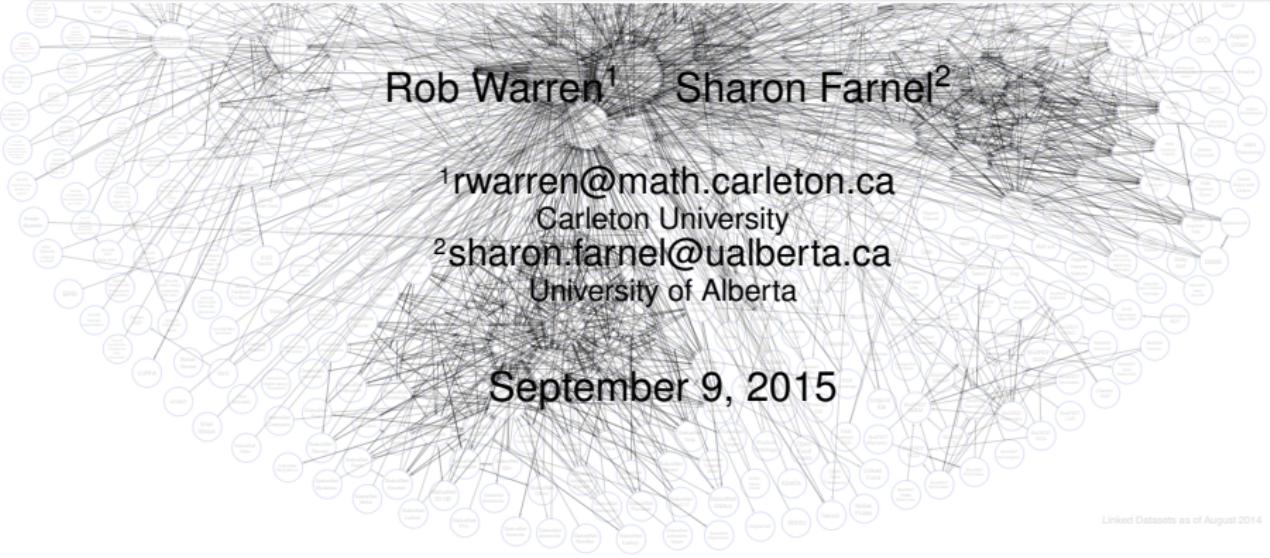




Linked Open Data - the value-add to a postcard collection



Rob Warren¹ Sharon Farnel²

¹rwarren@math.carleton.ca

Carleton University

²sharon.farnel@ualberta.ca

University of Alberta

September 9, 2015

Linked Datasets as of August 2014



1 Introduction

2 Places, Locations, Names

3 Facial Detection

4 Meta-data creation

5 Why this is important?

6 Closing notes

Peel Prairie Postcards Collection

Prairie Postcards

Activities & sports		Animals		Buildings	
Business & industry		Events		Land & land use	
Natural phenomena		Objects		Organizations	
People		Plants		Vehicles	

Background

- Enable the documentation of decisions / processes using RDF and Prov ontology.
- Promote organizational memory without too much overhead.
- Clouds of clouds -> from collections of collections to custom organizations.
- How much can we offload to machines?

Project Budget

Sharon's Budget



Rob's Budget (Matching Funds)



Manual Meta-data creation

- 'Stub' records (place, ID, etc.) from CSV.
- In-house guidelines, use of authorities (e.g., LC).
- Largely one individual over 2 year period; other staff and student assistance prior to launch.
- Final QA before website launch; minor cleanup and enhancements ongoing.

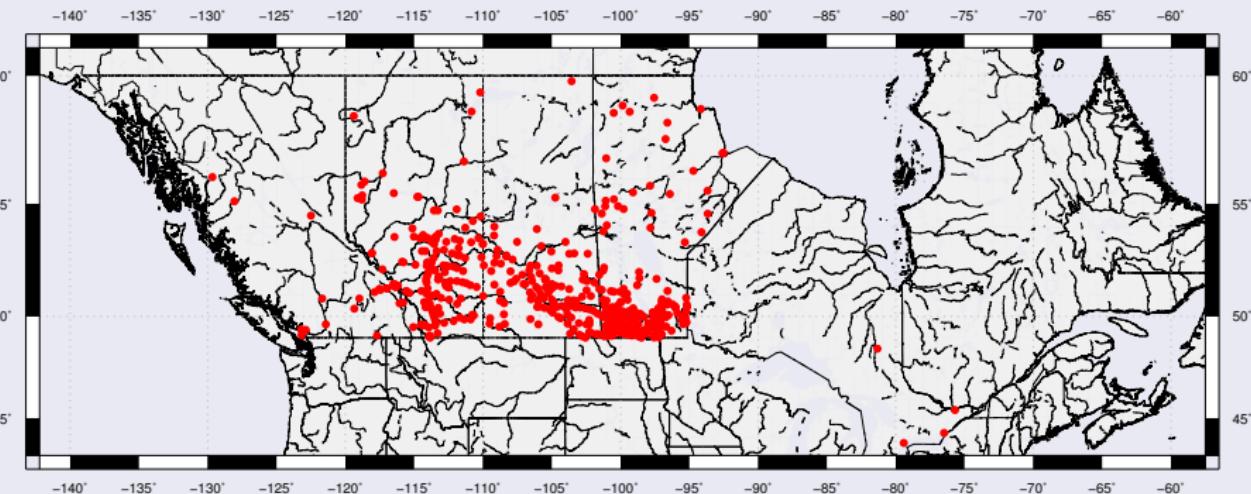
Tools and Processes

- Mods XML to rdf Mads via LC stylesheet (with minor edits).
- Additional RDF triples / linkages generated from raw XML Mods.
- Mostly perl, bash scripts and api's.

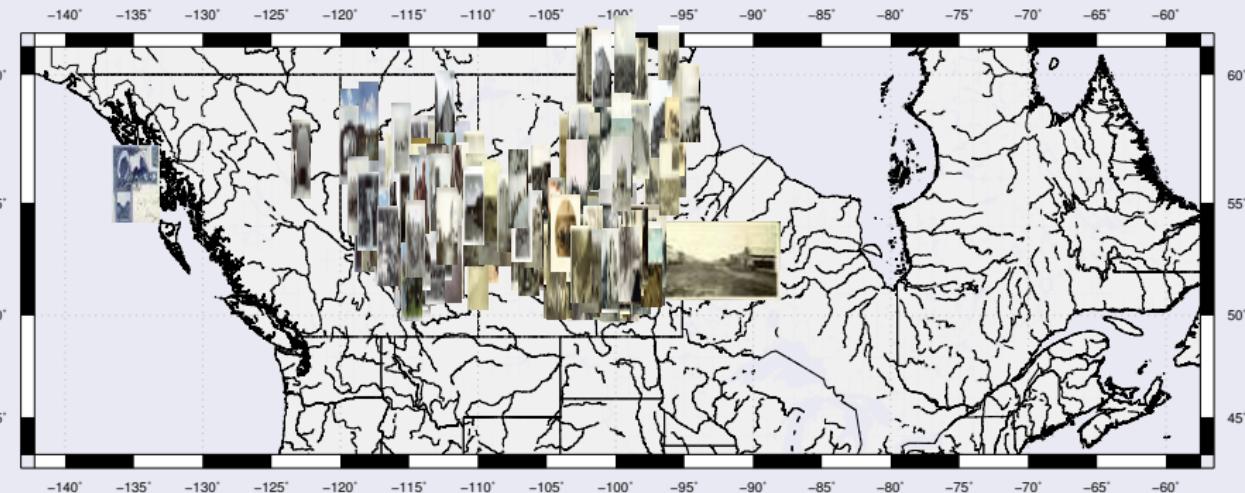
From strings to URIs

- Links to Geonames and Linked Geo Data (OSM).
- Links to LC Subject headings
- Try facial detection.
- Extract personal names and corporate names as foaf.
- Add full provenance information.
- ~~Extract colours from postcards: BW/Sepina?~~

URI's used



URI's used



Mads to geonames and linkgeodata

- Quality of location data ranges from great to plain wrong.
- Used geonames and linkedgeodata (OSM) api.
- Use OWL to deal with corner cases.
- Some data is likely unusable.

Data quality

People try very hard, but...

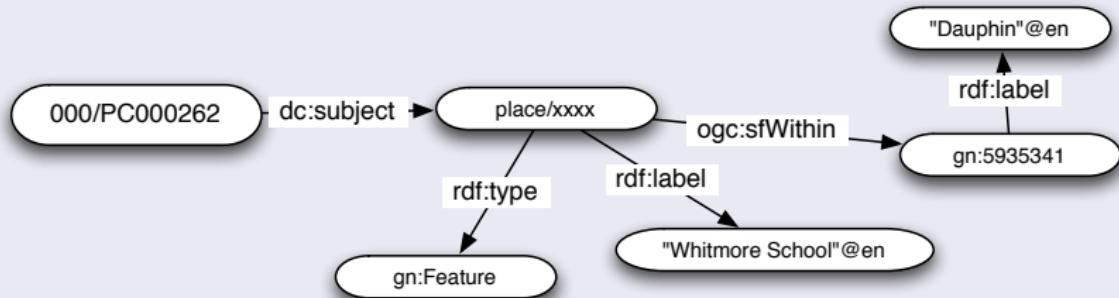
<city>Regina?</city>

??

<city>Trochu or Mannville</city>



<city>Whitmore School, Dauphin</city>



A running example

Postcard 8011

[ShareThis](#)

Jasper Wildcats, 1937-1938. [Jasper: c1937-1938.



Description: Group portrait of Wildcats basketball team players posing in two rows (standing and kneeling on their right knees) in front of a light colored background in a room.

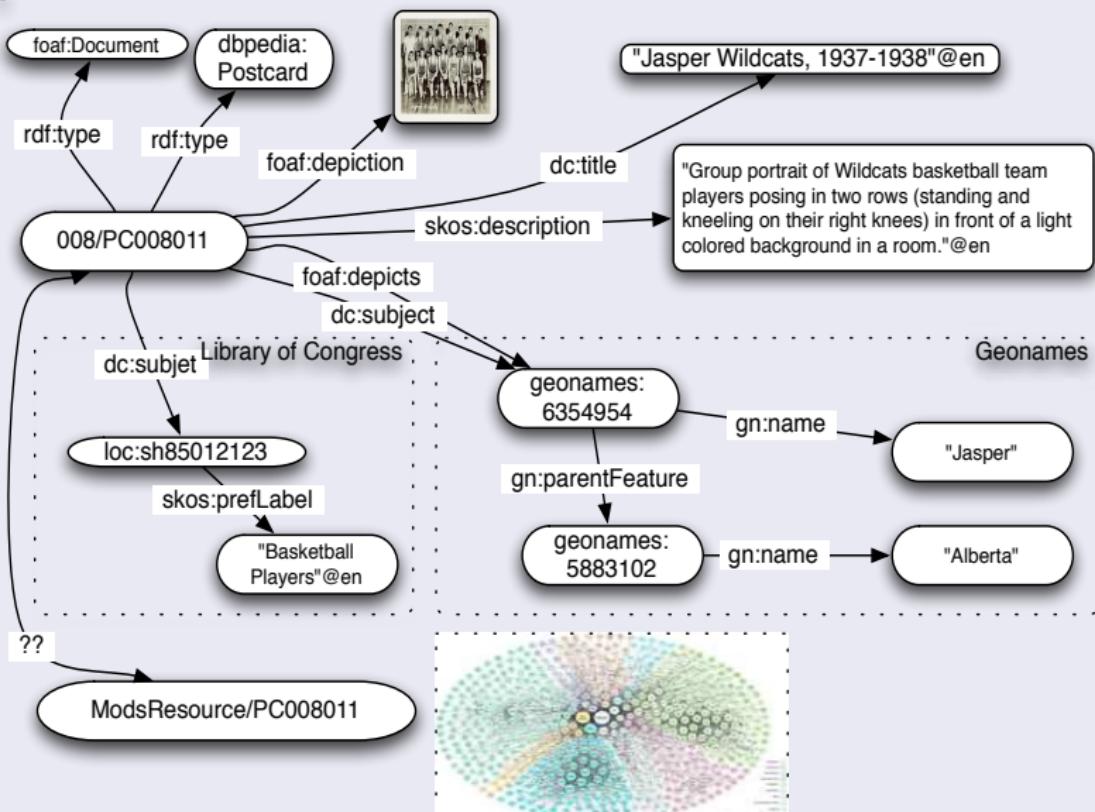
Physical description: 1 postcard : sepia ; 9 x 14 cm.

Language: English

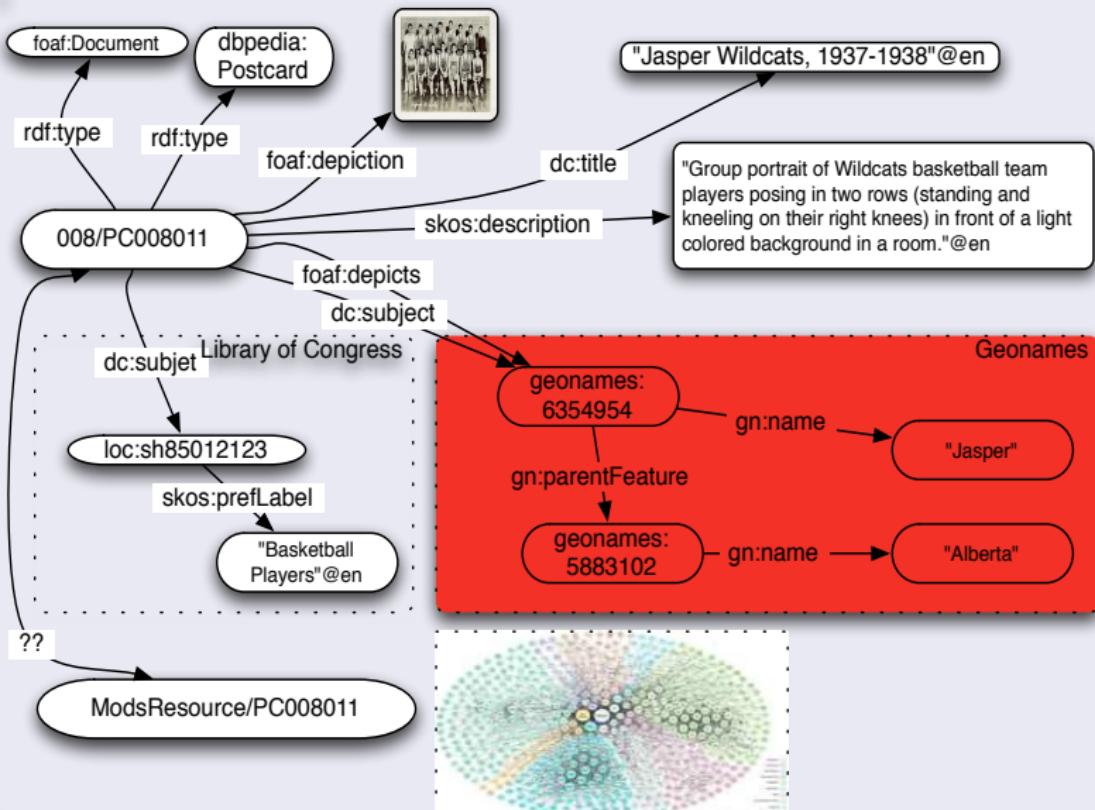
Subject headings:

- Canada--Alberta--Jasper
- Basketball players
- Sport clothes

Datasets / Process documents



Datasets / Process documents



Datasets / Process documents

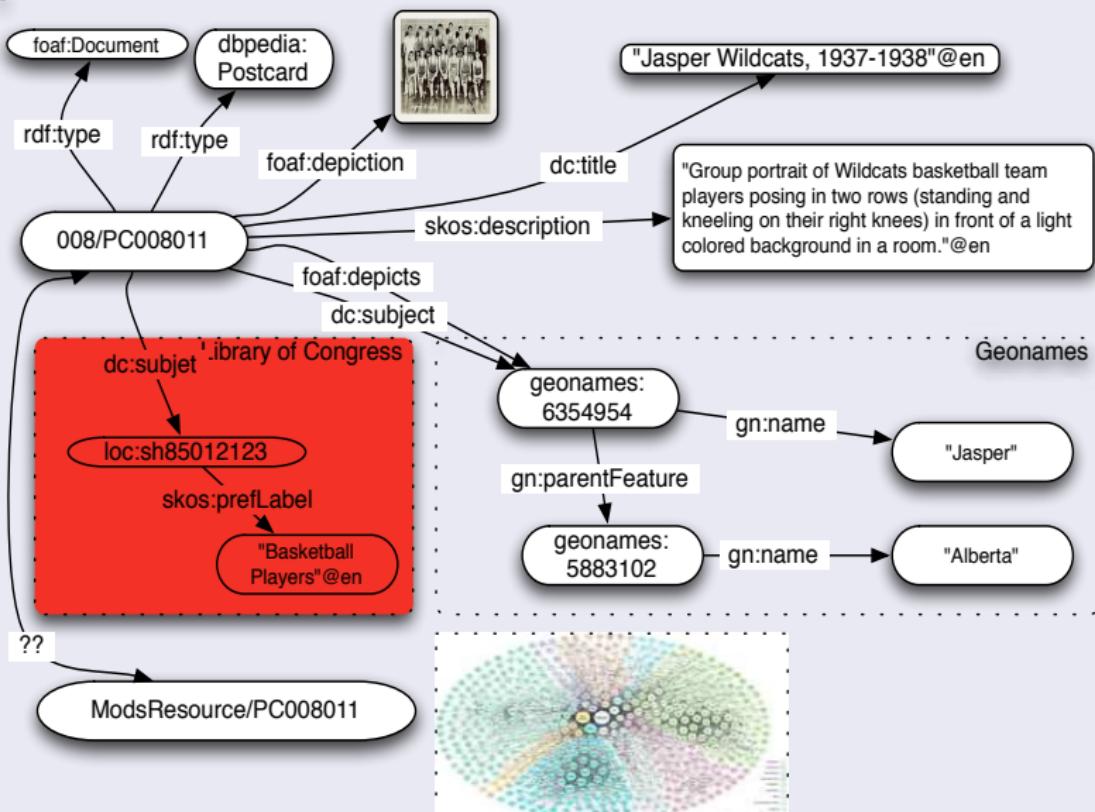


Image Analysis

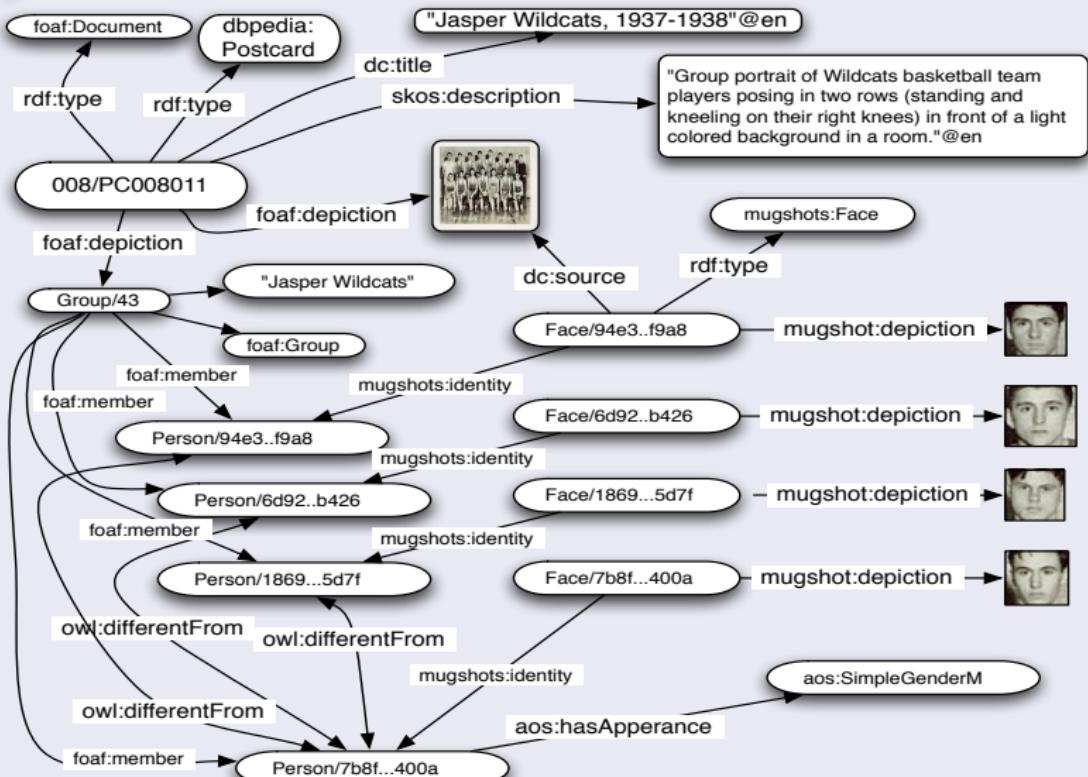


Image Analysis

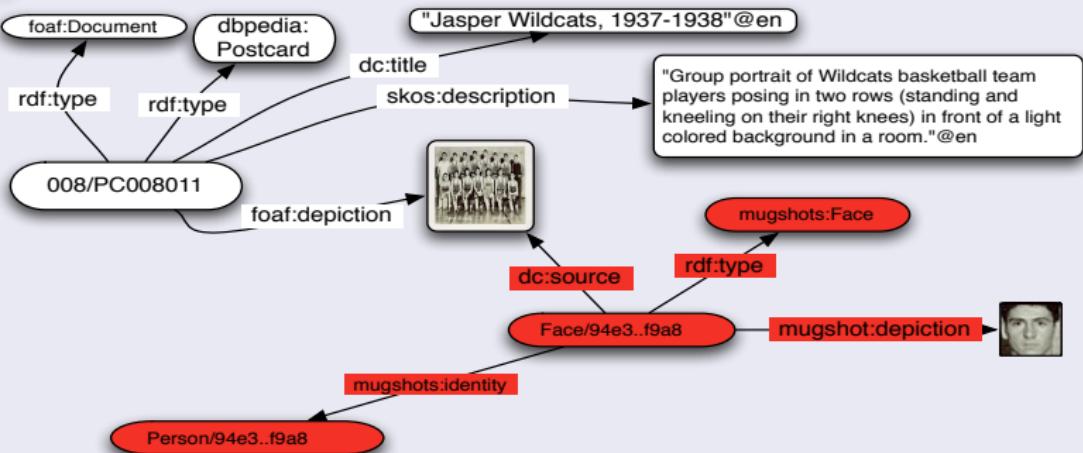


Image Analysis

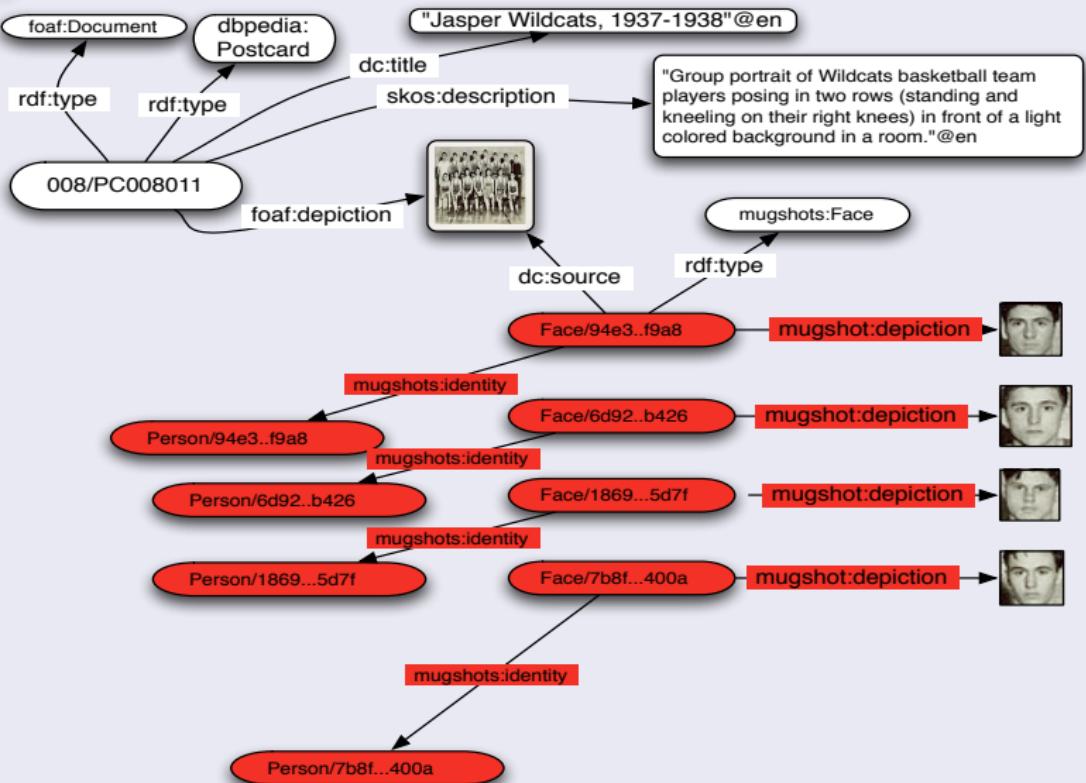


Image Analysis

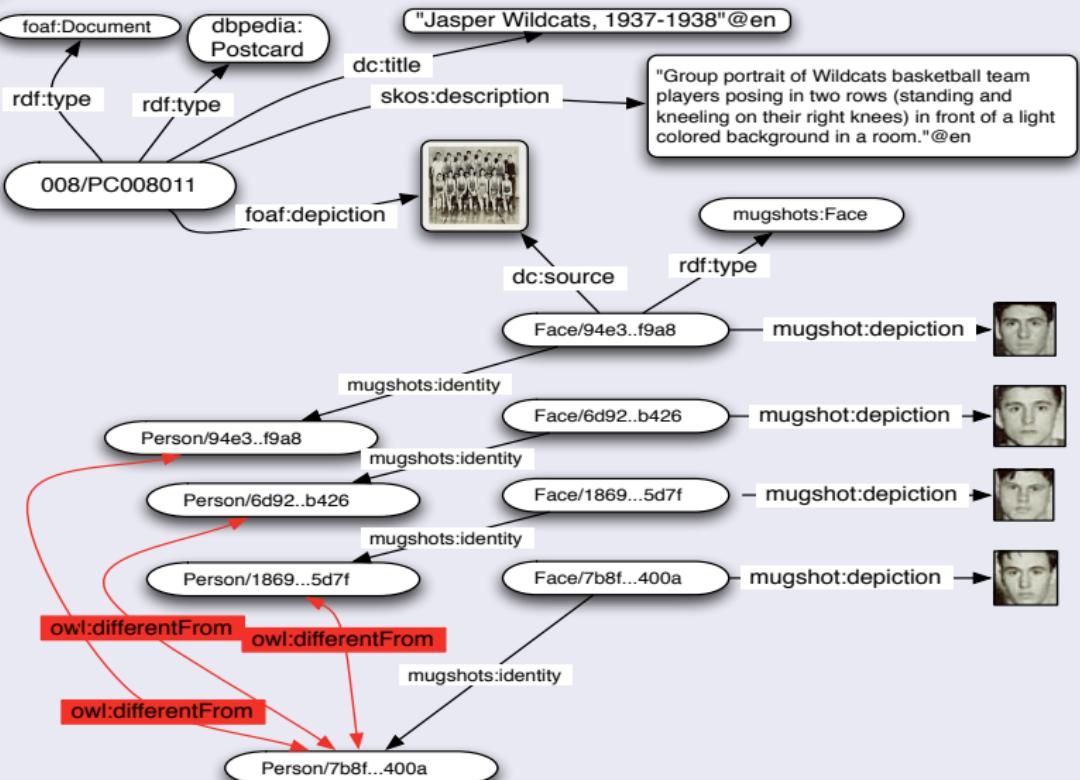


Image Analysis

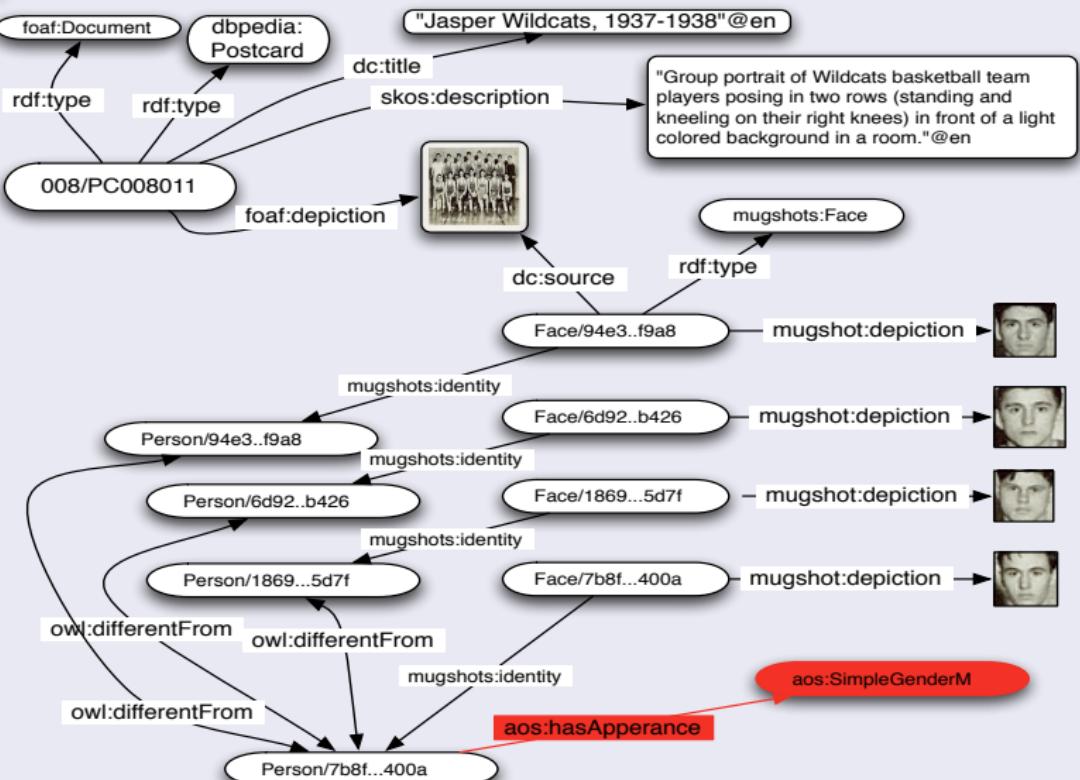


Image Analysis

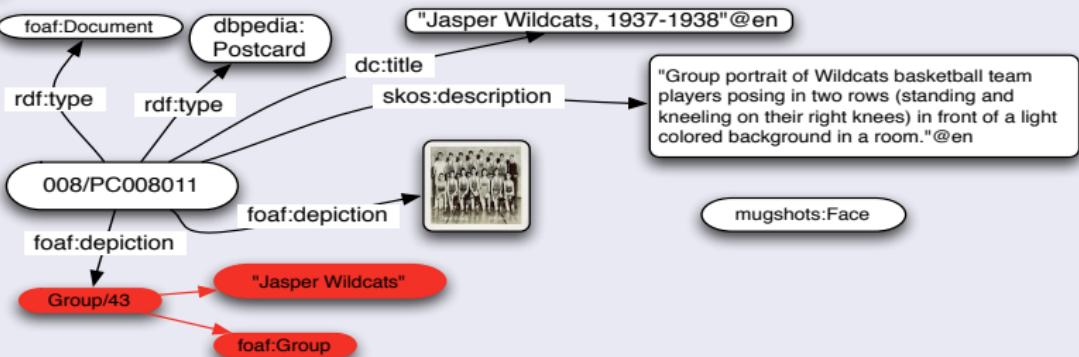


Image Analysis

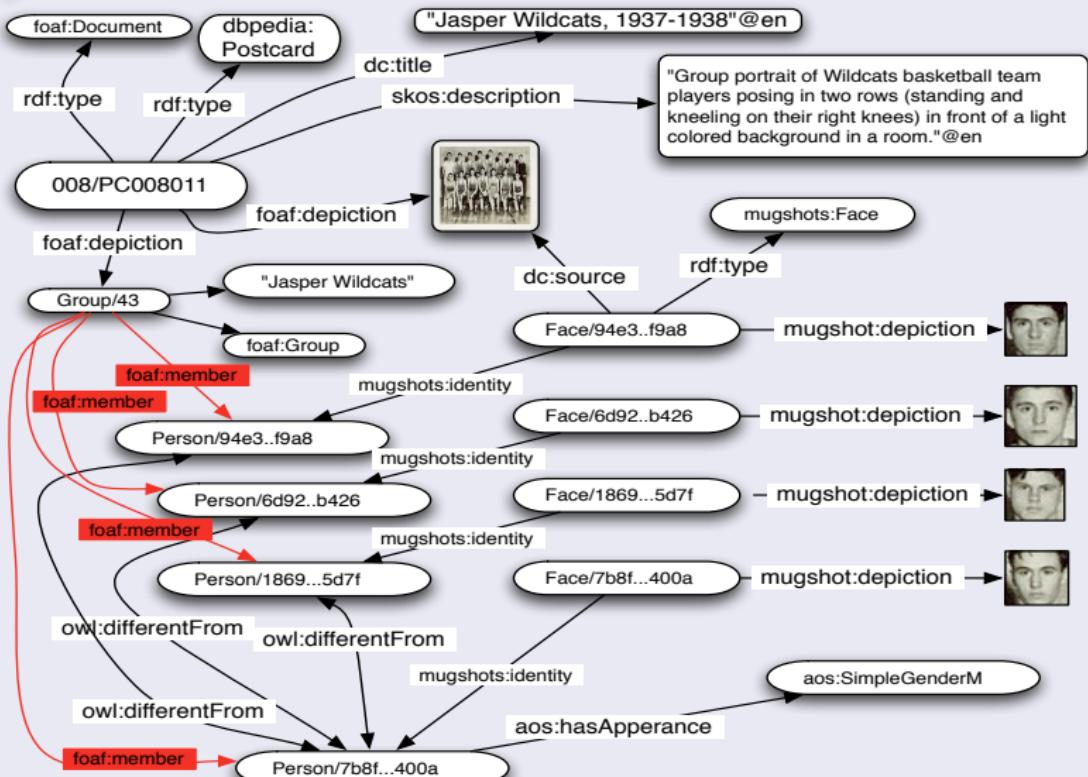
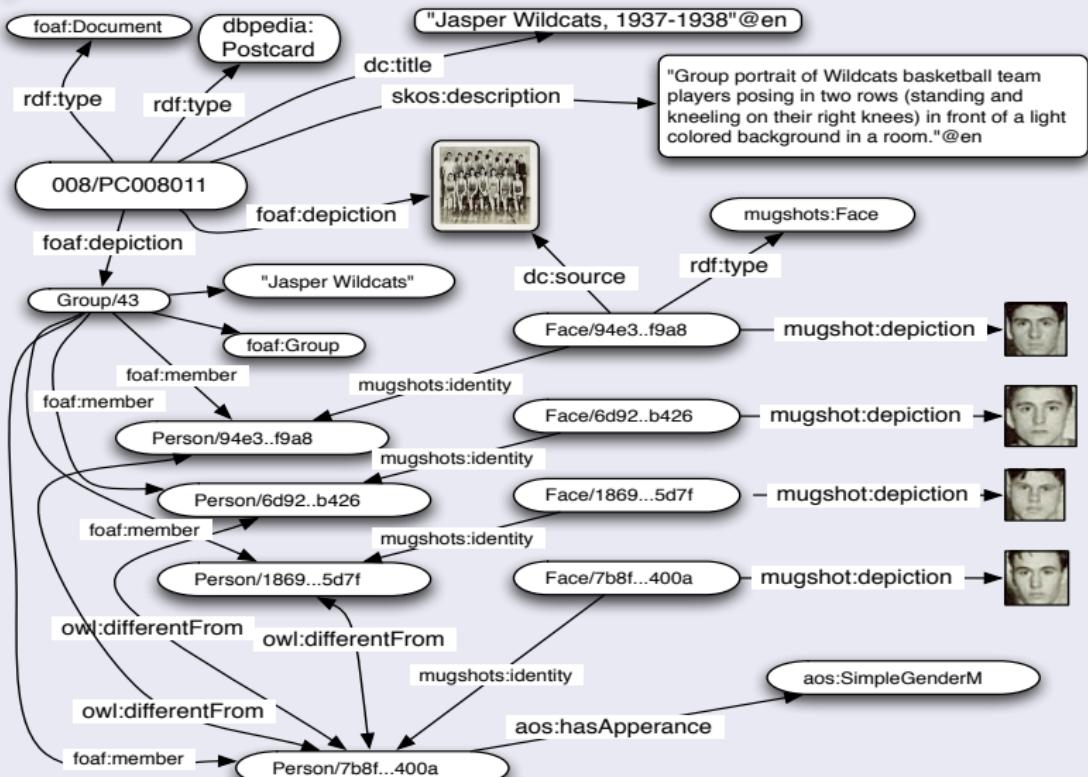


Image Analysis



Support for transcription?



A crowdsourcing campaign to describe McMaster's historical postcard collection.

[Home](#) [About](#) [Instructions](#) [Contact](#) [FAQ](#)

Click on an image to identify the postcard or [Reshuffle Postcards](#) | [Just Fronts](#) | [Just Backs](#).

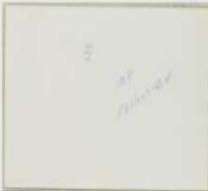
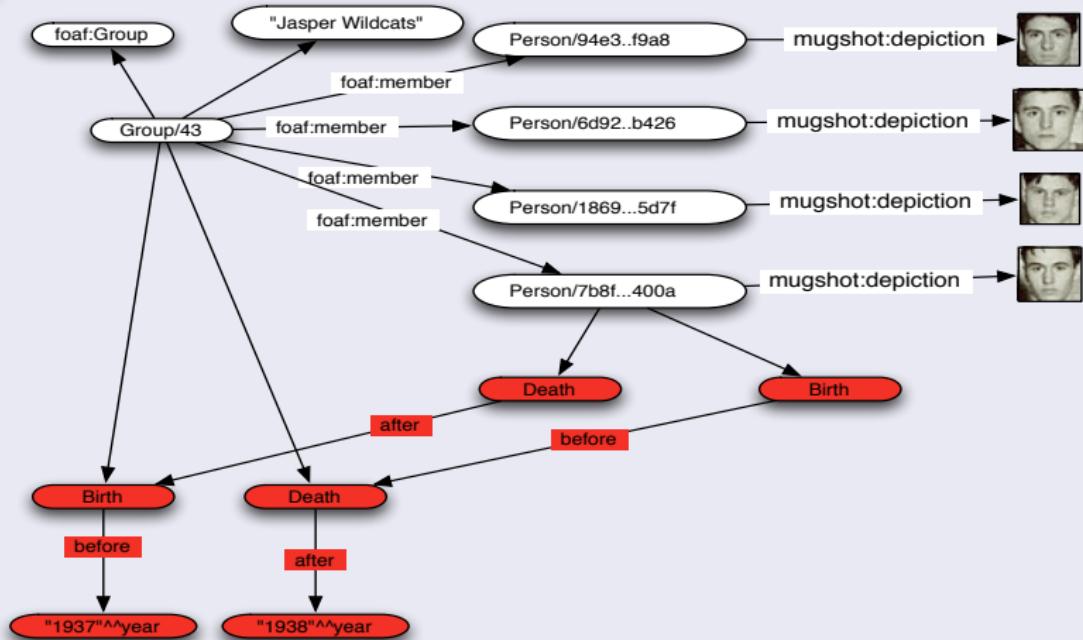
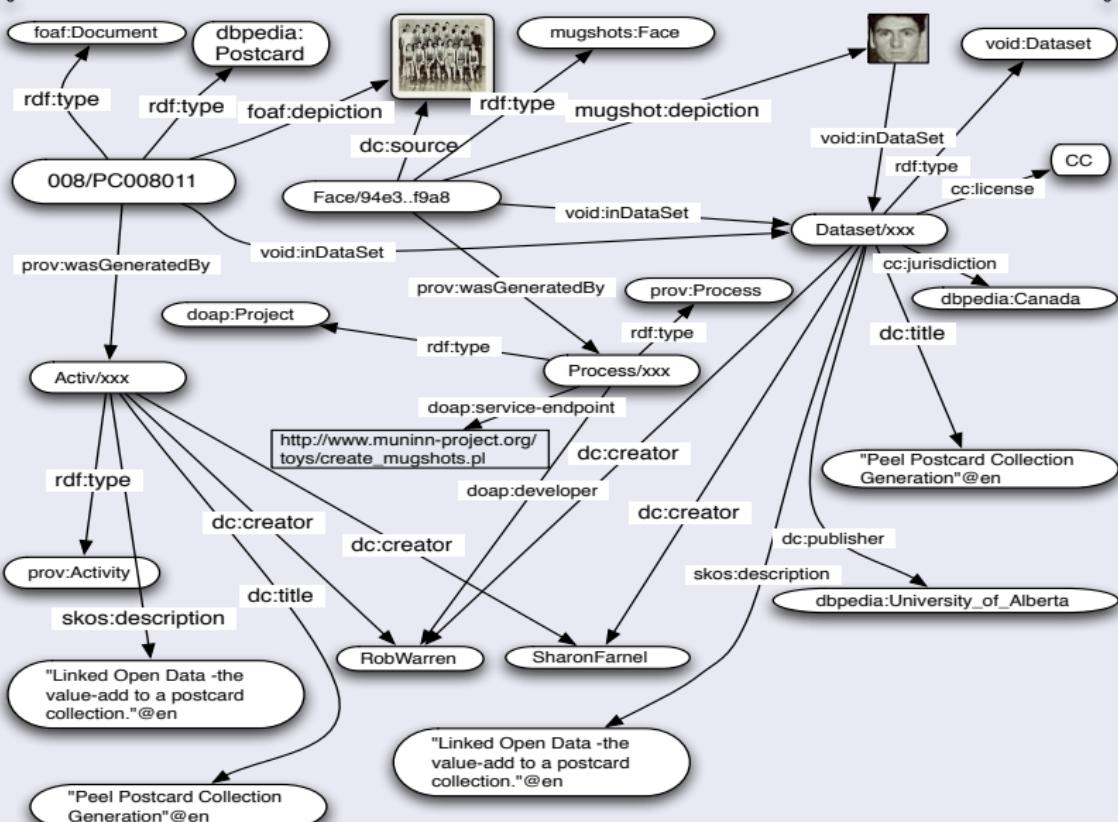


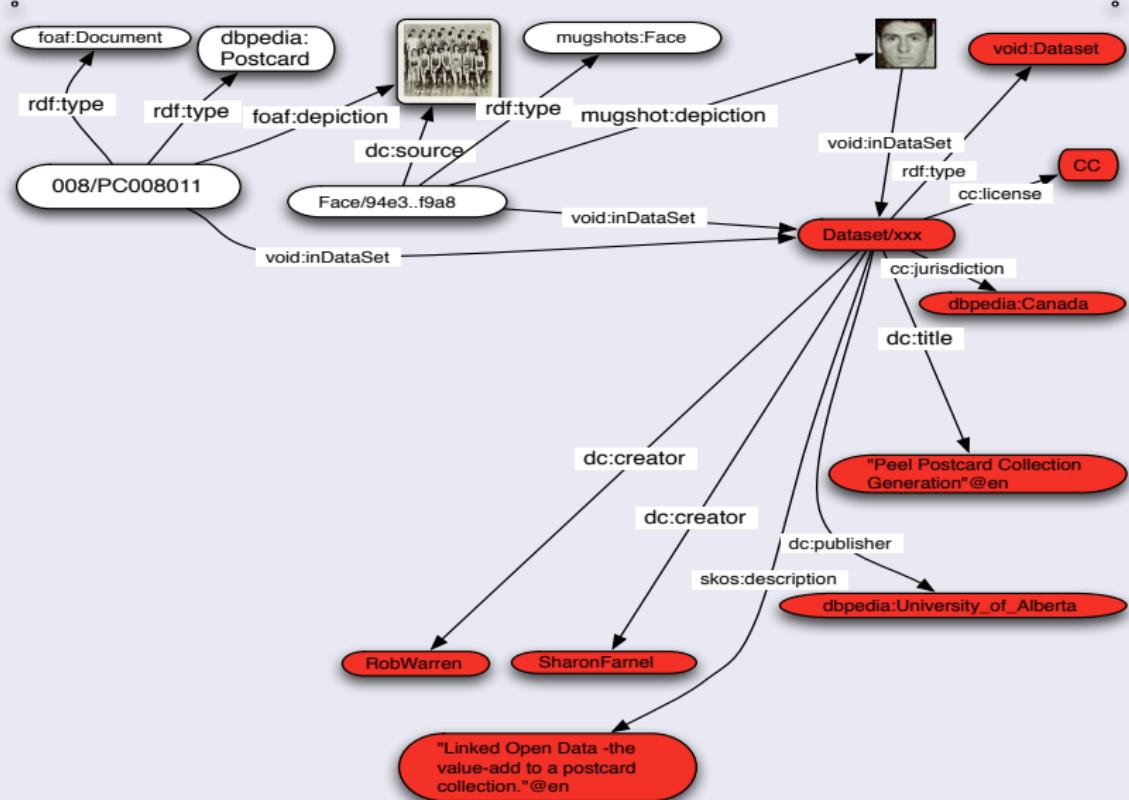
Image Analysis



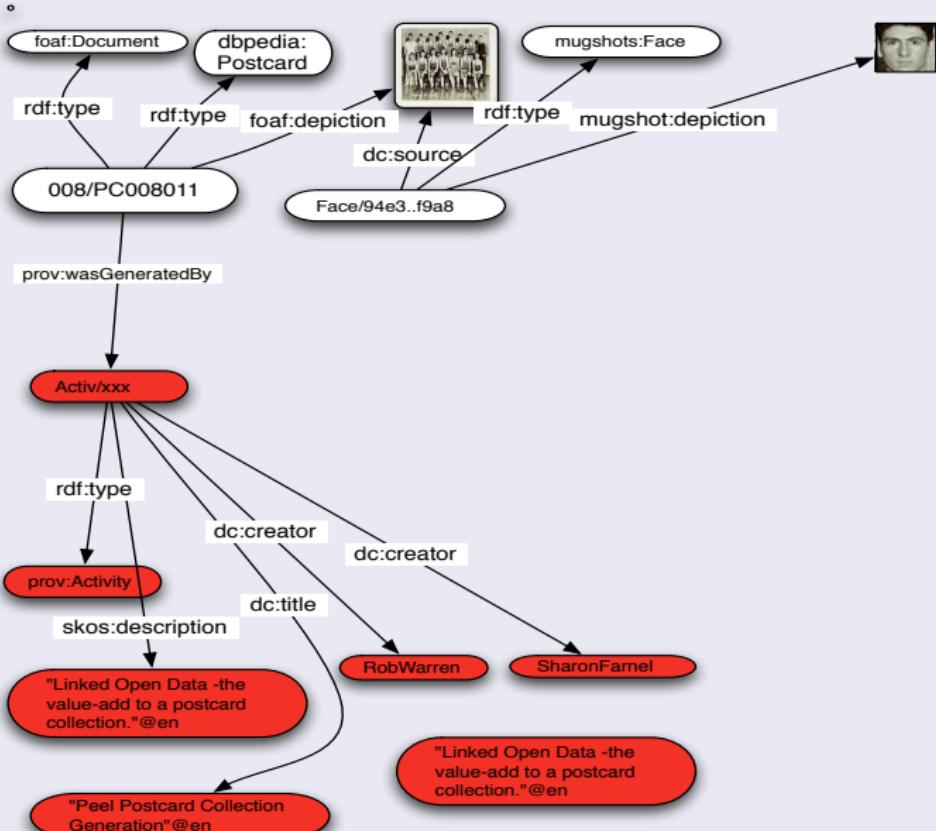
Datasets / Process documents



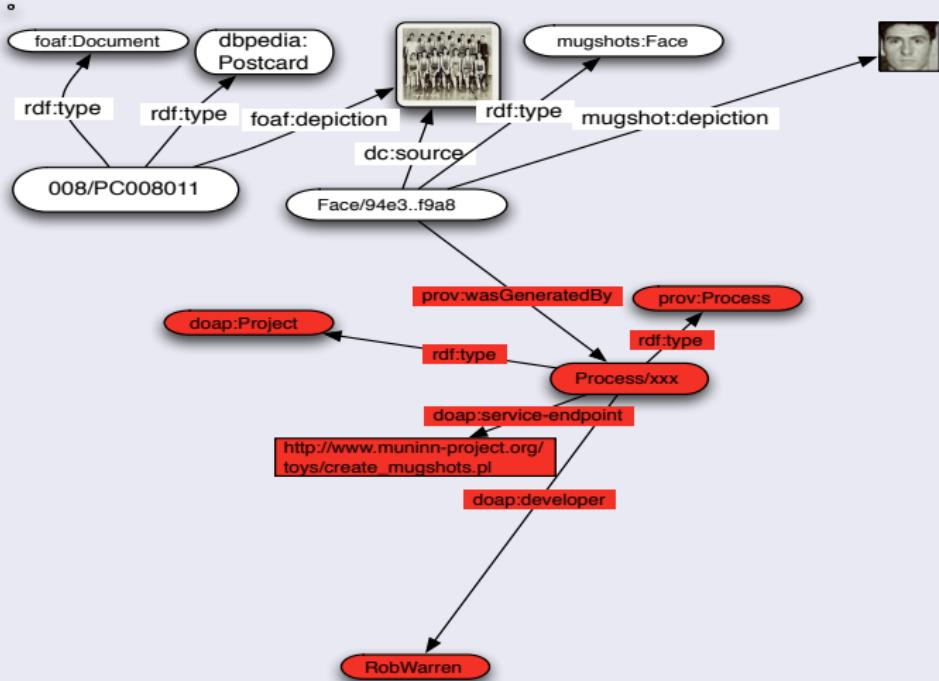
Datasets documentation



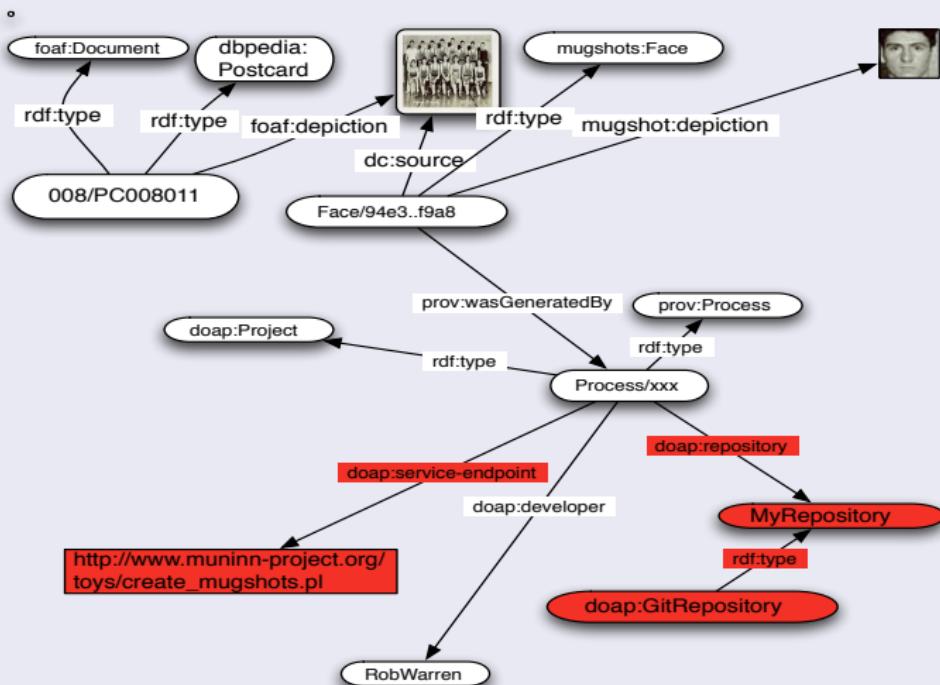
Datasets Creation



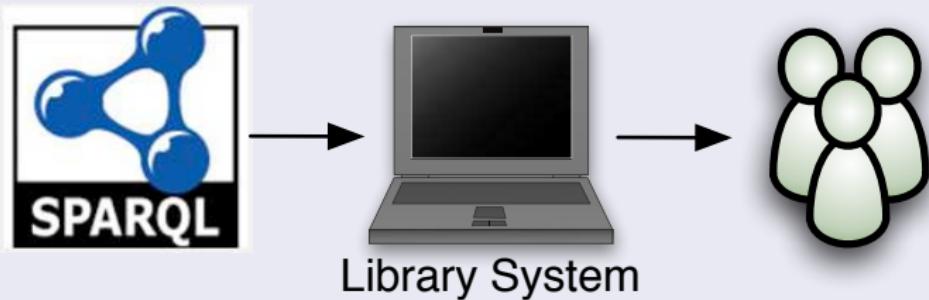
Datasets Process



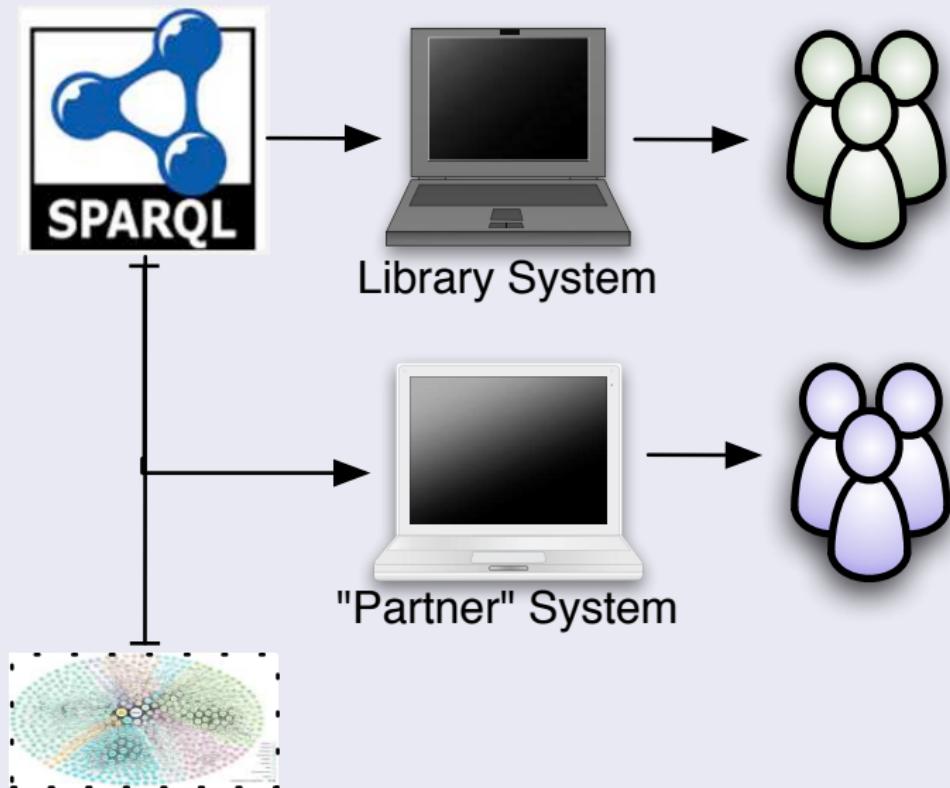
Datasets Process



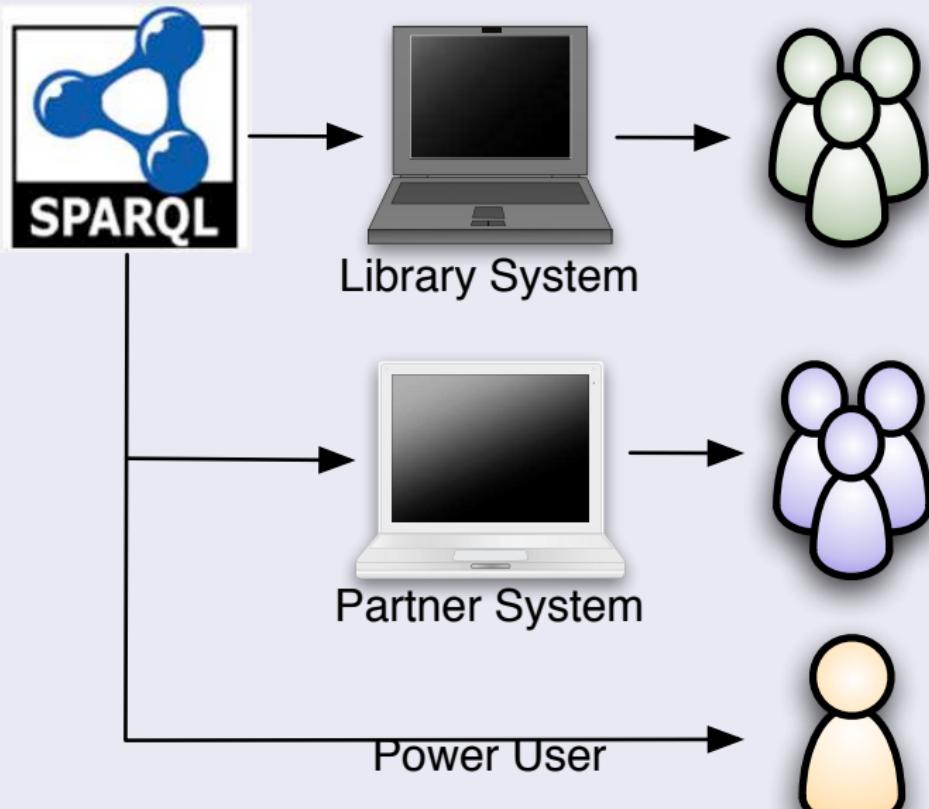
A possible future



A possible future

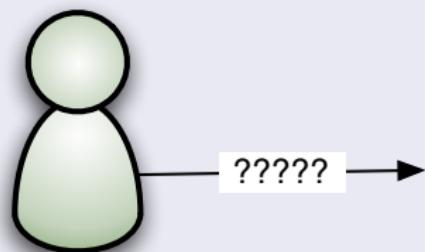


A possible future



Why is machine readable data important?

...but not for long!



The golden rule of Linked Open Data modelling:

The *thing* and the *name of the thing* aren't the same *thing*. ^a

^aSlide shamelessly stolen from the Canadian Writing Research Collaboratory slide deck.

Linked Open Data and OWL let you pick and choose standards

HOW STANDARDS PROLIFERATE:
(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

^a<https://xkcd.com/927/>

^bSlide shamelessly stolen from the Canadian Writing Research Collaboratory slide deck.

Lessons Learned, Next Steps

- Play with conversion to RDFa or Turtle and endpoints.
- Enhance other image collections in a similar way.
- Text-based collections - named entity recognition?
- Excel is the enemy of data quality.
- The artifact might be less important than the (meta?)data.
- RDF / LOD allows you to publish data, document it and (sometimes) fix mistakes later on.
- Separate the data from the application.

Code and scripts at:

<https://github.com/muninn/PC-Access2015>