

GRAPH-BASED ARCHIVAL DESCRIPTION

The Archives of Ontario's experience with Records
in Contexts

Aaron Hope and Pavel Zhelnov

ABOUT THE PROJECT

- **Objective:** a proof of concept to explore the possibilities and potential benefits of a graph-based data model for archival description, using the ICA's Records in Contexts standard
- **Timeline:** August 2023 - present
- **Project Support:** GLAM Incubator and SSHRC

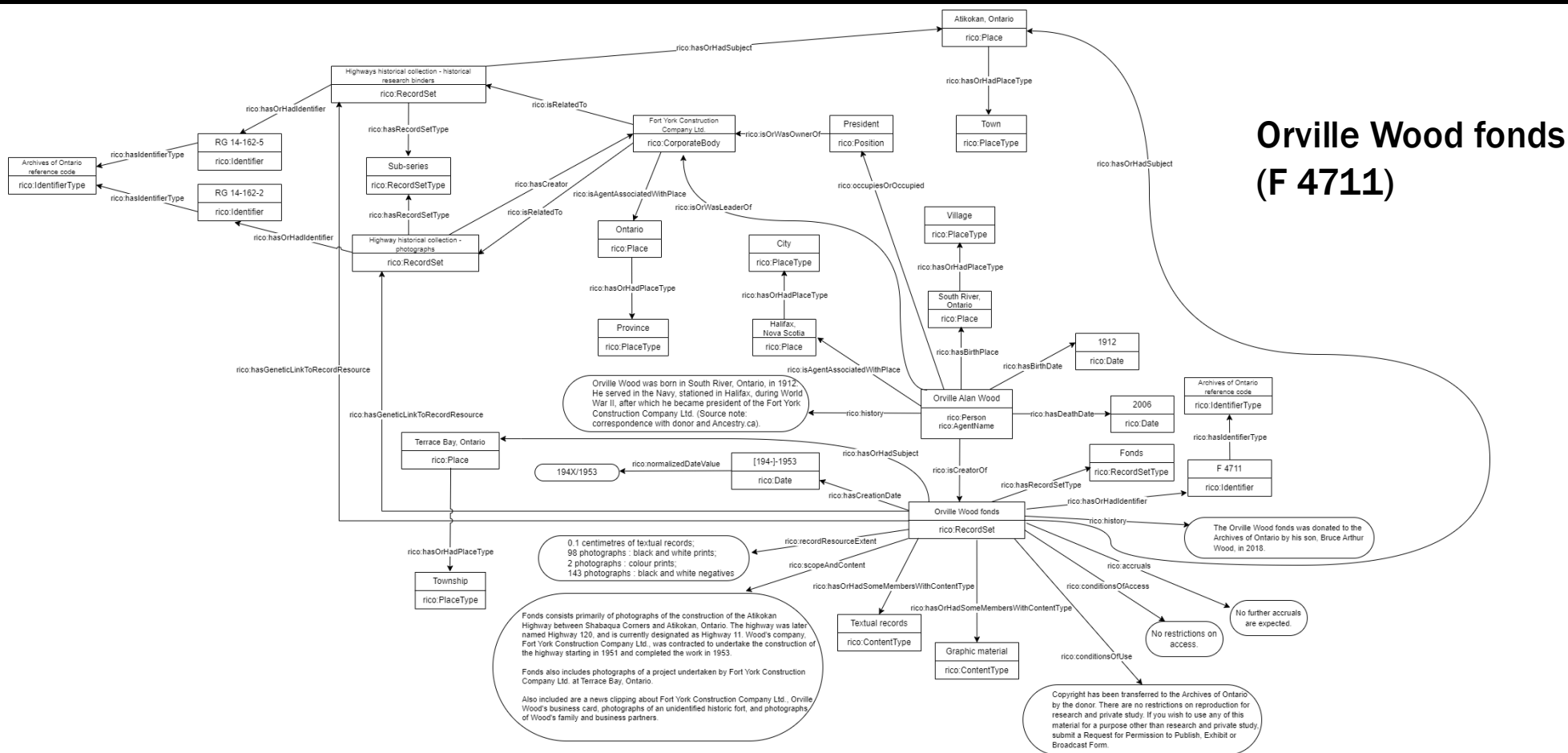
GLAM INCUBATOR

- A research and support hub of the University of Toronto's Faculty of Information
- Connects galleries, libraries, archives and museums with expertise and resources (at U of T and beyond)
- **New Call for 2025-2026 Projects:** <https://glam.ischool.utoronto.ca>
- **Deadline:** August 1, 2025

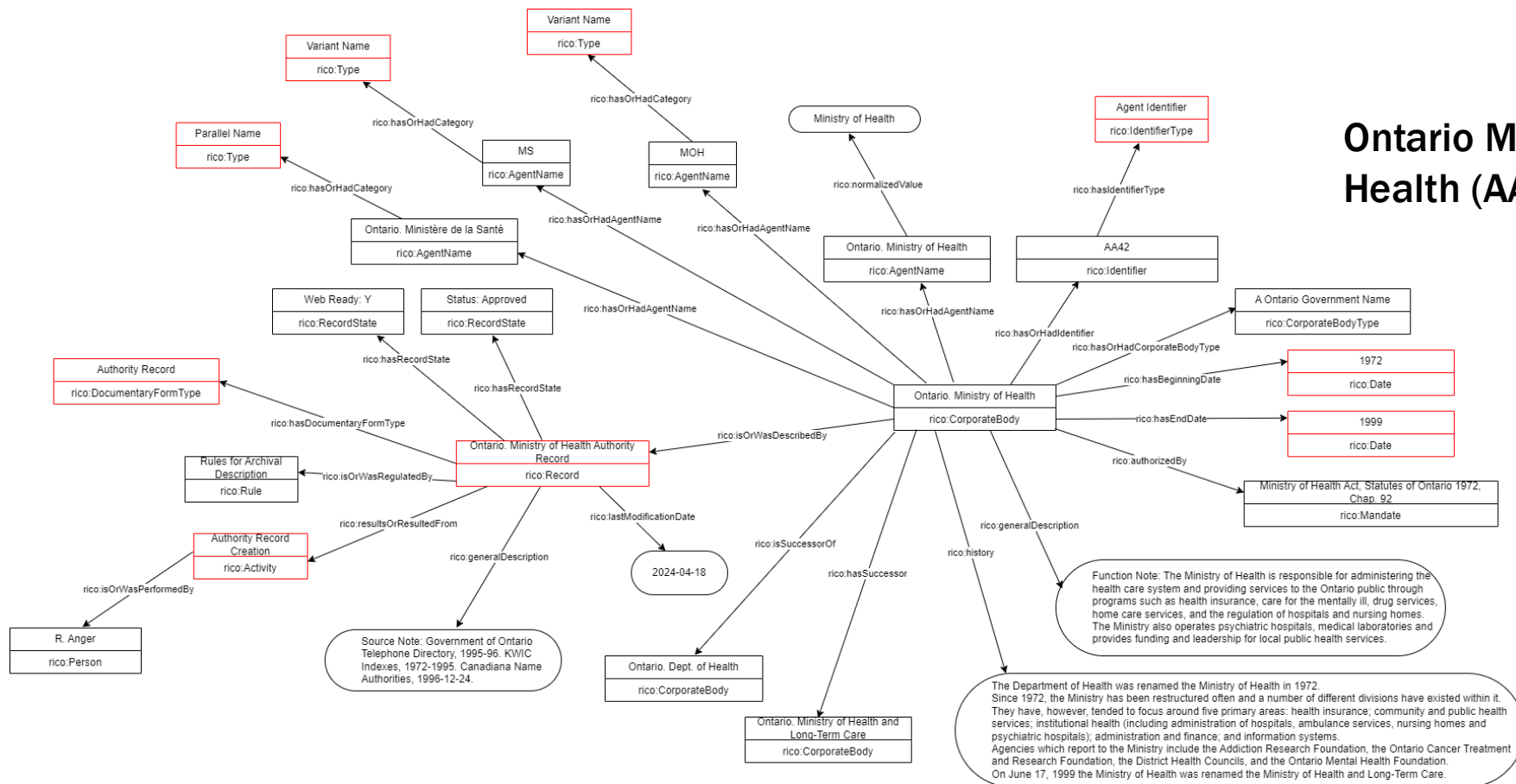
DATA MAPPING: PROCESS

- Mapping tables established crosswalks to RiC-CM and RiC-O
 - 57 archival description data elements
 - 22 authority record data elements
- RiC-O Shape Library for draw.io (diagrams.net)
 - Developed by Richard Williamson (EGAD, National Archives of Norway)
 - https://github.com/williamsonrichard/records_in_contexts_draw_io_shape_library

DATA MAPPING: PROCESS

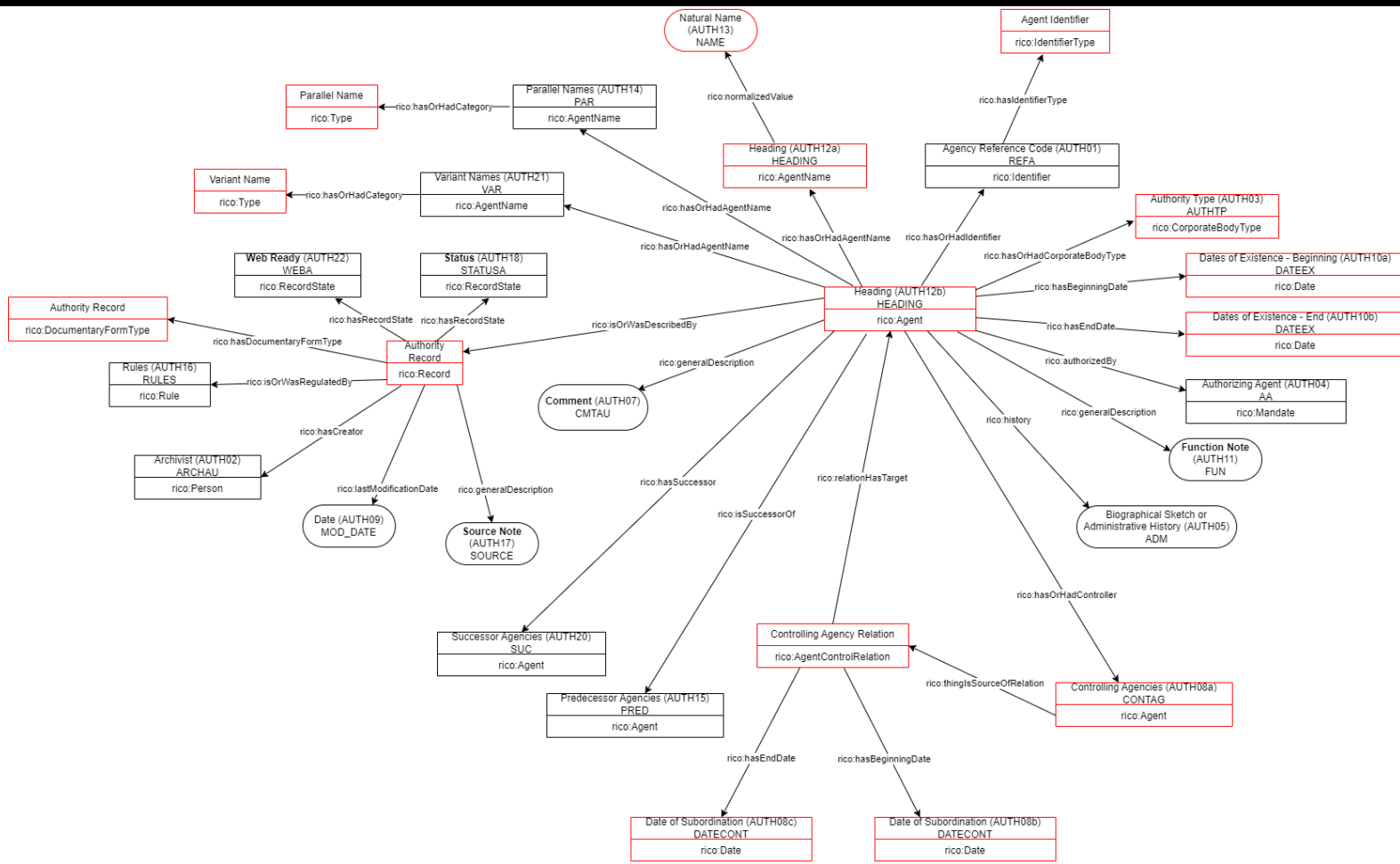


DATA MAPPING: PROCESS



Ontario Ministry of Health (AA42)

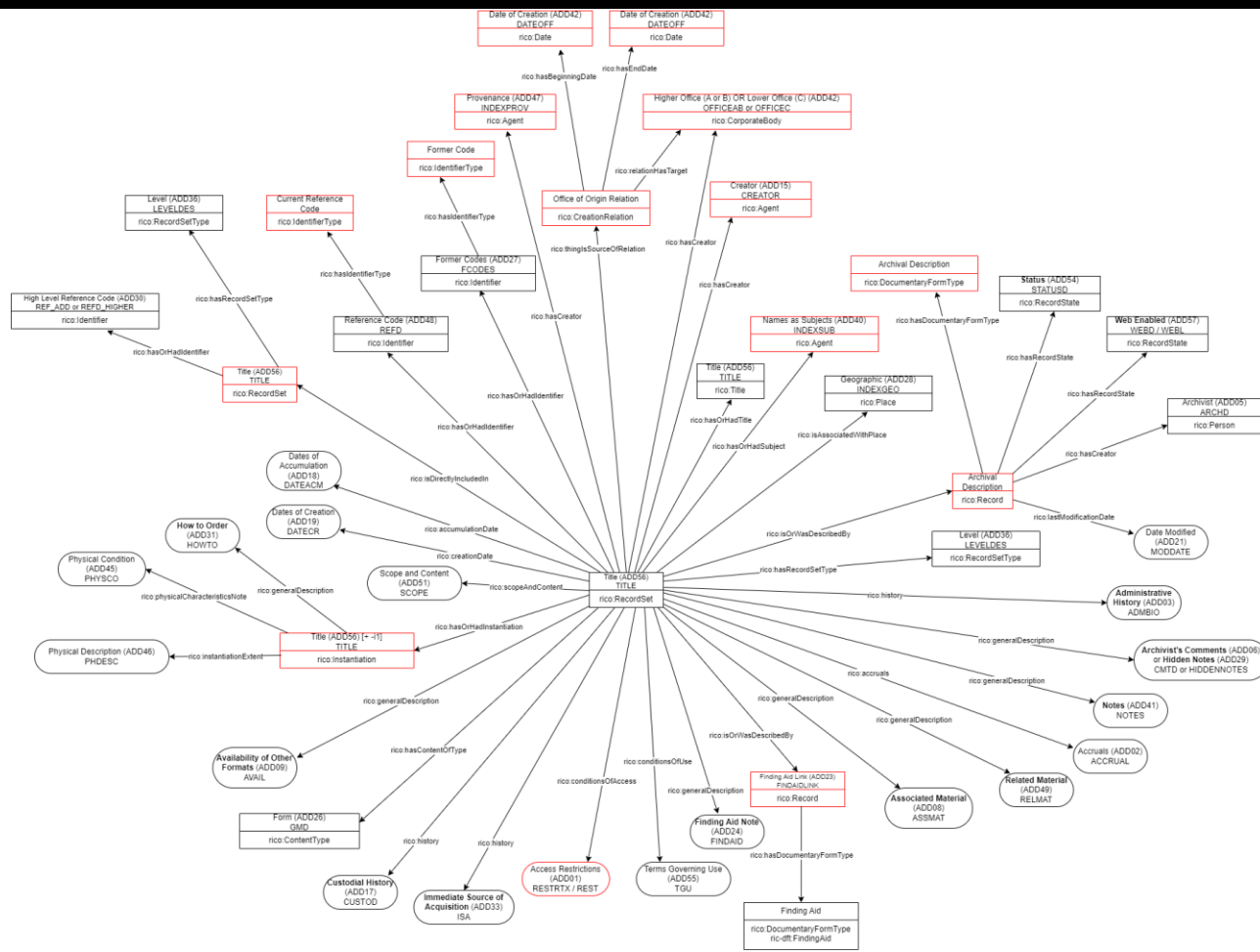
DATA MAPPING: PROCESS



General Authorities Model

DATA MAPPING: PROCESS

General Descriptions Model



DATA MAPPING: FINDINGS

- Most data elements could be translated successfully to RiC-O
- Some challenges:
 - Legacy data structures caused reliance on generic properties
 - Some descriptive data is metadata about the description
 - Some desired properties were not available (but are being added)

DATA CONVERSION



Step	Script Name	Input	Output
1	draw_io_parser.py	*.drawio	*.owl, *.ttl (intermediary map)
2	map_schema.py	*.ttl, *.csv (Excel, SQL)	*.rml (RDF Mapping Language), *.csv (preprocessed)
3	map_rml.py	*.rml, *.csv (preprocessed)	*.ttl, *.trig (final graph)

Result

Description and Authorities datasets = ca. one million triple statements (plus more inferred)
The Implementation Toolkit (early release) available from: <https://github.com/gbad-project>

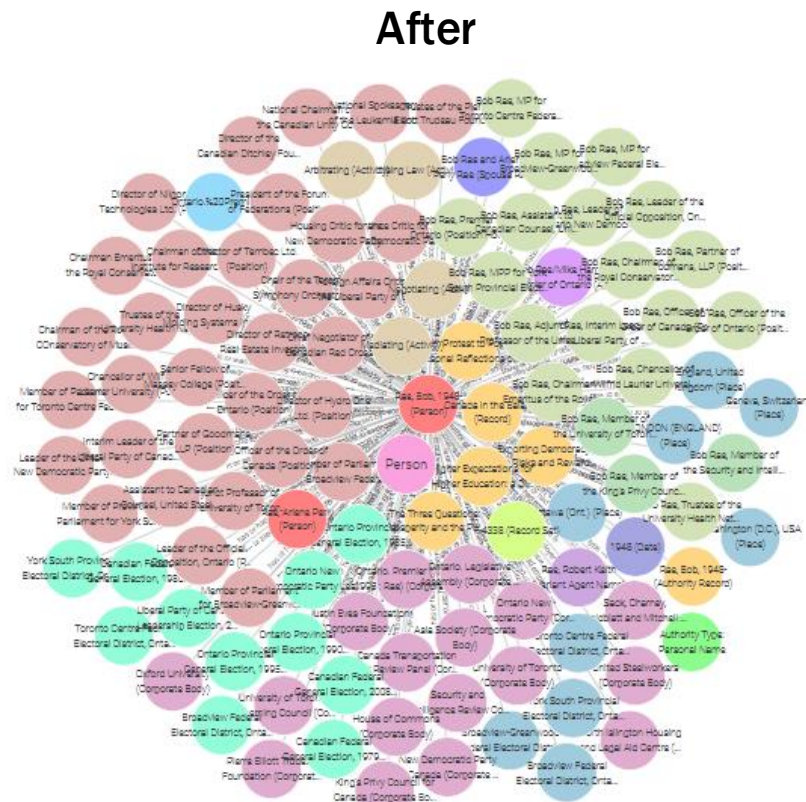
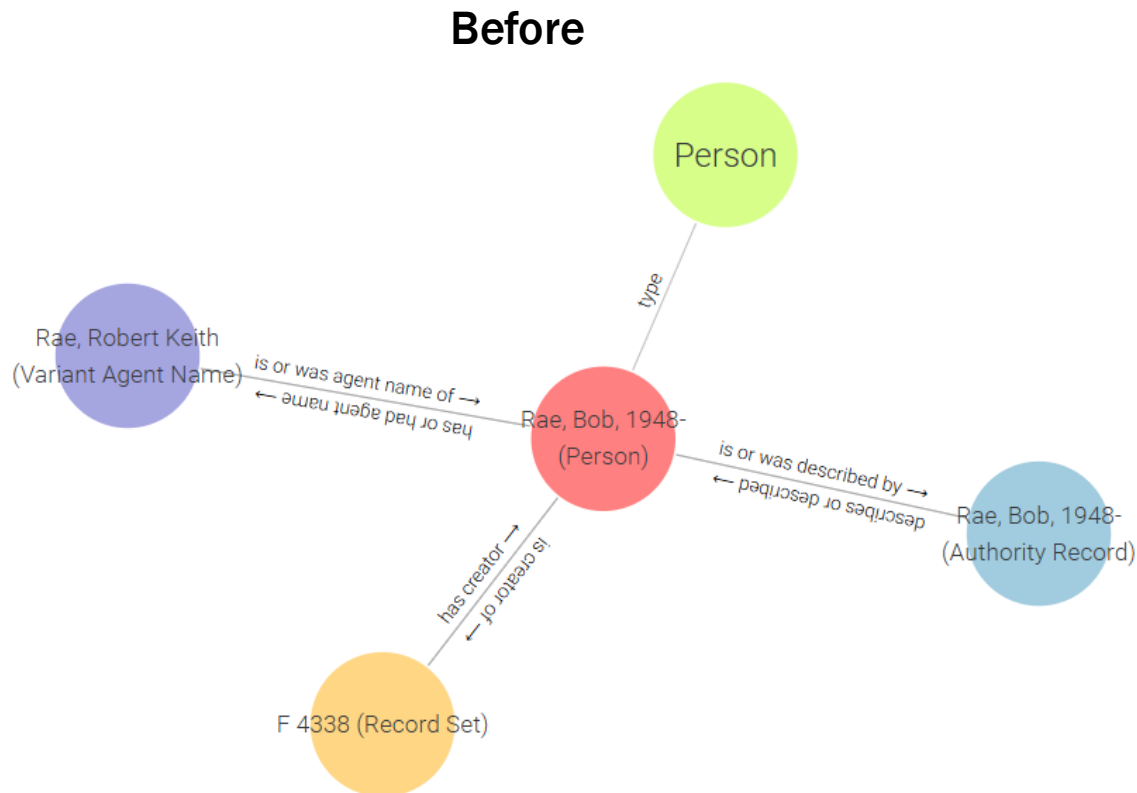
DATA ENRICHMENT

- Moving beyond conversion of legacy data to test new possibilities
 - Identifying new entities in existing unstructured information
 - Re-describing records

DATA ENRICHMENT

- Identifying new entities in existing unstructured information
 - Bob Rae's biography (authority record data element):
 - 9 place entities
 - 38 agent entities
 - 28 position entities
 - 11 event entities
 - 4 activity entities
 - 5 record entities

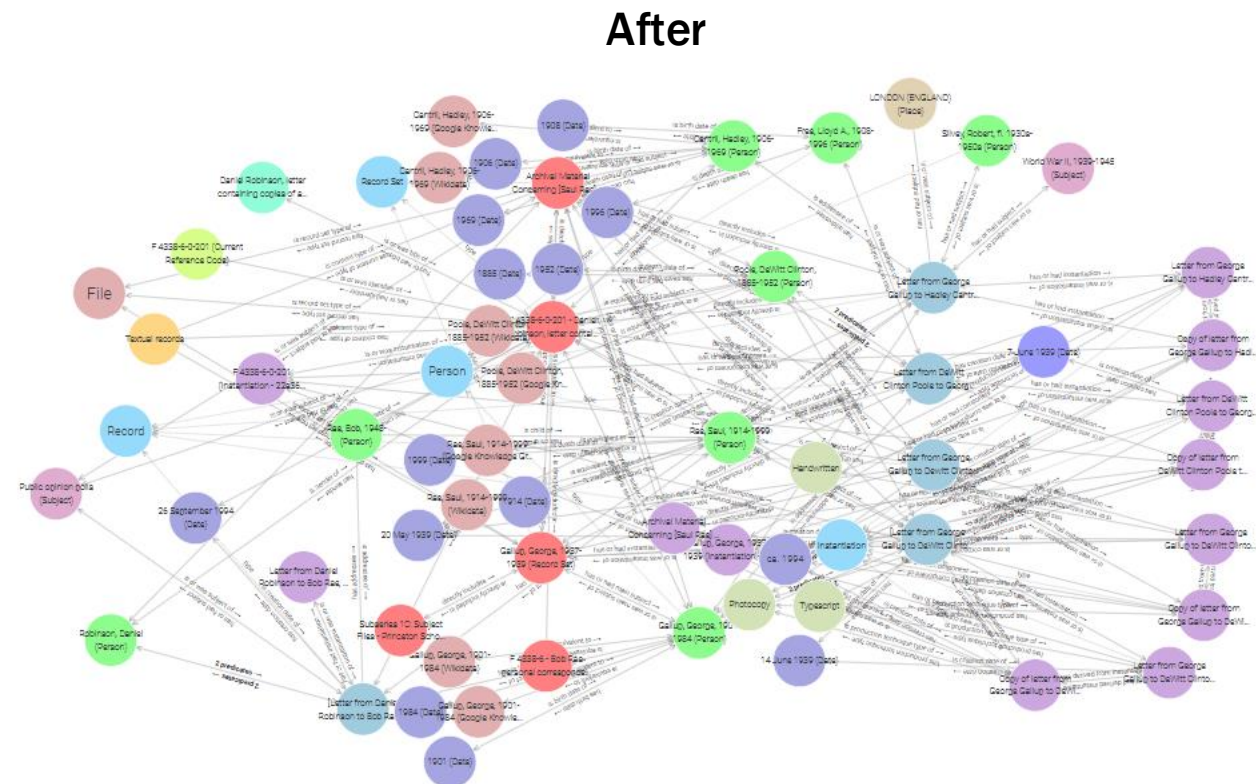
DATA ENRICHMENT



DATA ENRICHMENT

- Re-describing records
 - Describing the complexity of a single file in the Bob Rae fonds
 - 5 records (items)
 - 3 record sets (files – including one at Princeton University Archives)
 - 12 instantiations (3 of record sets, 9 of records)
 - 10 agents (8 persons, 2 corporate bodies)
 - 9 subjects (topics)
 - 3 production techniques
 - 16 Linked Open Data links (7 Wikidata, 6 Google Knowledge Graph, 2 LCSHs, 1 GeoNames)

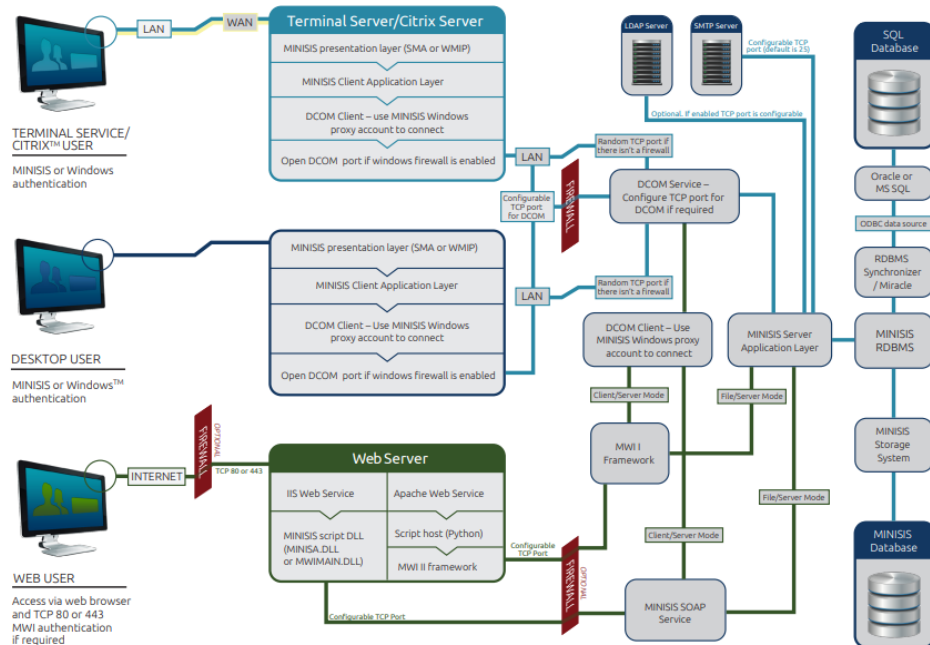
DATA ENRICHMENT



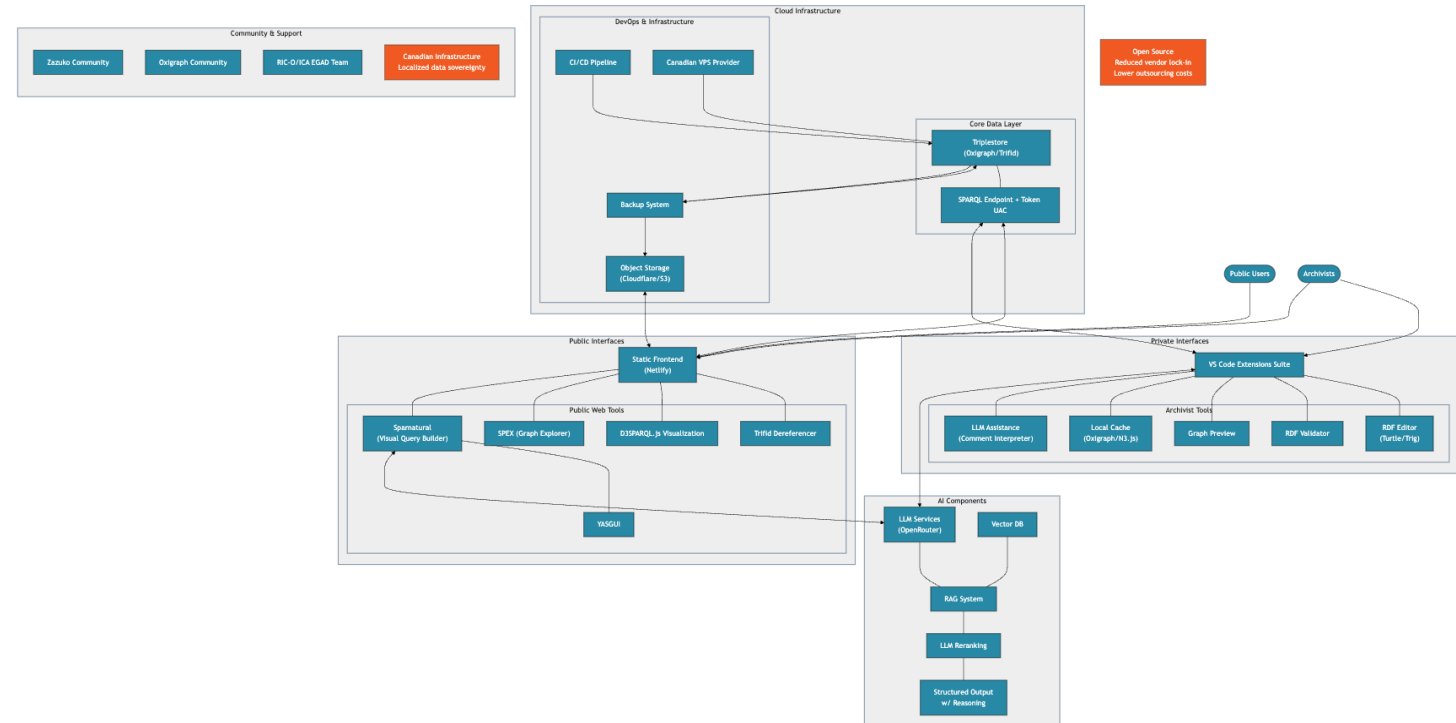
TOOLS FOR DEPLOYMENT AND DISCOVERY



Before



After



THANK YOU!

PLEASE CONNECT:

Aaron Hope – aaron.hope@ontario.ca

Pavel Zhelnov – p.zhelnov@mail.utoronto.ca

ontario.ca/archives



English: @ArchivesOntario

French: @ArchivesON (X & FB)
@archivesontario (IG)

