



Accelerating Environmental Permitting Through AI- Driven Data Enrichment

The NEPATEC v2.0 Framework

Kaustav Bhattacharjee

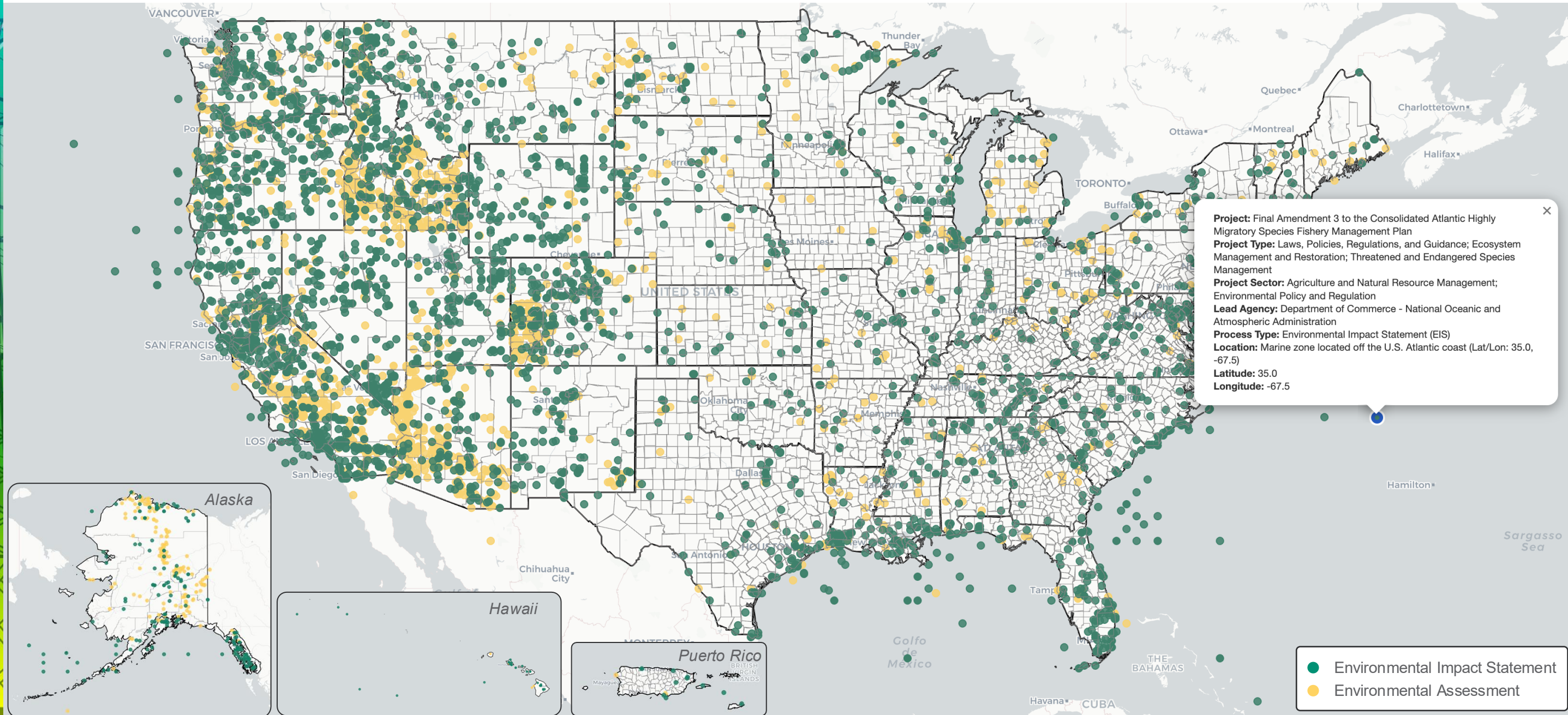
Data Scientist
Pacific Northwest National Laboratory



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

Vision: Mapping Six Decades of NEPA Projects

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





PermitAI - Faster Federal Permitting using AI



FAST FEDERAL PERMITTING



AGENCIES ARE UNDER INCREASING PRESSURE TO COMPLETE HIGH-QUALITY REVIEWS FASTER & CHEAPER

What is PermitAI?

PermitAI is an AI-powered data and application platform developed by PNNL to accelerate federal environmental permitting processes, starting with NEPA, for critical infrastructure projects.

Why it matters

PermitAI addresses fragmented, document-heavy permitting systems. It unlocks insights from millions of legacy documents (such as, CEs, EAs, EISs) using large language models and structured metadata.

Mission

PermitAI will develop and deploy next-generation AI tools and technology to modernize environmental reviews and permitting across federal government agencies.

Impact

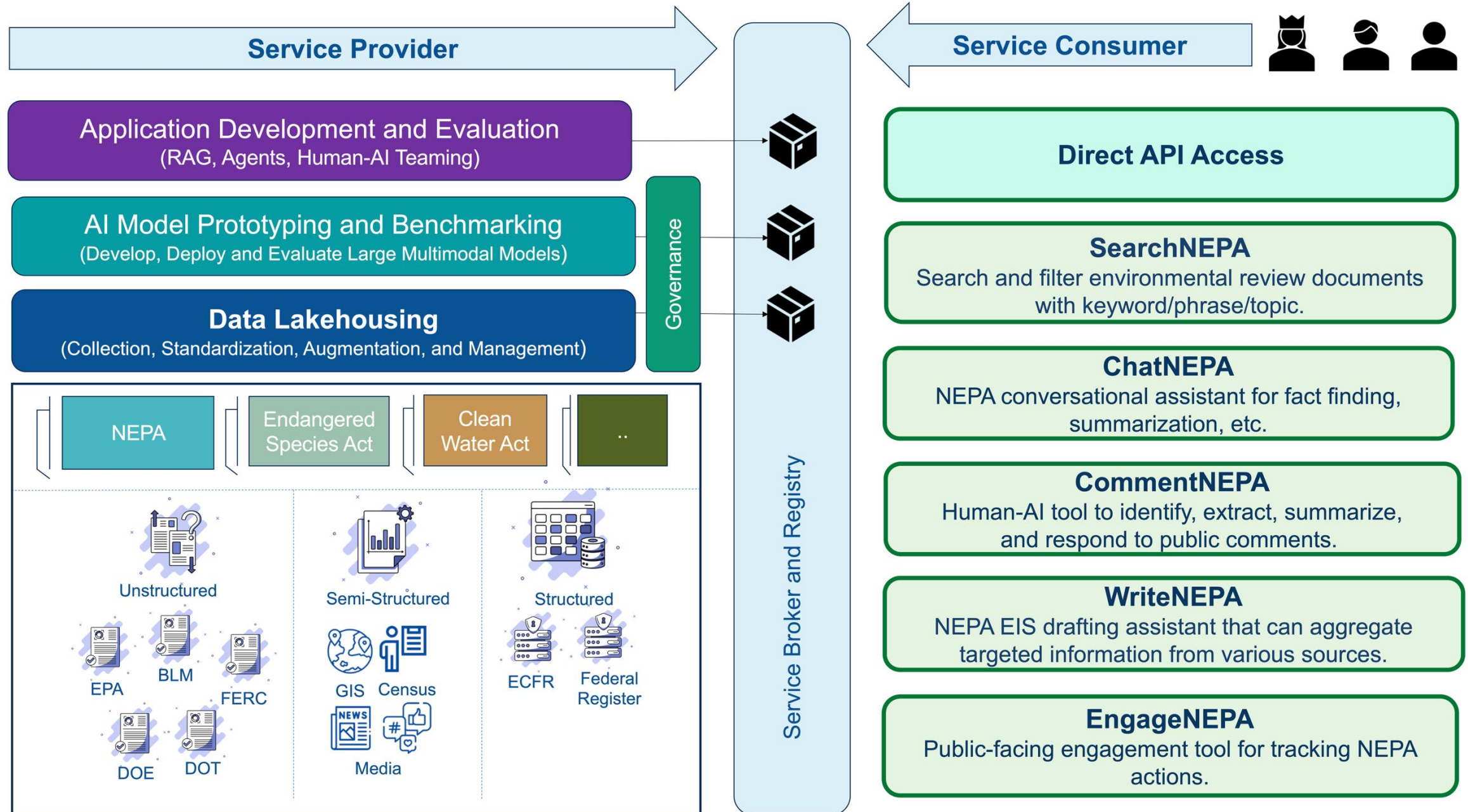
PermitAI applications are already showing up to 50% efficiency gains for federal NEPA SMEs.

<https://www.pnnl.gov/projects/permitai>

[Presidential Actions, Updating Permitting Technology for the 21st Century](#), April 15, 2025

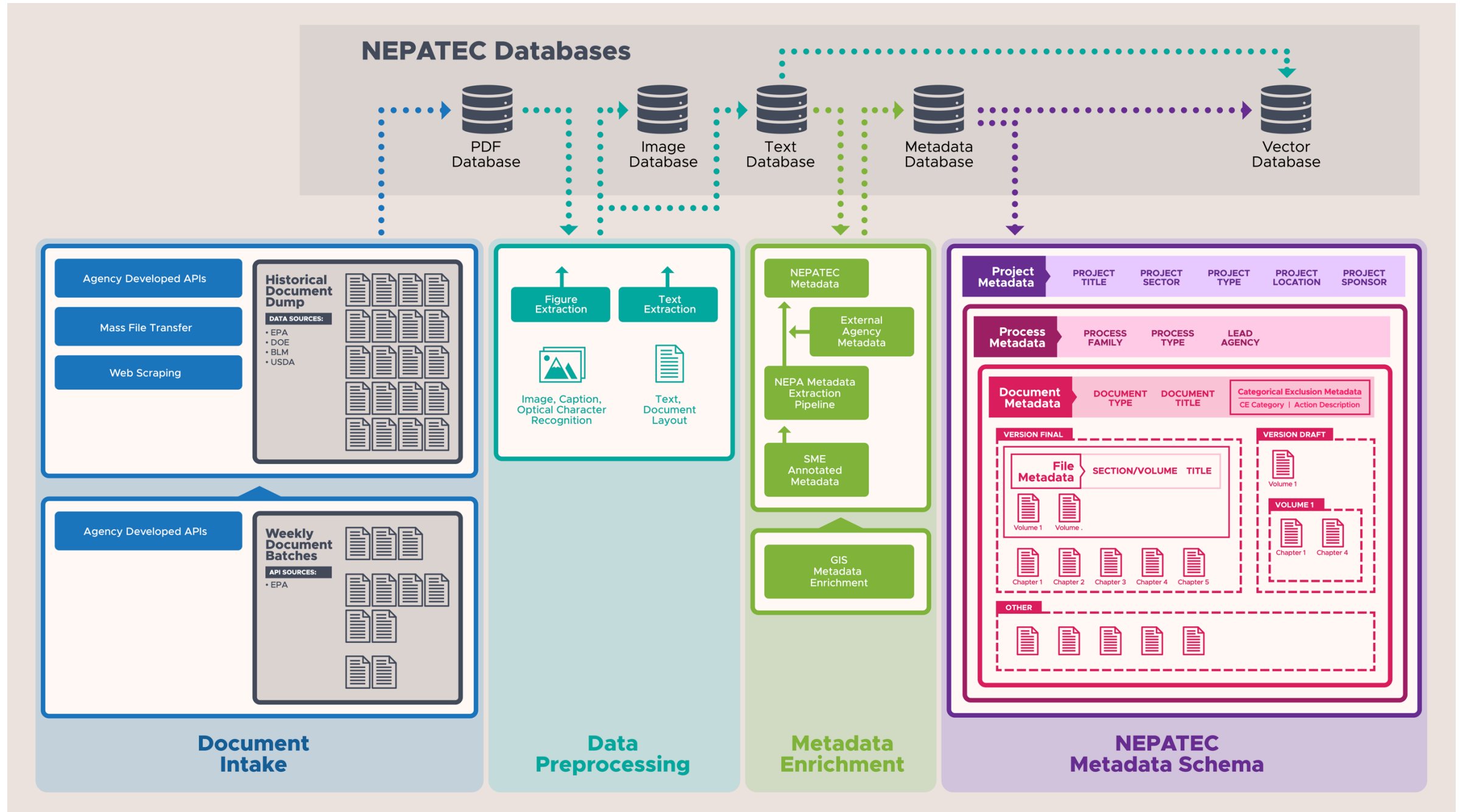
PermitAI Services

Accessible Structured Environmental Data



Service Consumers can access services (eventually) via APIs without necessarily going only through our native developed applications

NEPA Text Corpus (NEPATEC)





Pacific Northwest
NATIONAL LABORATORY

Rich Multimodal Data of the NEPA Documents

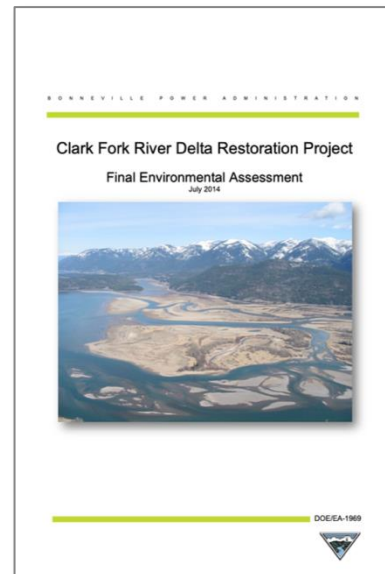
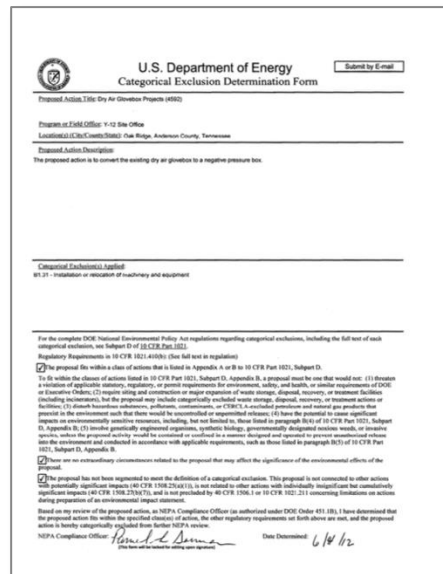
DOCUMENTS

140,000+ files

7 million pages

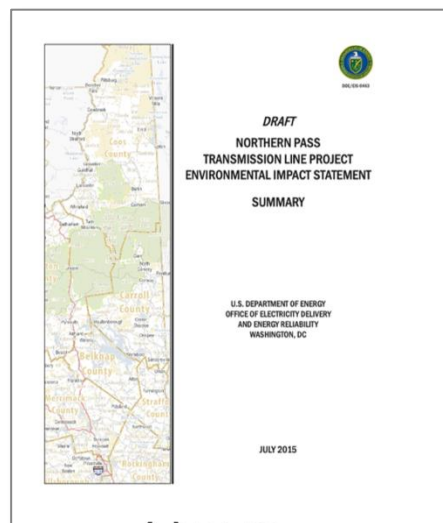
MAPS

1+ million

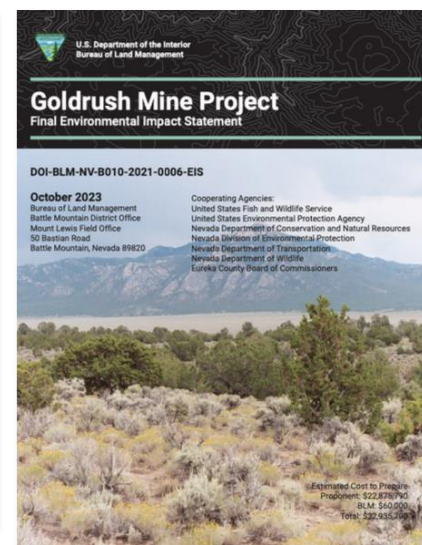


CE

EA



DEIS



FEIS

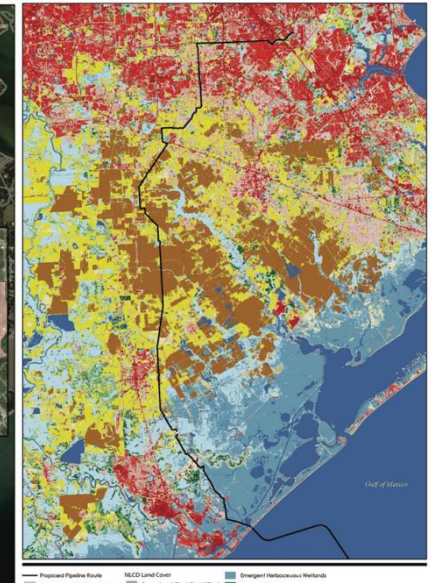
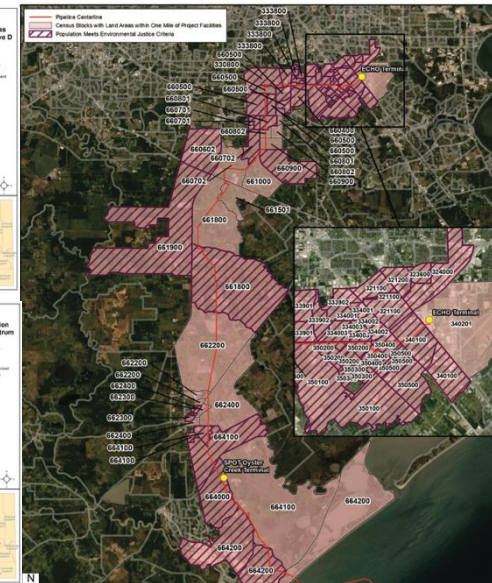
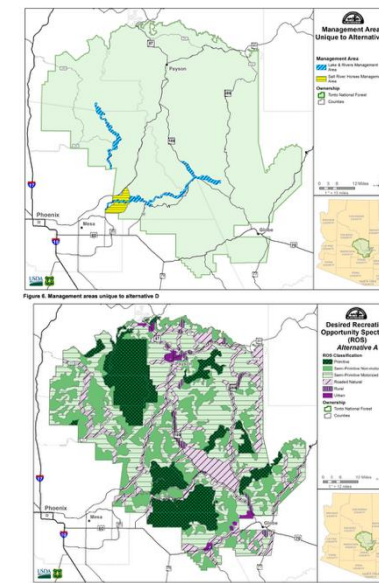


Figure 2.2-1: Aerial View of the Existing ECHO Terminal and Proposed Modifications

Figure 3.13-1. Historic Properties

Figure 3.13-2. Historic Viewshed Analysis

NEPATEC Releases

NEPATECv3.0

ONGOING

Standardized 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
- Multipoint Document Summary

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

NEPATECv2.0 Document Intake & Statistics

NEPATEC Dataset

 **60+**
AGENCIES

 **61,881**
TOTAL PROJECTS

 **142,083**
TOTAL FILES

 **6,967,739**
TOTAL PAGES

Categorical Exclusion (CE)

 **54,668**
PROJECTS

 **73,544**
FILES

 **366,876**
PAGES

Environmental Assessment (EA)

 **3,083**
PROJECTS

 **14,242**
FILES

 **469,106**
PAGES

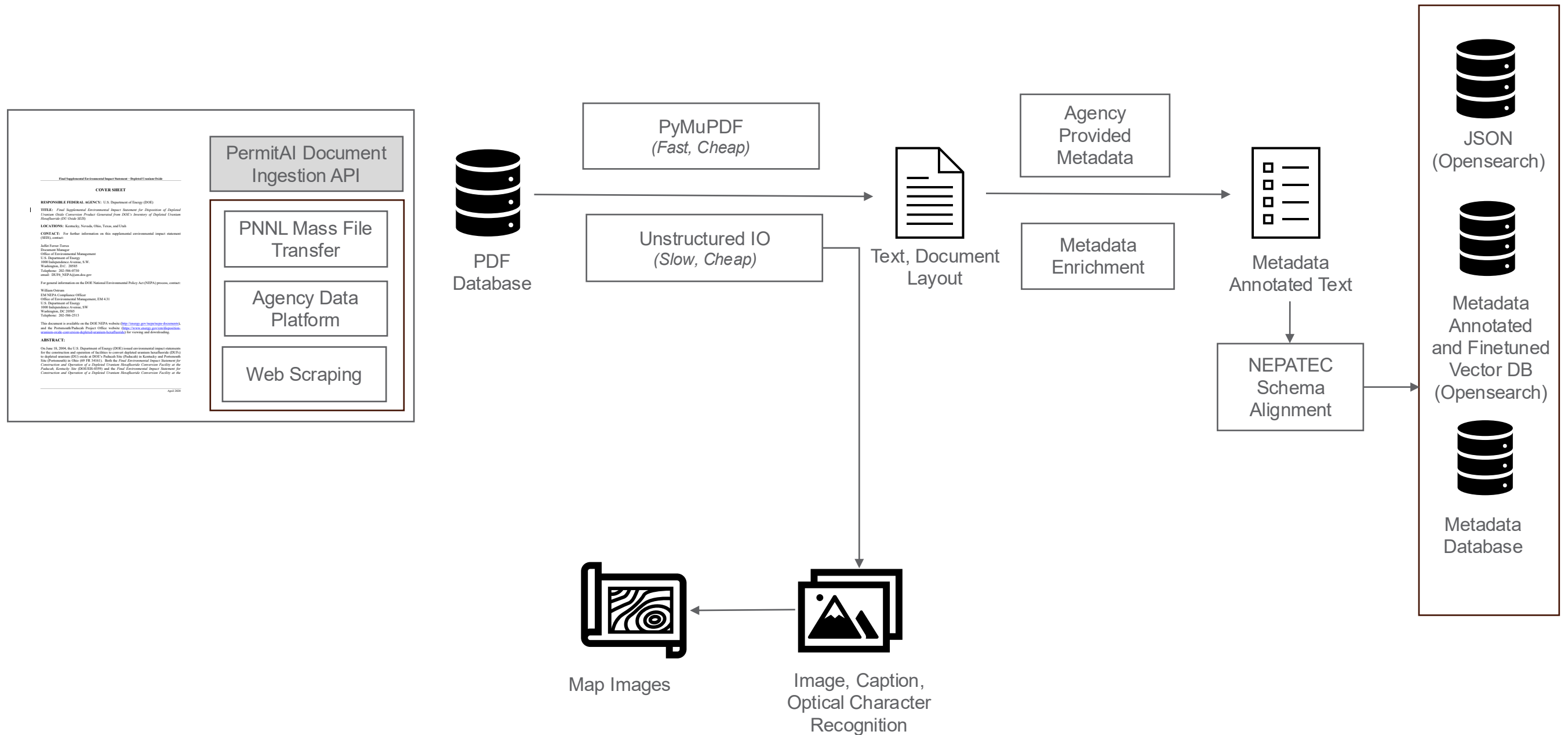
Environmental Impact Statement (EIS)

 **4,130**
PROJECTS

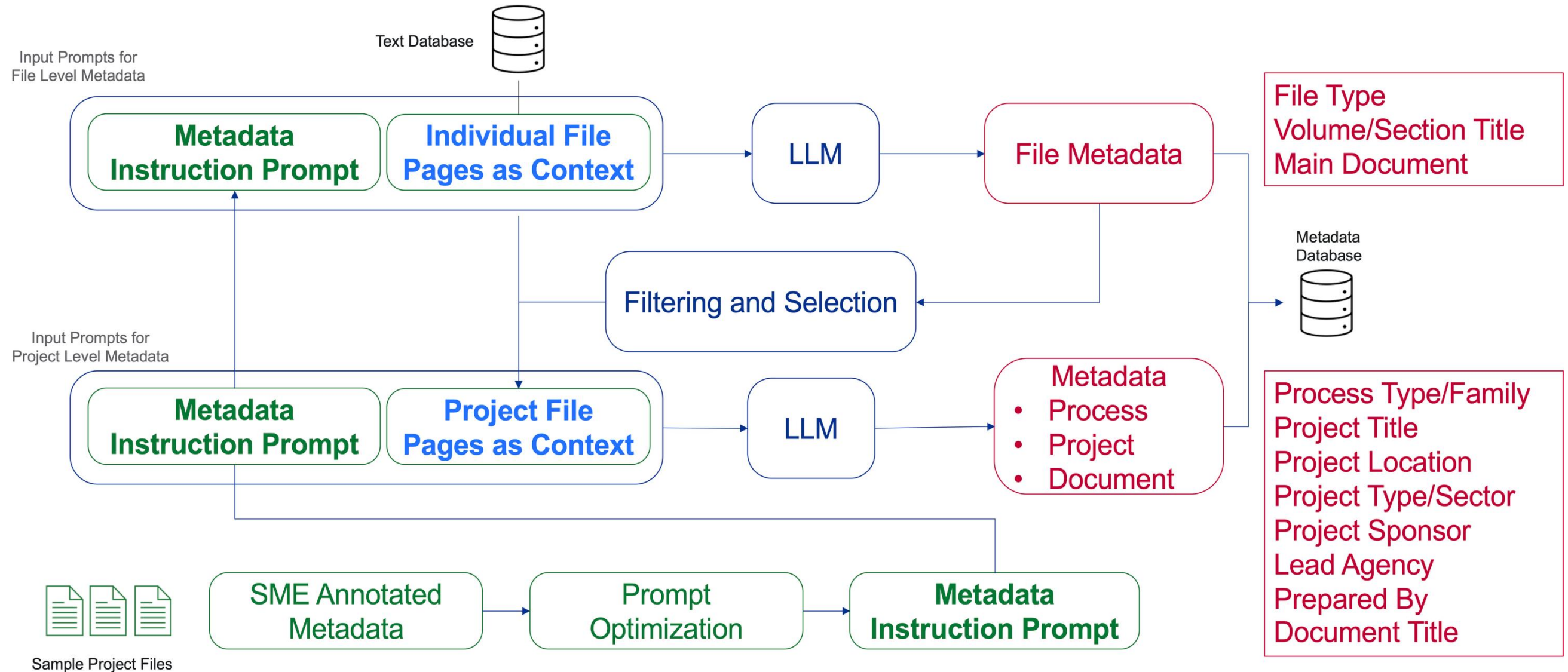
 **54,297**
FILES

 **6,131,757**
PAGES

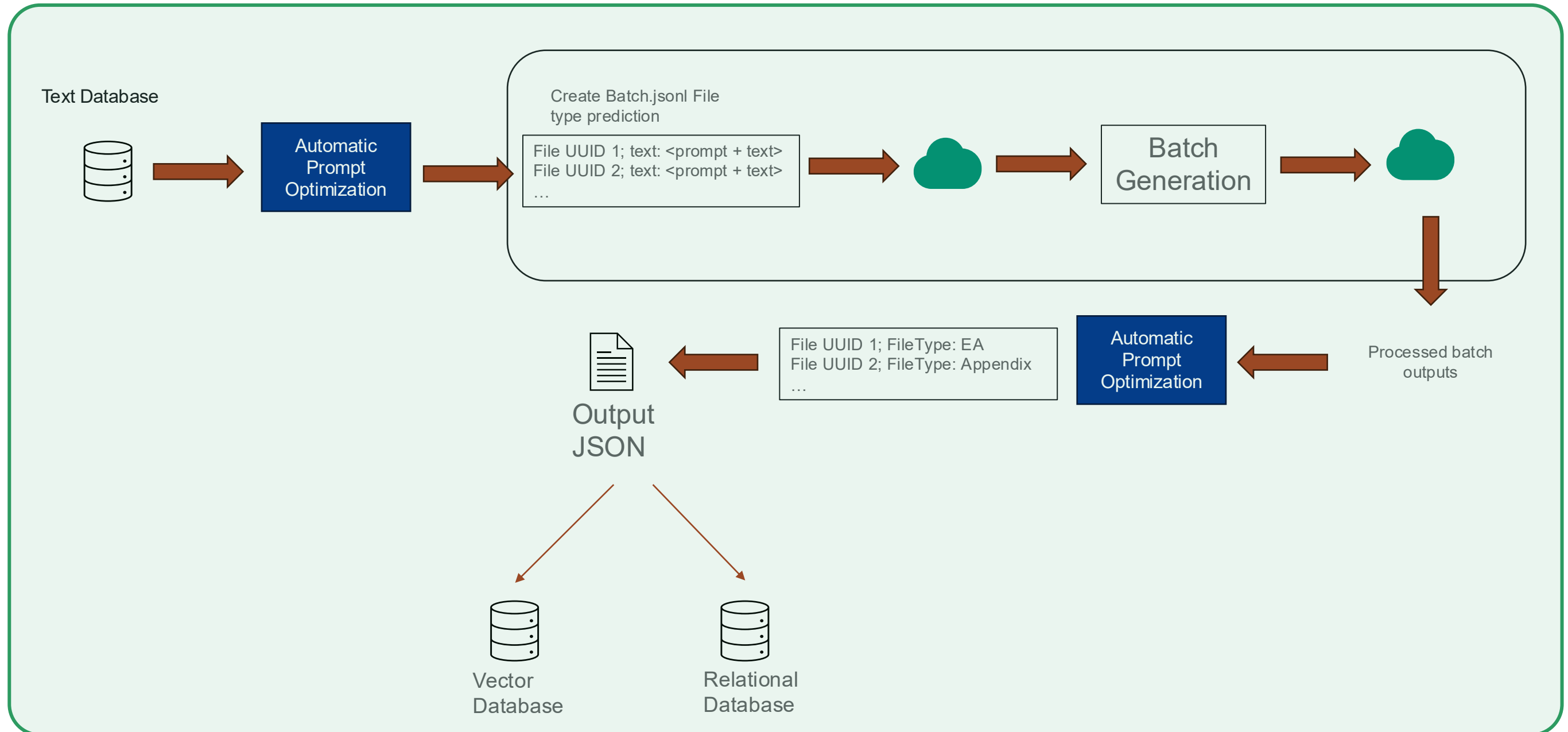
Data Ingestion and Processing Pipeline



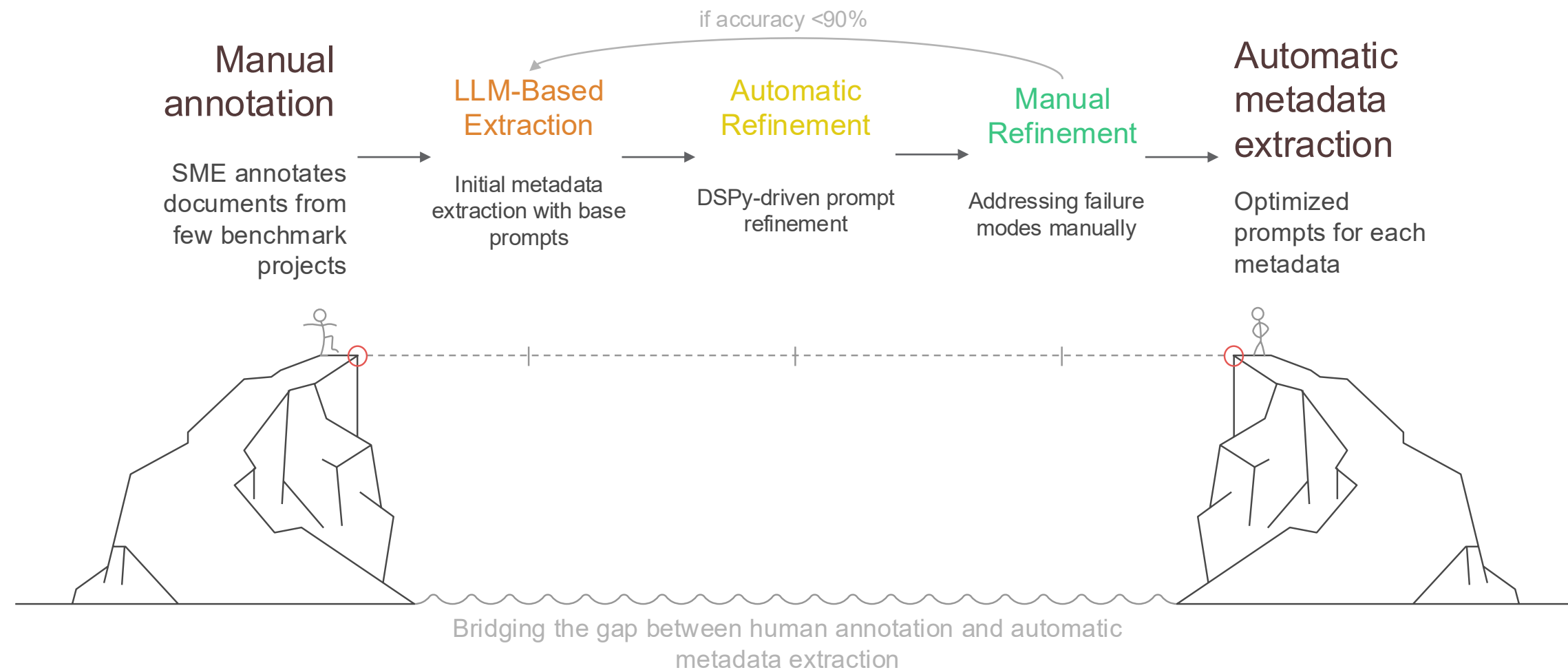
Text-Centric Metadata Extraction and Enrichment



Text-Centric Metadata Extraction and Enrichment



Iterative Prompt Refinement Using SME Feedback and Prompt Optimization



Prompt Refinement Example: Project Title

Base prompt

Accuracy: 27%

What is the title of this project? To assist in this determination, consider that a project title is a descriptive name for a project. Project titles are often written on the cover page or title page of a document, but are not always explicitly labeled. A project title should not include the type of document (e.g., categorical exclusions, environmental assessment, environmental impact statement). If a project title cannot be clearly identified from the source document, you may infer a title, typically based on the proposed action described in the document.

Multiple rounds of automatic refinement using DSPy

Input: base prompt + annotated datasets
Output: improved prompts

Added role

Improved Prompt 2

Accuracy: 30%

You are an expert document analyst. Your task is to identify the project title from the provided text. To assist in this determination, consider that a project title is a descriptive name for a project. Project titles are often written on the cover page or title page of a document, but are not always explicitly labeled. A project title should not include the type of document (e.g., categorical exclusions, environmental assessment, environmental impact statement). If a project title cannot be clearly identified from the source document, you may infer a title, typically based on the proposed action described in the document.

Multiple rounds of manual refinement focusing on the failure modes

Continued on next slide

Prompt Refinement Example

Added protection
against failure modes

Improved Prompt 6

Accuracy: 62%

You are a NEPA Document Specialist and senior analyst for Federal environmental documents. You have deep expertise in identifying official project titles in Environmental Impact Statements (EIS) and related NEPA materials. You will be given an excerpt from an EIS or other NEPA document. Project titles in these documents usually appear on the cover page, title page, or a clearly marked header—but they are not always explicitly labeled. A true “Project Title” is a concise, descriptive name for the proposed action itself and should never include document classification terms such as “Environmental Impact Statement,” “EIS,” “Draft,” “Final,” “Environmental Assessment,” “EA,” “Report,” or “Study.”

Instructions:

1. Scan the text for an explicit, standalone name that represents the project or proposed action.
2. If you find one, extract it verbatim (minus any document-type labels).
3. If no explicit title is present, infer the most accurate, concise title based on the primary scope and action described.
4. Exclude any appended or prepended classification words (EIS, EA, Draft, Final, Appendix, etc.).
5. Return only the project title as a single line of plain text—no additional commentary, formatting, or labels.

Multiple rounds of both automatic and manual refinement

Added concise instructions

Improved Prompt 8

Accuracy: 91%

You are an environmental document specialist. Your task is to identify and extract the precise, formal project title of the provided Environmental Impact Statement. A project title is a descriptive name for a project. Project titles may be written on the cover page or title page of documents available for this project, but are not always explicitly labeled.

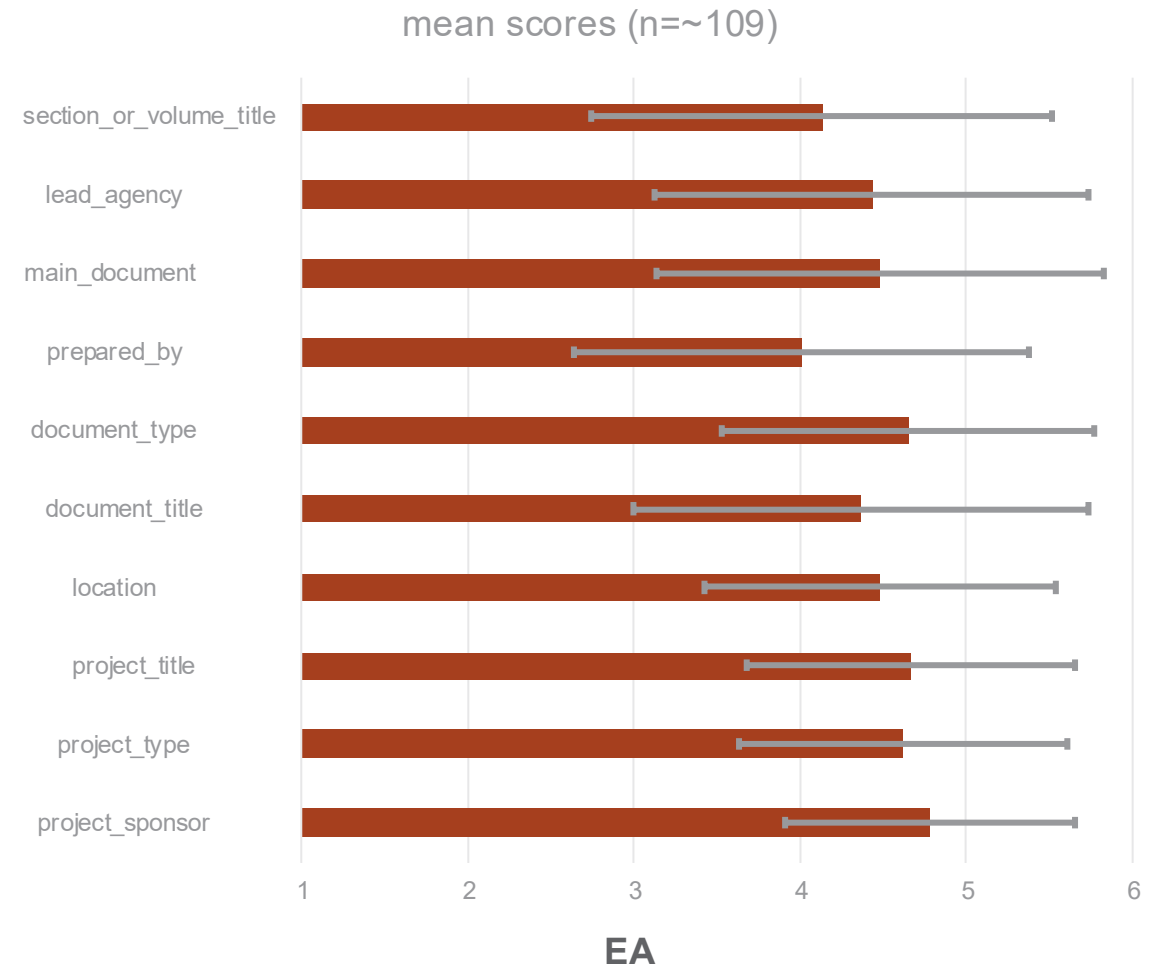
Instructions:

1. Ensure the extracted project title does not include any file names, volume numbers, or internal section titles.
2. A project title should be revised to omit the type of document (e.g., categorical exclusions, environmental assessment, environmental impact statement), if present.
3. Do not have abbreviations in the project title, use their full version.
4. If you find multiple acceptable titles at different places, return the one that appears earlier in the project text.
5. If a project title cannot be clearly identified from the available documents, you may infer a title, typically based on the proposed action described in one or more of the documents (preferentially the most recent).

Validating Prompts Refinement on SME Annotated Dataset

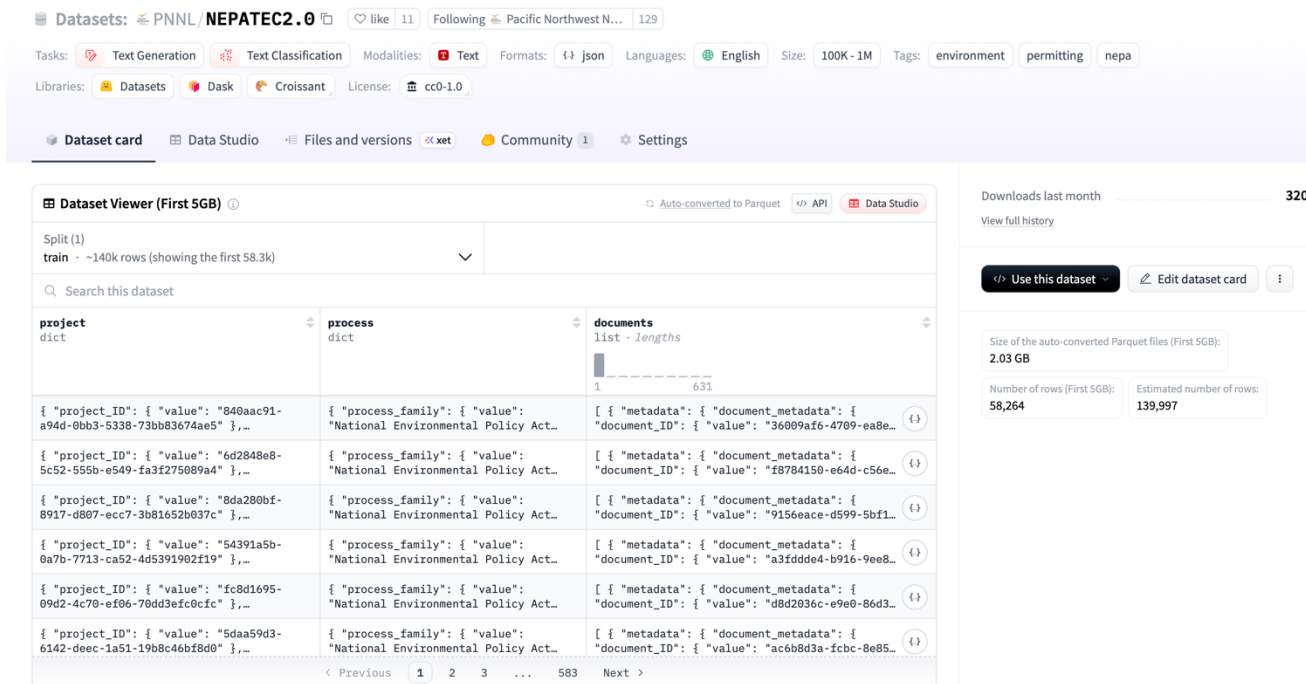
	Environmental Impact Statements	Environmental Assessments
Metadata Attribute	Accuracy (#correct/#total)	Accuracy (#correct/#total)
project_title	91% (111/122)	100% (12/12)
project_type	90.83% (99/109)	91.67% (11/12)
project_sponsor	91.67% (110/120)	91.67% (11/12)
location	95.00% (112/118)	91.67% (11/12)
lead_agency	92.31% (96/104)	91.67% (11/12)
document_title	96.92% (63/65)	91.67% (11/12)
document_type	96.92% (283/292)	100% (12/12)
prepared_by	100% (11/11)	91.67% (11/12)
volume_or_section_title	81.25% (78/96)	91.67% (11/12)
main_document	98.99% (98/99)	100% (12/12)

Validating Extracted Metadata through SME Evaluation



Score	Interpretation	Description
1	Unusable	Mostly incorrect or missing, preventing downstream use.
2	Poor	Major errors or omissions; partial utility only.
3	Acceptable	Correct for key fields, with minor issues.
4	Good	Accurate and complete; slight formatting inconsistencies.
5	Excellent	Fully accurate, complete, and formatted ideally.

NEPATEC Public Release



Dataset Viewer (First 5GB) | Auto-converted to Parquet | API | Data Studio

Split (1)
train · ~140k rows (showing the first 58.3k)

Search this dataset

project	process	documents
{ "project_ID": { "value": "840aac91-a94d-0bb3-5338-73bb83674ae5" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "36009af6-4709-ea8e..." } } }]
{ "project_ID": { "value": "6d2848e8-5c52-555b-e549-fa3f275089a4" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "f8784150-e64d-c56e..." } } }]
{ "project_ID": { "value": "8da280bf-8917-d807-ecc7-3b81652b037c" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "9156eace-d599-5bf1..." } } }]
{ "project_ID": { "value": "54391a5b-0a7b-7713-ca52-4d5391902f19" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "a3fddde4-b916-9ee8..." } } }]
{ "project_ID": { "value": "fcd81695-09d2-4c70-ef96-70dd3efc0cfc" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "d8d2036c-e9e0-86d3..." } } }]
{ "project_ID": { "value": "5daa59d3-6142-deec-1a51-19b8c46bf8d0" }, ... }	{ "process_family": { "value": "National Environmental Policy Act..." }	[{ "metadata": { "document_metadata": { "document_ID": { "value": "ac6b8d3a-fcbc-8e85..." } } }]

Downloads last month: 320

Size of the auto-converted Parquet files (First 5GB): 2.03 GB

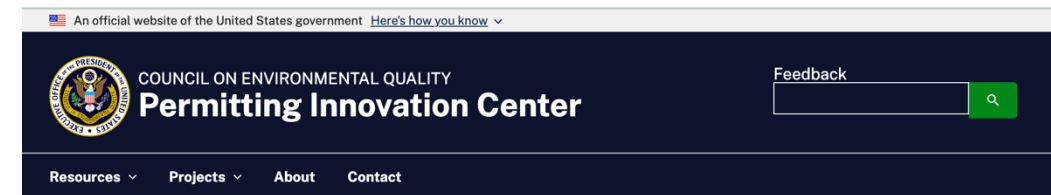
Number of rows (First 5GB): 58,264 | Estimated number of rows: 139,997

[HuggingFace](https://huggingface.co/datasets/PNLL/NEPATEC2.0)

(<https://huggingface.co/datasets/PNLL/NEPATEC2.0>)

[Technical Report](https://www.osti.gov/biblio/2584716)

(<https://www.osti.gov/biblio/2584716>)



An official website of the United States government | Here's how you know

COUNCIL ON ENVIRONMENTAL QUALITY
Permitting Innovation Center

Feedback

Resources | Projects | About | Contact

Developer Tools

Developer Resources

NEPA and Permitting Data and Technology Standard

NEPA and Permitting Data and Technology Standards GitHub Resources

Example Swagger OpenAPI Page

CE Explorer - Categorical Exclusion Data

NEPATEC 2.0

OPEN API documentation

Developer Resources

The following tools provide information, data, or other support for developers working with environmental review and permitting systems.

NEPA and Permitting Data and Technology Standard

[The data standard](#) establishes a government-wide data and technology standard for NEPA, and other related permits and reviews. This data standard provides agencies with a common digital language to facilitate interoperability and automatic data exchange among systems. These pages provide an overview of the data standard, description of entities and their properties, and example relationship diagrams.

[CEQ's Permitting Innovation center – Developer Resources](https://permitting.innovation.gov/)

(<https://permitting.innovation.gov/>)

More talks from PermitAI

- 10:00 -10:30 AM
NEPA GIS: Digitizing Historic NEPA Layers Using GeoAI
Presenter: Anastasia Bernat

- 11:00 -11:30 AM
WriteNEPA: Human AI Collaboration Tool for Drafting and Auditing Environmental Review
Presenter: Sarthak Chaturvedi



**Pacific
Northwest**
NATIONAL LABORATORY

Thank you

kaustav.bhattacharjee@pnnl.gov
permitai@pnnl.gov

<https://www.pnnl.gov/projects/permitai>

*This document was cleared through PNNL's
Information Release as PNNL-SA-223037*