

PDN2RDF: RDF Knowledge Graph Construction

Panayiotis Smeros, Imane Lboukili, David Yuan, Douglas Fils

NIAID-BRC AI Codeathon

November 12-14, 2025

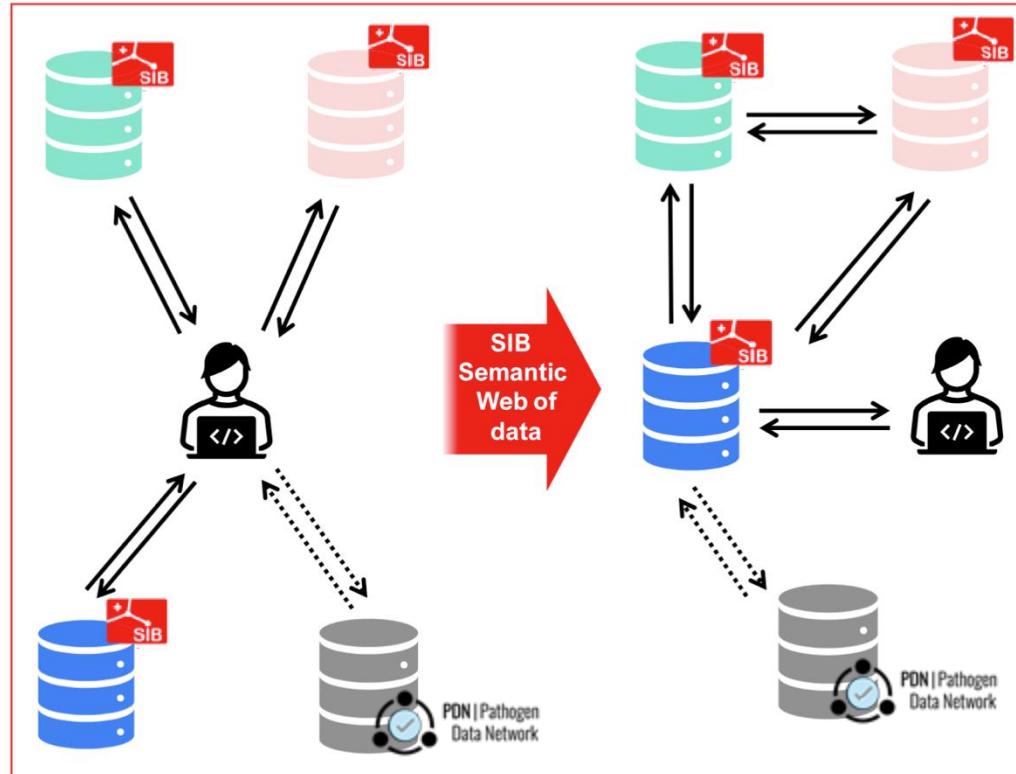


National Institutes
of Health



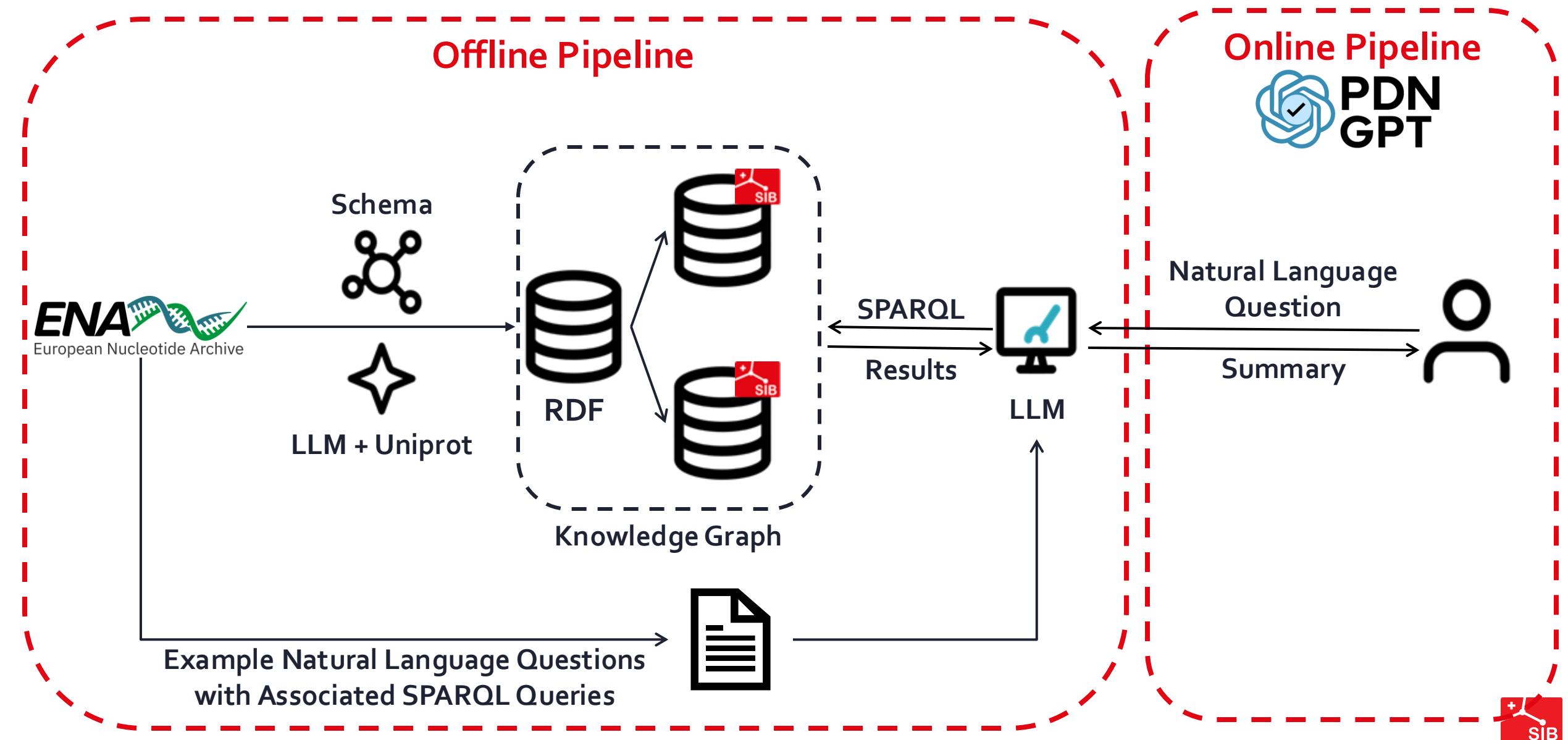
PDN | Pathogen
Data Network

PDN2RDF – Overall goal



- Enhance the interoperability of Pathogen Data Network (PDN) resources within the Linked Data ecosystem of the SIB
 - Improved data accessibility and interoperability
 - Streamlined data harmonization

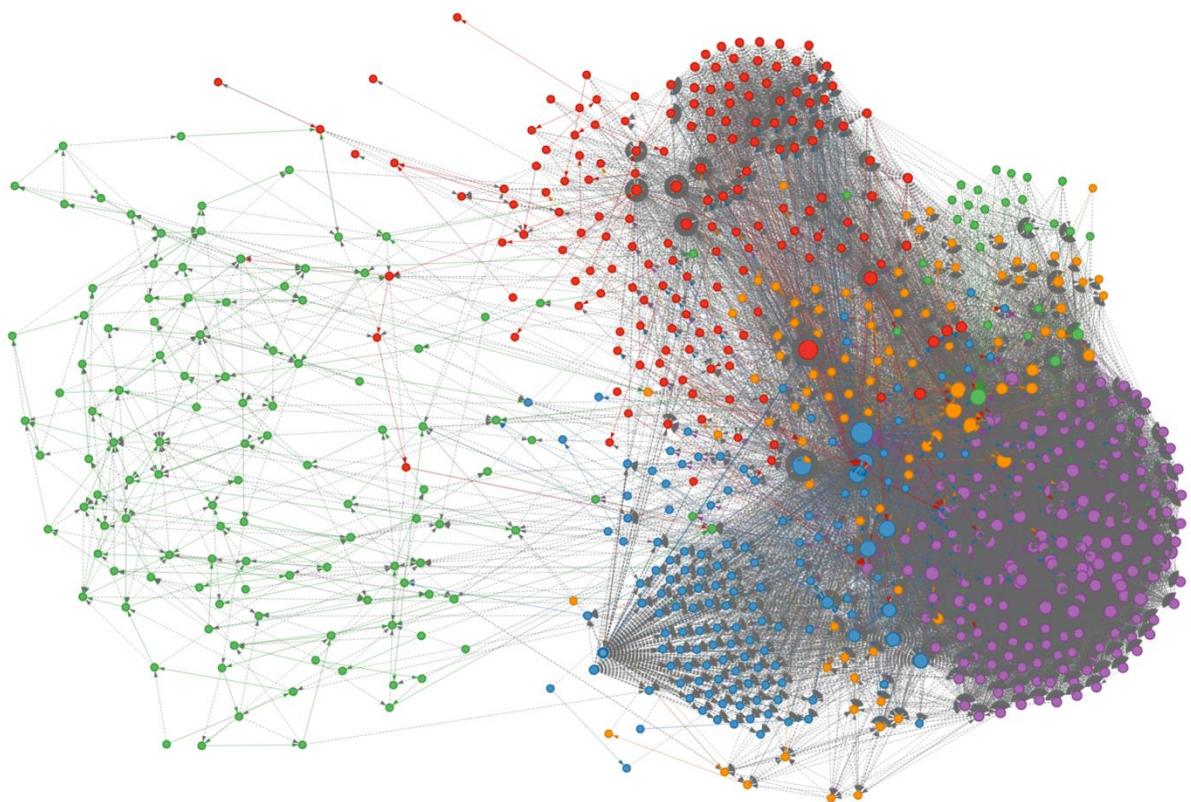
PDN2RDF – Workflow



PDN2RDF – Outcome

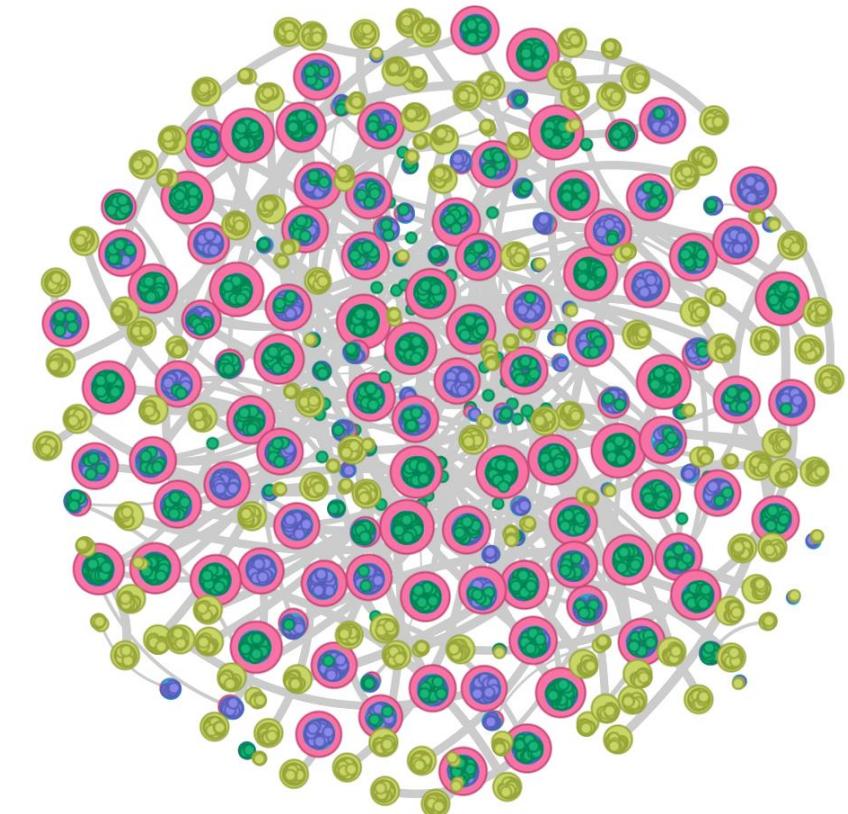
Before

Knowledge Graph of Monkeypox data from Japan from ENA.



After

PDN2RDF Knowledge Graph for the 100 first hits of the 198 PDN priority pathogens.



PDN2RDF - Our objectives for the codeathon

OFFLINE PIPELINE

1. Example queries and use-cases

Propose research questions and related SPARQL queries demonstrating extractable insights



2. Curated schema profile informing the choice of relevant SIB resources

Enable cross-resource integration



3. Prototype RDF transformer powered by LLMs

Develop an LLM-based system for automated RDF triple generation



4. ONLINE PIPELINE



DEMO TIME!

Future work

-  Incorporate additional PDN data sources to improve coverage and insights
-  Optimize process to limit data duplication
-  Ensure schema coverage and align with established schemas
-  Curate additional examples (NL Queries and associated SPARQL Queries)
-  Turn the online pipeline into an MCP server for improved utility

Questions ?