

PDN₂RDF: RDF Knowledge Graph Construction

Panayiotis Smeros, Imane Lboukili, David Yuan, Douglas Fils

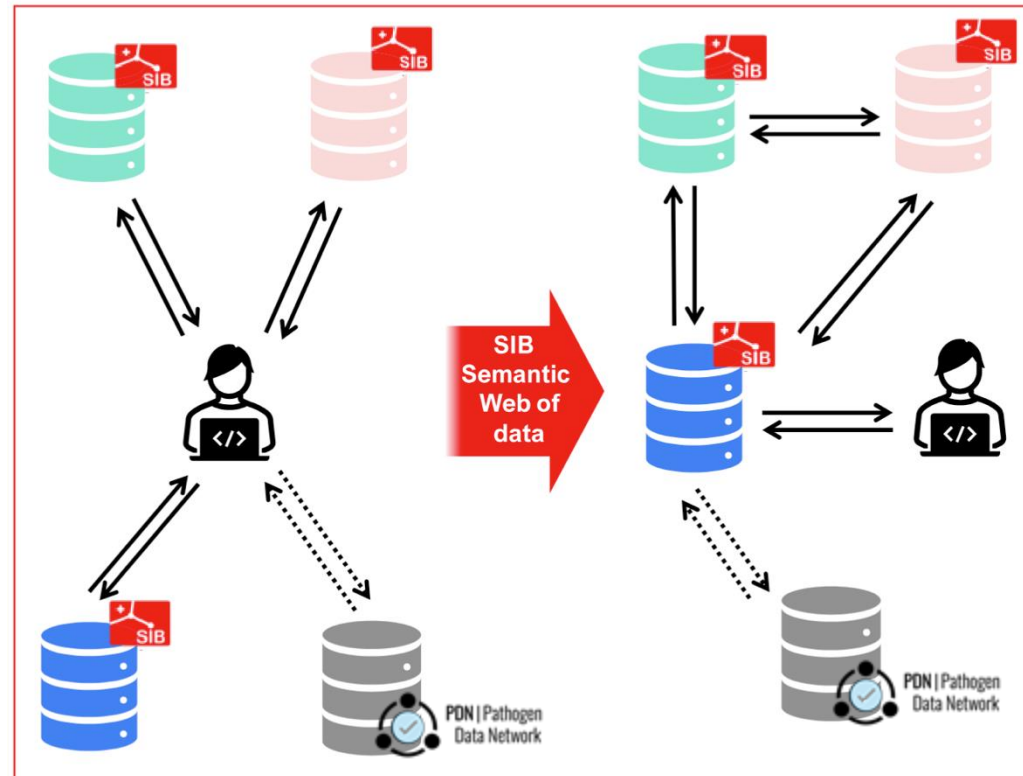
NIAID-BRC AI Codeathon

November 12-14, 2025



PDN | Pathogen
Data Network

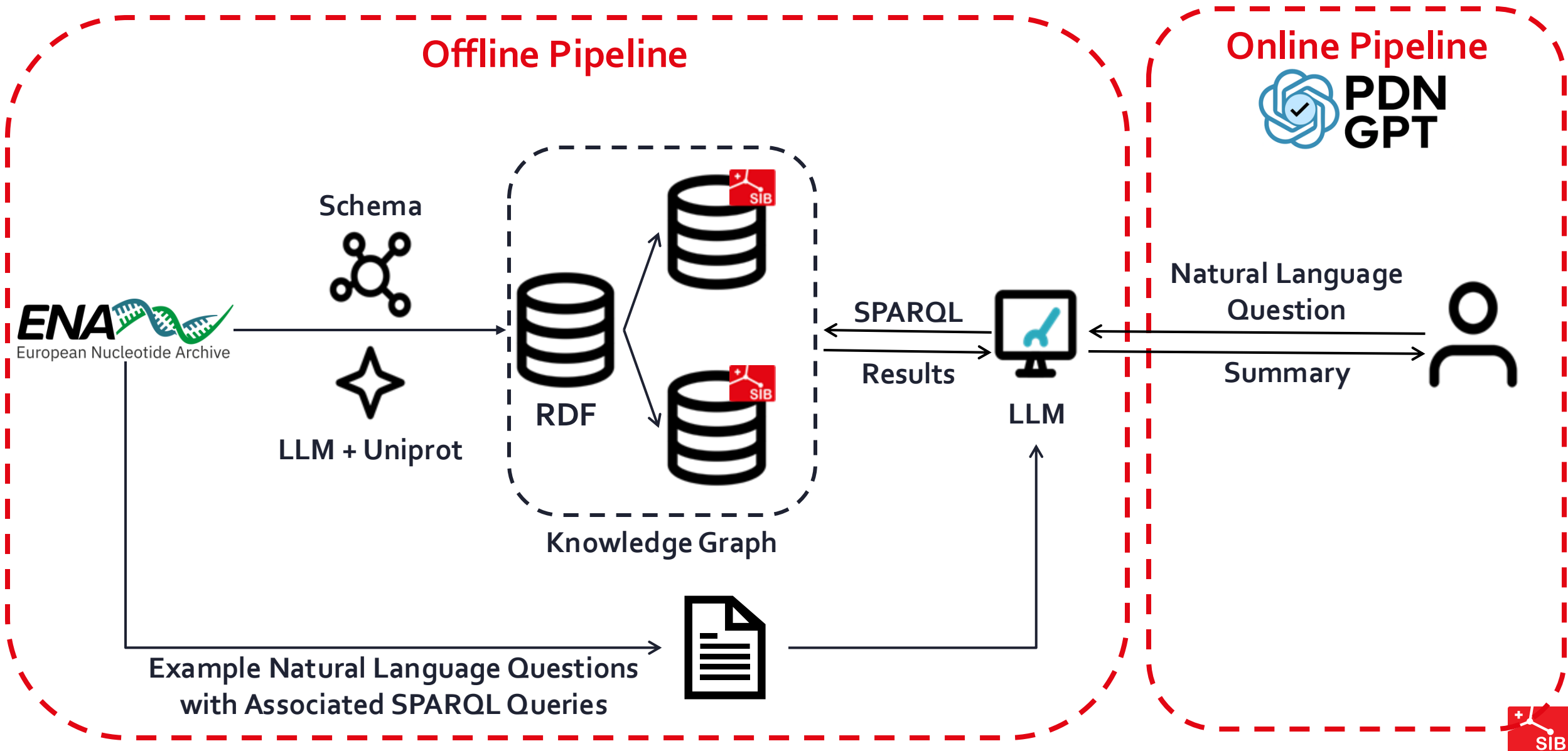
PDN₂RDF – Overall goal



<https://academic.oup.com/nar/article/52/D1/D44/7329473>

- 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

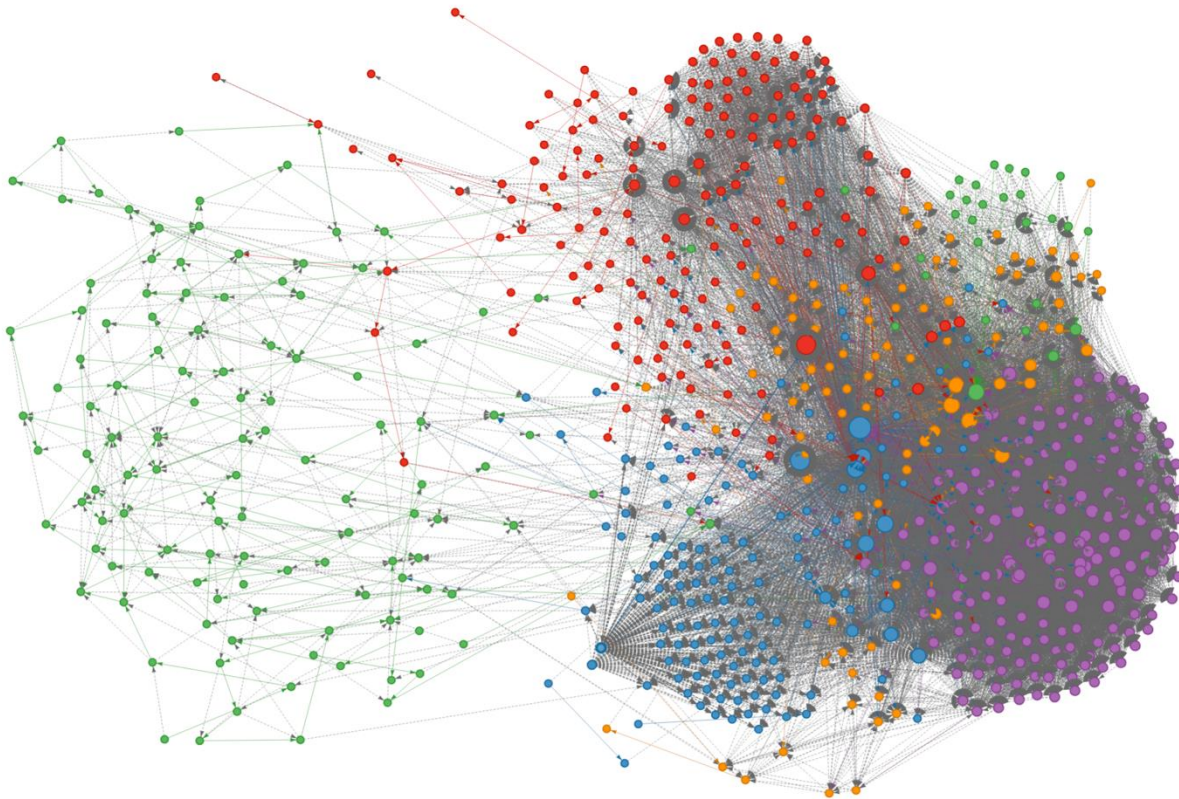
PDN₂RDF – Workflow



PDN2RDF – Outcome

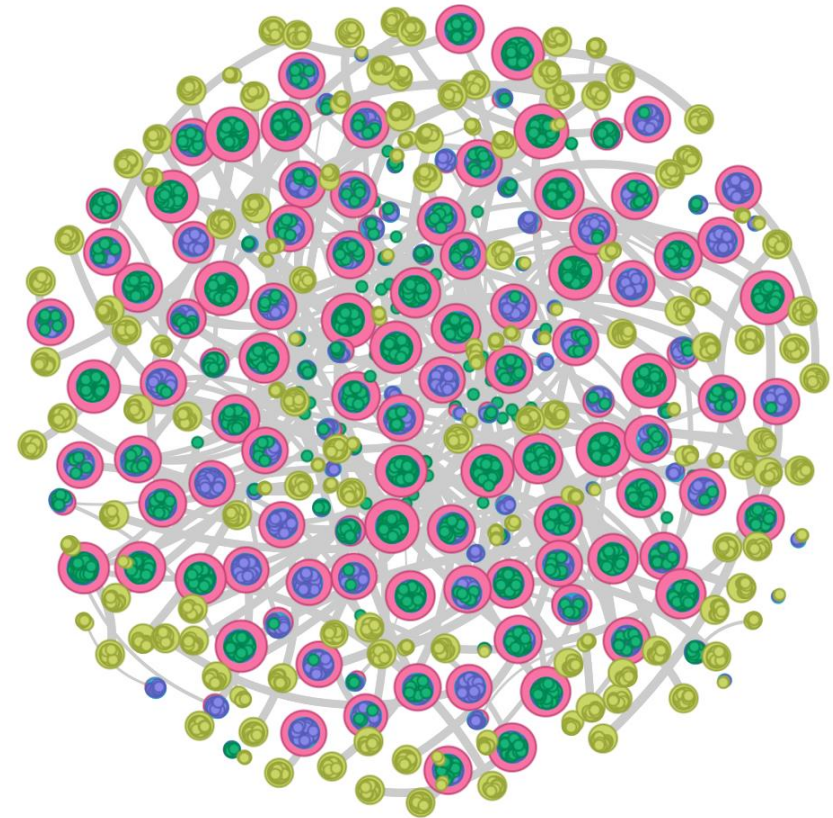
Before

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

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



OFFLINE PIPELINE

DEMO TIME!

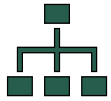
Future work



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 ?