Supporting Digital Scholarship in Islandora CLAW

Linked Data Modeling for Humanities Data

Open Repositories 2018, Bozeman, Montana

github: digitalutsc
Project Background

- Prof. Natalie Rothman
- Dragomans, diplomatic interpreter-translators ca. 1550-1750
Why Linked Data?

- Codifying, querying, visualizing complex relationships
- Reuse and sharing of data
- Enriching data with other data
- Collective development of a highly specialised knowledge domain
Project Stages

1. ontology development  
2. CLAW implementation  
3. data migration, refinement  
4. data analysis
What is Linked Data?

Tim Berners-Lee’s Rules:
1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names.
3. When someone looks up a URI, provide useful information, using the standards (RDF*, SPARQL)
4. Include links to other URIs, so that they can discover more things.
Developing an Ontology

- Scope
- Classes
- Properties
- Domain and range
- Instances

Preliminary Dragomans Ontology Decisions

Outline

Goal of this document

About Ontologies

Knowledge engineering

Part I: Domain...

Part II: Scan of...

Part III: Enumerating...

Part IV: Defining...

Part V: Define...

Part VI: Define...

Part VII: Creating...

Advanced ontology

Link classes and...

Defining properties

**DocumentPersonRelation**
- Document Source (1: instance class:Document)
- Person Target (1: instance class:Person)

**DocumentDocumentRelation**
- Document Source (1: instance class:Document)
- Document Target (1: instance class:Document)
- Document to Document relationship Properties as defined in the Vocabulary with their inverse relationships
- Production Date

**DocumentGenreRelation**
- Document Source (1: instance class:Document)
- Genre Target (1: instance class:Genre)
- Notes (1: text facet) (1: text facet)
- Mediating form (1: enumerated facet)
- Person (1: instance class:Person)

**DocumentPersonHonorificRelation**
- Document Source (1: instance class:Document)
- Person (1: instance class:Person)
- Honorifics (N: instance class: Honorific)

**Event**
- Name: Event (1: text facet)
- Event Type (1: instance class: EventType)
- Description (1: text facet)
Why Islandora CLAW?

- Linked Data Application Platform
- Existing skills and expertise
- Leverage content modelling features
- Leverage flexible views and search in Drupal
- Experiment with different modules in Drupal
CLAW Architecture

**Islandora CLAW Data Flow**

Create a resource

- User creates a node in **Drupal**
- Drupal Rule gets executed and messages are sent to **ActiveMQ**
- Alpaca **islandora-indexing-fcrepo** picks up the message and invokes Milliner microservice.
- Milliner creates a resource in **Fedora**
- Gemini provides path mapping service for Milliner.
- Alpaca **islandora-indexing-triplestore** picks up the message and uses **SparqlUpdateProcessor** to index it in **Blazegraph**.
Implementing in CLAW

- Content Modeling based on ontology
  - Classes and Taxonomies into Content Types
  - Views to show related content
  - Search configuration for faceting
- RDF Mapping
- Inferences
Browse Research Materials

Content Type
- Document (2113)
- Archival Object (631)
- Person (570)

Topic
- ahdnname (755)
- bailate (2)
- biographical (21)
- border conflict (44)
- border disputes (85)
- borderlands (9)
- borders (6)
- captivity and slavery (19)
- captivity and slavery (2)
- categories (1)

AdC, Misc. Civil box 235, fascicle 2

Archival Path
AdC, Misc. Civil box 235, fascicle 2

Archival Identifier
Venetian State Archives

AdC, Misc. Civil box 260, fascicle 20
Aggiondrite, Zuanne election

Event Type

election

Description

Modoneo, was elected by Prov.r della Cefalonia as interprete della lingua greca in the
examine case and respond that supplicant doesn't seem to have v thorough
impedirsi in altri che nella sola semplice interpretazione riportandosi'

Source Person

Aggiondrite, Zuanne

Related People

Role

Parent

Person

Aggiondrite, Andrea

Notes

Father
targetEntityType: node
bundle: person
types:
  - 'schema:Person'
fieldMappings:
  field_family_name:
    properties:
      - 'schema:familyName'
  field_given_name:
    properties:
      - 'schema:givenName'
  field_language:
    properties:
      - 'schema:inLanguage'
  field_event:
    properties:
      - 'schema:event'
  field_p2p_role_relation:
    properties:
      - 'dragomans:personToPersonRelation'
  field_associated_position:
    properties:
      - 'dragomans:positionRelation'
Implementation Issues

- N-ary Relations
  - Indexing triples of N-ary relations (custom code needed)
- RDF mapping is static, lacks support for complex fields
  - Possibly addressed by context
- Inference/inverse relations are not in Drupal
Data Migration

- Data cleanup in spreadsheets (OpenRefine)
- Drupal’s Migration Tools: migrate_plus, migrate_tools and migrate_source_csv
- CSV to Nodes and Fields
- Relations established by migrations_update
Dragomans

Content Type
- Document (2113)
- Archival Object (631)
- Person (570)

Browse Research Materials

Operator
Search
Contains any of these words

Topic
- ahname (755)
- bailate (2)
- biographical (21)
- border conflict (44)
- border disputes (85)
- borderlands (9)
- borders (6)
- captivity and slavery (19)
- captivity and slavery (2)
- categories (1)

AdC, Misc. Civil box 235, fascicle 2
Archival Path
AdC, Misc. Civil box 235, fascicle 2
Archival Identifier
Venetian State Archives

AdC, Misc. Civil box 260, fascicle 20
prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

SELECT ?sub WHERE {
}

LIMIT 20
Next Steps...

- Interoperability with other linked data platforms (enrichment, expansion of data)
- Addition of flexible scholarship tools that address needs of multiple projects
- How to speed up deployment of new projects on the platform without increasing staffing needs within the unit (i.e., creating a digital-scholarship-on-demand platform?)
- How to simplify on-ramping of researchers who may not be experts in linked-data?
- Other thoughts and suggestions from the audience?