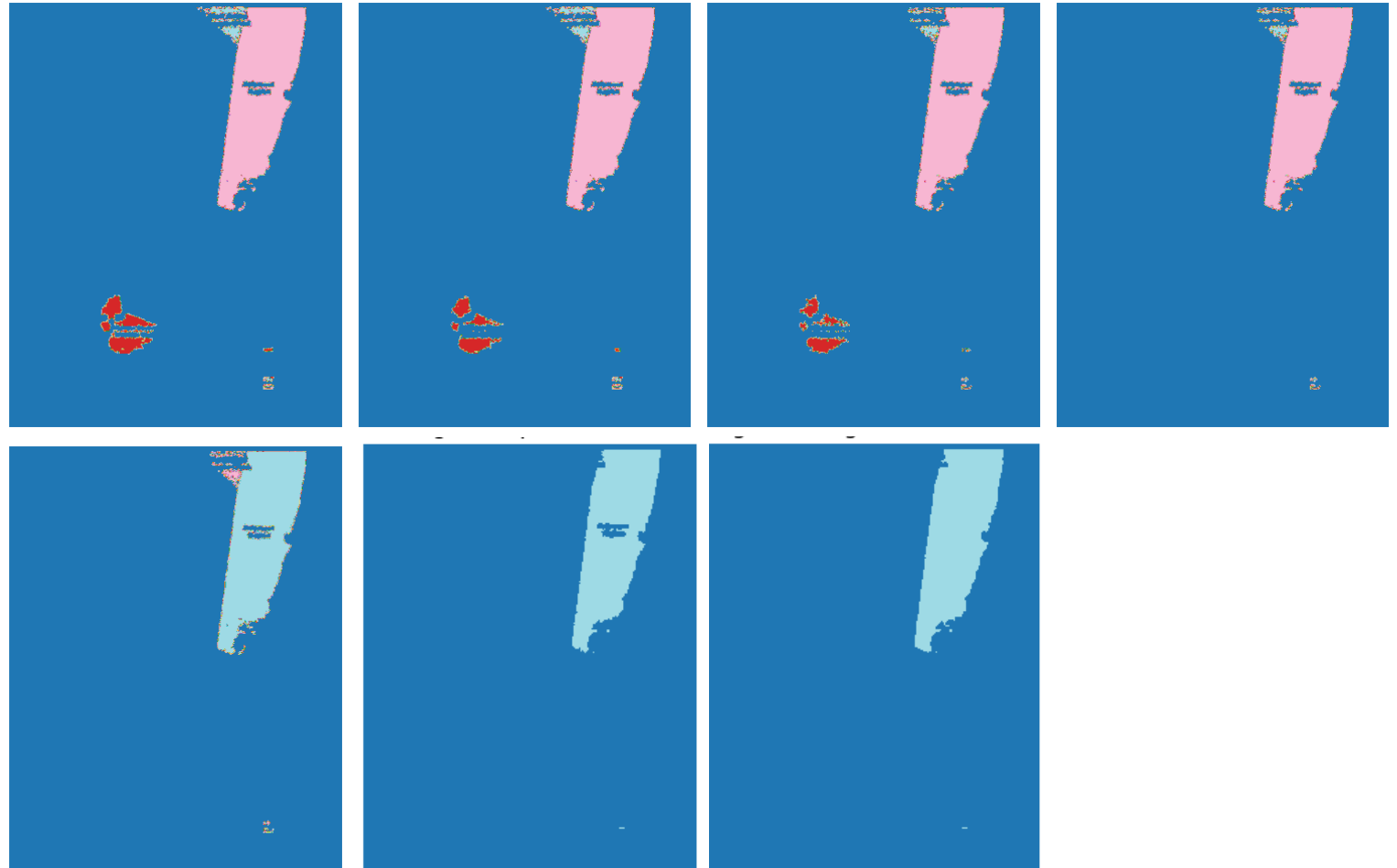
★ Turkey Point Site

Dominant Colors



#C0DFE8 #A3CF77 #ADC5C9 #EDE839 #78A253

Iterations





**Pacific  
Northwest**  
NATIONAL LABORATORY



# GeoBench



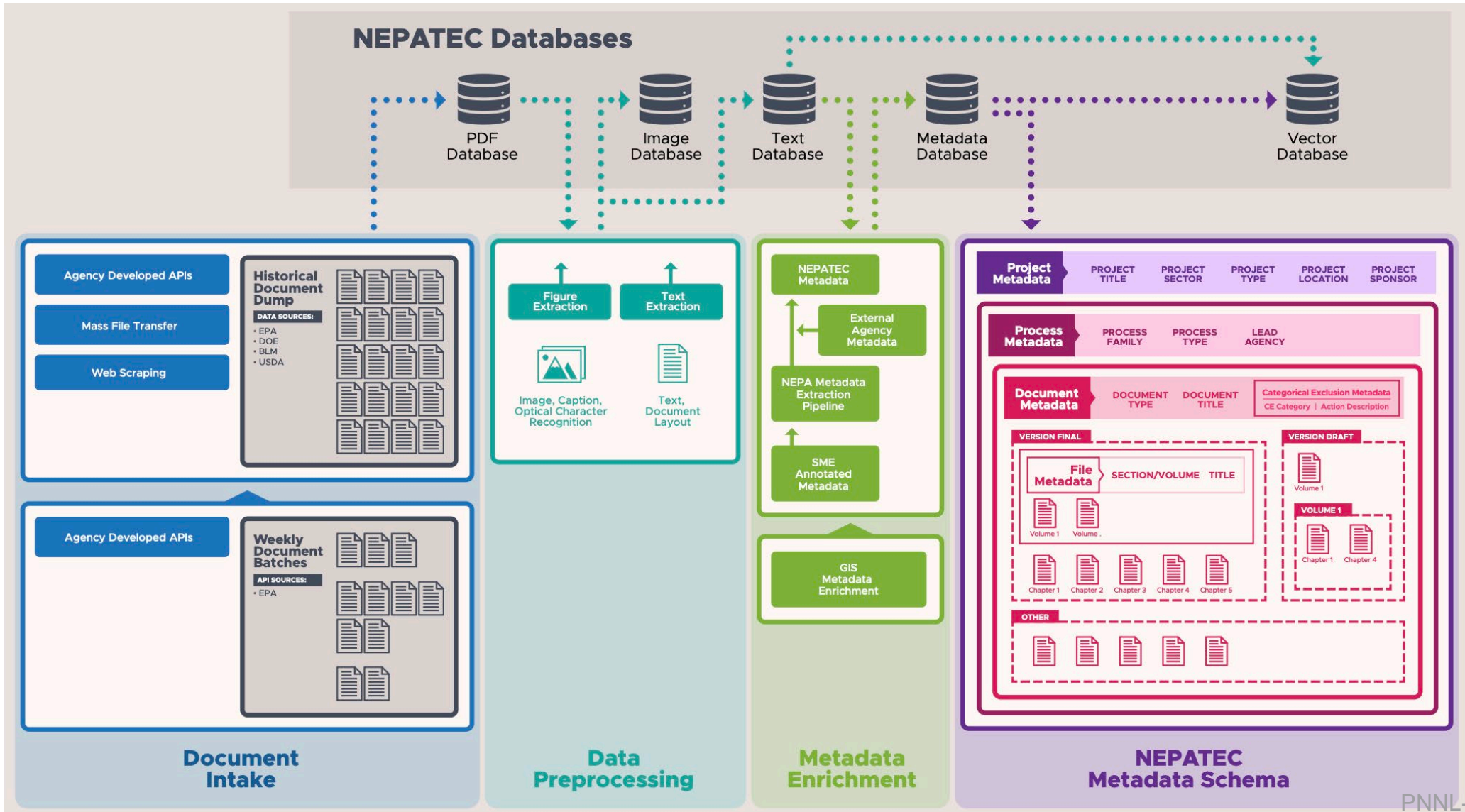
# Evaluations

Metadata Attribute	Accuracy (#correct/#total)	Completed Human Annotations
Project Location	95.00% (112/118)	Read through document and interpret project centroid
Image Classification	97.01% (2,726/2,810)	Bin maps into map/non-map
Map Type Classification	64.96% (2,748/4,230)	Bin maps into map types [15 options]
Map Theme Classification	51.73% (1,992/3,851)	Bin maps into map themes [33 options]
Figure Caption Extraction	60.00% (20/21)	Transcribe figure captions
Scale Bar Presence	97.44% (114/117)	Check presence of scale bar in the map
Compass Presence	99.15% (116/117)	Check presence of compass in the map
Legend Presence	98.3% (115/117)	Check presence of legend in the map
Basemap Identification	<i>Underway</i>	Bin maps into basemap types [4 options]
Inset Map Presence	99.15% (116/117)	Check presence of inset map in the map
Legend Extraction	51.85% (14/27)	Collect pixel coordinates of corners of legend
Legend Unit Extraction	<i>Underway</i>	Transcribe legend text, crop out symbols
Point Extraction	<i>Underway</i>	Georeference using ground control points in QGIS
Polyline Extraction	85.00%	Trace and export polyline features with QGIS
Polygon Extraction	92.00%	Trace and export polygon features with QGIS
Bounding Box Extraction	74.19% (46/62)	Export bounding box from georeferenced map
Georeferencing	52.77% (353/669)	Georeference map using QGIS



NEPATEC

# NEPATEC Production



# NEPATEC3.0+ Geospatial Data

## NEPATEC “Map Gallery”

### DOCUMENT

1+ million maps arriving to NEPATEC

```

"pages": [
  {
    "page number": "1-4",
    "page text": "US. Department of the Interior
    Bureau of Land Management Carson City District
    Office CATEGORICAL EXCLUSION REVIEW..."
  },
  {
    "map image": base64 encoded image, MAP IMAGE
    "cartographic_elements": {
      legend: yes/no,
      inset_map: yes/no,
      north_arrow: yes/no,
      scale_bar: yes/no,
      "basemap": "satellite"
    },
    "figure caption text": "Figure 7. Recreation
    spectrum classifications
    for alternative A",
    "figure references": ["There are 4 alternatives
    as shown in Figure 7."],
    "map type": "political administrative",
    "map theme": "alternatives",
    "legend_units": {
      "unit_1": {"geometry": "line/point/polygon"},
      "unit_2": {"geometry": "line/point/polygon"},
      "unit_3": {"geometry": "line/point/polygon"},
    }
  }
]

```

## NEPATEC “One Image of NEPA”

### PROJECT

Georeferenced features (GeoJSON)

```

{
  "type": "FeatureCollection",
  "bbox": [-125.0, 24.4, -66.9, 49.4],
  "features": [
    "common legend [image]":
    {
      "type": "Feature",
      "properties": {
        "label": "Affected Area",
        "marker-symbol": "circle",
        "marker-color": "#0a84ff",
        "stroke": "#003366",
        "stroke-width": 1.5,
        "stroke-opacity": 1.0,
        "fill": "#66cc66",
        "fill-opacity": 0.9
      },
      "geometry": {
        "type": "Point",
        "coordinates": [
          [-124.4096, 40.0], [-123.0, 42.0],
          [-114.1315, 41.0],
          [-114.1315, 32.53], [-124.4096, 40.0]
        ]
      }
    }
  ]
}

```

**LEGEND UNIT**

**FEATURE**

# NEPATEC Releases

## NEPATECv3.0

ONGOING FY26

### Structured Schema

#### Process

- Lead Agency
- Participating Agencies
- Process Family
- Process Type

#### Project

- Project Title
- 📍 **Site Bounding Box(es)**
- Project Sponsor
- Project Sector
- Project Type
- Date

#### Document

- Document Type
- Programmatic Documents
- Supplemental Documents
- Document Title
- Pages

- Text
- 📍 **Map Images/GeoTIFFs**
- **Figure Captions**

- Prepared By
- CE Category
- Action Description

#### File

- Section Or Volume Title
- Main Document

## NEPATECv2.0

😊 AUGUST 2025 | 142,083 Documents

### Structured Schema

#### Process

- Lead Agency
- Process Family
- Process Type

#### Project

- Project Title
- 📍 **Location**
- Project Sponsor
- Project Sector
- Project Type

#### Document

- Document Type
- Document Title
- Prepared By
- CE Category
- Action Description

#### File

- Section Or Volume Title
- Main Document

## NEPATECv1.0

😊 JUNE 2024 | 26,085 Documents

### Structured Schema

#### Project

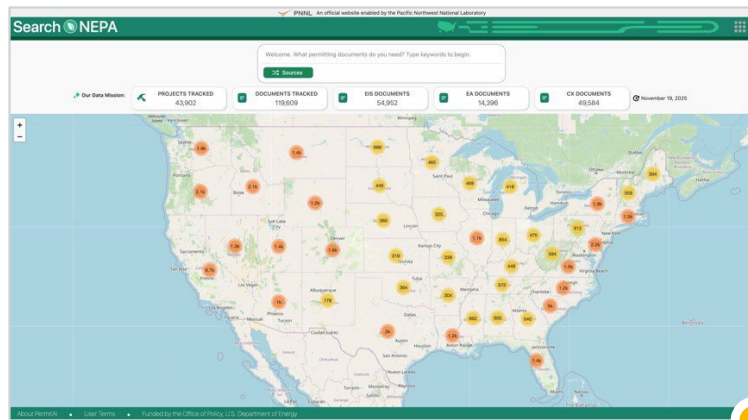
- Project Title
- Project Process
- Agency
- Location

#### Document

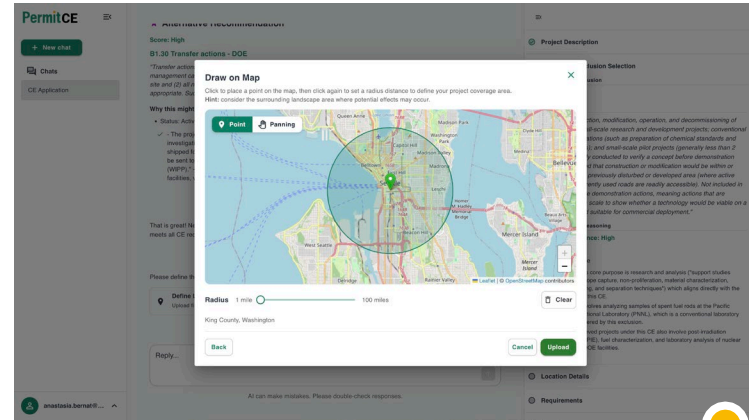
- Document Title
- Document Version
- Table of Content
- Pages
  - Text
  - Named Entities

# Building Blocks

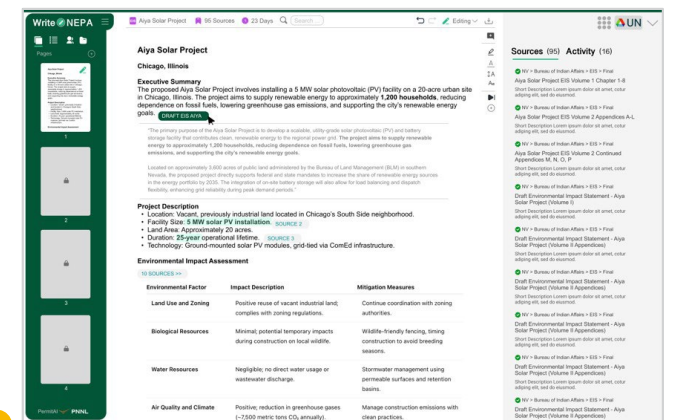
## THE PERMIT-AI APPLICATION SUITE AND PERMIT ATLAS TESTBED FOR ACCELERATING PERMITTING



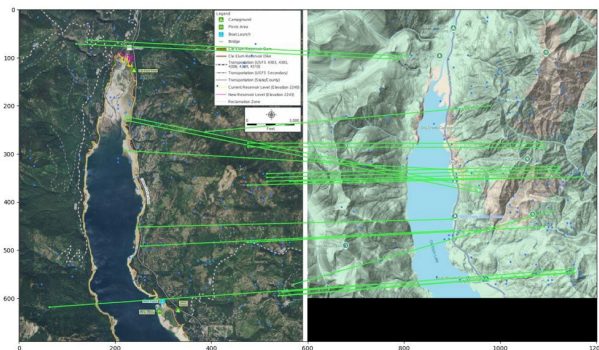
**SearchNEPA.** A document discovery tool that geovisualizes the most comprehensive, AI-ready database of NEPA documents.



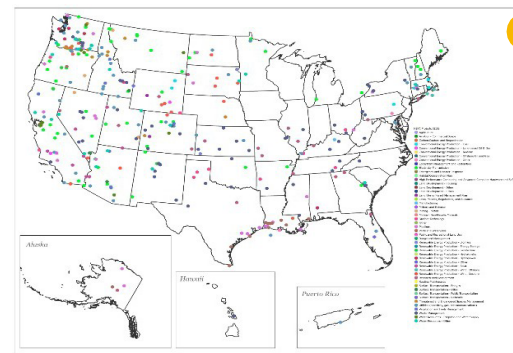
**PermitCE.** A multi-agentic tool for preparing CE documents with automated environmental screening of project ROI.



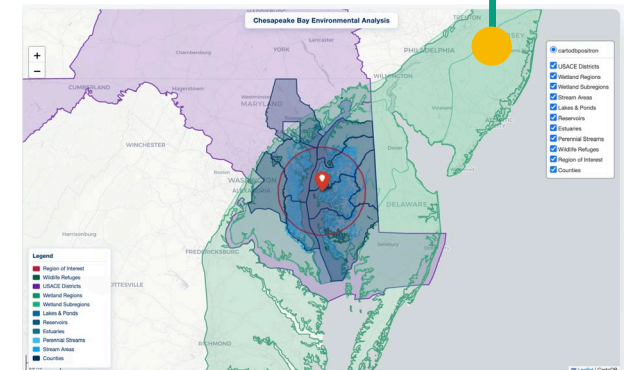
**WriteNEPA.** An AI-drafting tool that automatically produces accurate, well-structured first drafts of EIS sections.



**GeoAi Pipeline.** Intakes NEPA map images and documents to generate classified and georeferenced, geospatial layers.



**NEPAtec.** A high-quality, standardized dataset that collates and tags the NEPA corpus and its more than one million maps.



**NEPA-MCP.** A multi-step MCP tool chain exposing 42 tools through one protocol for rapid map generation.

# PermitAI Contributors



**Johnny Chen**  
DATA SCIENTIST

Ask me how I use vision models to georeference maps.



**Siddhartha Das**  
POST DOCTORATE RA

Ask me how I use frontier VLMs to georeference maps.



**Ben Chauhan**  
TECH STUDENT

Ask me how I use frontier models to classify and georeference maps.



**Sridevi Wagle**  
ML ENGINEER

Ask me how I use frontier parsers to extract maps from documents.



**Cleve Davis**  
EARTH SCIENTIST

Ask me how I advise as a domain SME and build the GeoBench.



**Micah Taylor**  
COMPUTATIONAL SCIENTIST

Ask me how I design and build the GeoBench and extract geotext.



**Heng (“Alfred”) Wan**  
EARTH SCIENTIST

Ask me how I use frontier models to extract map features and legends.



**Tim Vega**  
SOFTWARE ENGINEER

Ask me how I use cloud to parse maps from documents at scale.



**Renuka Chintalapati**  
DATA SCIENTIST

Ask me how I use evaluate agentic systems.



**Dan Nally**  
NEPA EXPERT AND DATA THRUST LEAD

Ask me how I manage NEPATEC development and releases.



**Sai Koneru**  
POST DOCTORATE RA

Ask me how I perform metadata enrichments from NEPA text.



**Sameera Horawalavithana**  
SENIOR DATA SCIENTIST

PermitAI PI



**Sai Munikoti**  
DATA SCIENTIST

PermitAI co-PI



**Pacific  
Northwest**  
NATIONAL LABORATORY

# Thank You

**Anastasia Bernat**

[anastasia.bernat@pnnl.gov](mailto:anastasia.bernat@pnnl.gov)



**PermitAI**



NEPATEC2.0