AI-Powered Exploration of Unique Multimodal Collections: Surfacing Knowledge for the Next Manhattan Project

Juliane Schneider, Sridevi Wagle, Miriam Blake, Sutanay Choudhury, Nicole Critchley, Olivia Hess, Martin Klein, Janice Scarano, William Zhang

AI-ready Data Transformation

- **Challenge:** Hanford's multimodal Manhattan Project archives contain critical scientific and environmental knowledge, but limited metadata and fragmented formats restrict discovery and analysis.
- **Solution:** An AI-powered multimodal search and chat platform combining semantic and fuzzy search, page-level indexing, and LLM-based metadata extraction to curate and explore 1.5 TB of archival text and images.
- **Impact:** Enables knowledge graphs and guided discovery that surface hidden technical ad environmental insights, while providing a scalable AI framework for unlocking historical scientific collections across DOE.

Dataset

- Size: 1.5 TB
- 4M TIFs (condensed to 115K PDFs)
- 70K images
- 1.3K PDFs

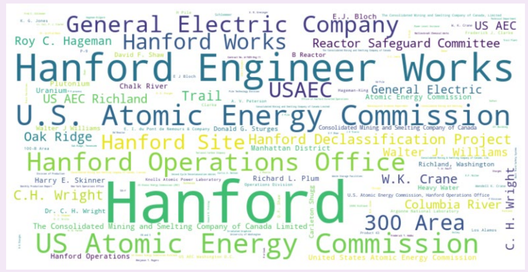

Figure 1: Word Cloud of AI-Generated PDF Metadata

Data Pipeline

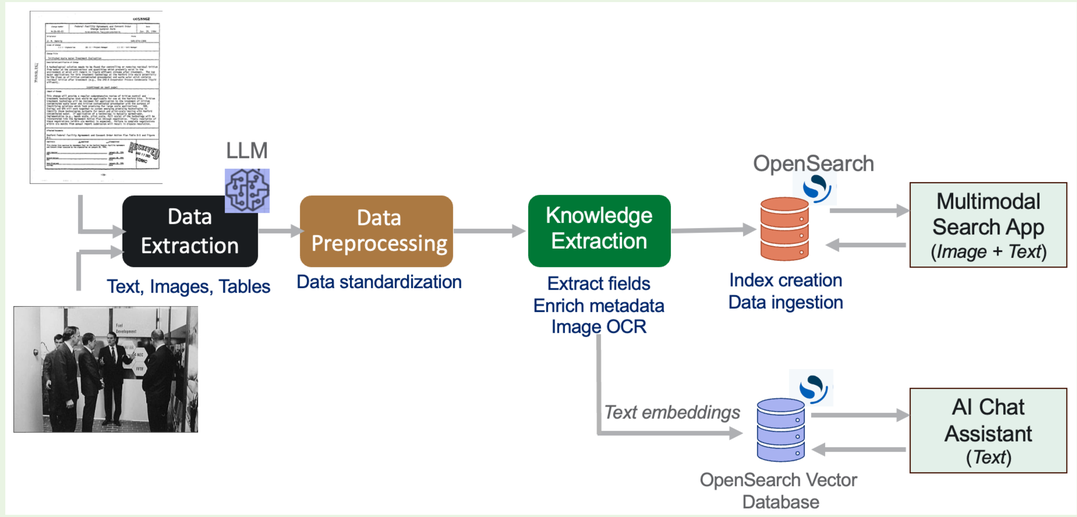

Figure 2: Multimodal Data Processing Pipeline

Multimodal Search App

Figure 3: Multimodal Search App Interface

- Multimodal Search App uses a basic NLP preprocessing pipeline for user queries and indexed text content with partial matching and fuzzy search.
- PDFs are indexed at page level allowing page specific search results and improved precision and contextual retrieval.

AI Chat Assistant

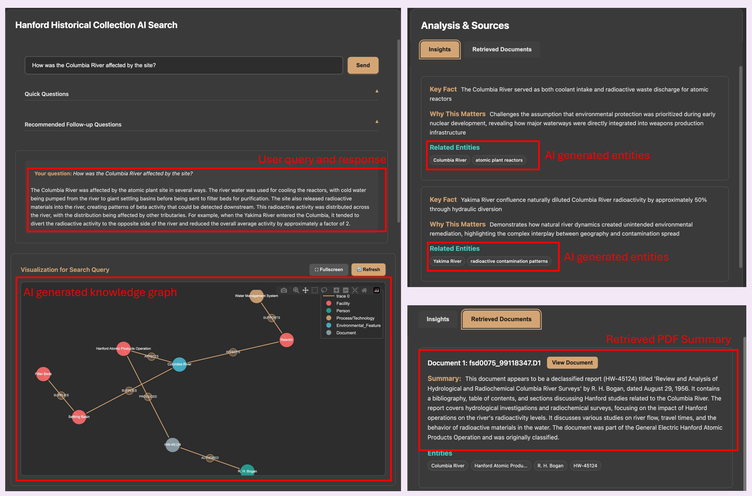

Figure 4: AI Chat Assistant

- Uses AWS Bedrock (Embed English v3) to generate 1024 dimensional embeddings for semantic search.
- Applies Claude Sonnet 3.5 to extract key entities from documents.
- Extracts organizations, projects, facilities, and individuals for deeper data insights.
- Identifies relationships and connections between extracted entities for better understanding of the data.

For additional information, contact: juliane.schneider@pnnl.gov

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

Pacific Northwest NATIONAL LABORATORY
www.pnnl.gov