Ontotext CH/DH Projects

(A lot more details in LOD for CH webinar, 2016-09, 130 slides)

Vladimir Alexiev, PhD, PMP

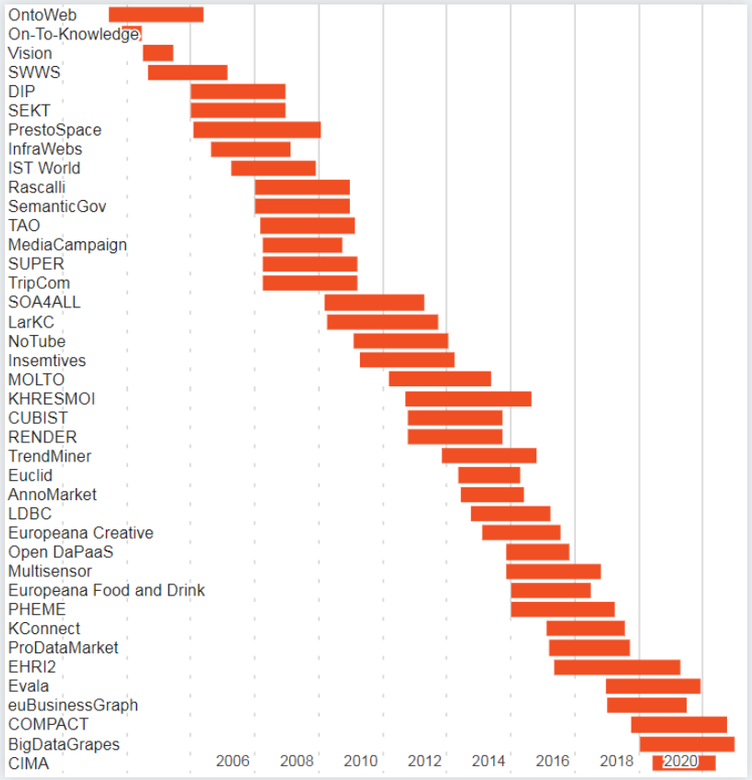
2018-06-13, CLADA-BG Meeting, Sofia

ontotext

About Ontotext

- Founded 2000, part of Sirma Group (400 people, BSE:SKK, part of SOFIX), venture funding 2008
- 65 people: 7 PhD, 30 MS, 20 BS, 6 university lecturers. Offices in Sofia, Varna, London
- Core part of Sirma Strategy 2022 with focus on cognitive computing
- Working on: semantic technologies, semantic repositories, semantic text analysis, machine learning
- Semantic Graph Database: Ontotext GraphDB
- Semantic data integration and building of Knowledge Graphs
- Semantic text analysis: entity, concept, relation extraction, document classification
- Recommendations, sentiment analysis
- Machine learning: entity disambiguation, deep learning in graphs, etc

Research Projects



Current Projects:

- EHRI2: European Holocaust Research Infrastructure (H2020 RI): **CH**
- Evala: Congnitive and Semantic Links Analysis and Media Evaluation Platform (EuroStars)
- euBusinessGraph: Innovative Data Products and Services for Company Data (H2020 BigData Experimentation)
- COMPACT: From Research Through Policy on Social Media and Convergence (H2020 CSA)
- BigDataGrapes: BigData to Enable Global Disruption of the Grapevine-Powered Industries (H2020 BigData Research)
- CIMA: Company Intelligent Matching and Linking (BG OPC ISIS)

Research and Innovation Awards

Arguably, Ontotext is the most innovative Bulgarian software company.

- Innovative Enterprise of the Year 2017
- EU Innovation Radar Prize 2016 nomination
- BAIT Business Innovation Award 2014
- Innovative Enterprise of the Year 2014
- Washington Post “Destination Innovation” Competition 2014 Award
- Pythagoras Award 2010 for most successful company in EU FP6 projects

We have more EU research projects than some universities combined

Industries and Clients

80% of our sales are in the UK and US

- Media: BBC, UK Press Association, NL Press Association (NDP)…
- Financial Info: S&P Global Platts, Euromoney, Financial Times, Nikkei…
- STEM Publishing: IET, Oxford University Press, Wiley, Elsevier, Springer Nature…
- Life Science: AstraZeneca, Novartis…
- Government: UK Parliament, The National Archives, Natural Resources Canada…
- Cultural Heritage and Digital Humanities (see next)

CH/DH Projects

- ResearchSpace: British Museum, Yale Center for British Art. Largest museum collection, CIDOC CRM, semantic search…
- (with Sirma Enterprise) ConservationSpace, Sirma MuseumSpace
- Medieval Cultures and Technological Resources (VCMS) COST action
- Europeana: Creative, Food and Drink (sem app), OAI PMH, SPARQL, members council, 5 work groups, Data Quality Committee
- Bulgariana national aggegator: initiator
- Getty Research Institute: vocabularies LOD and helping on Getty Museum LOD
- Carnegie Hall LOD
- American Art Collaborative consulting: 14 US museums integrating data using CIDOC CRM
- European Holocaust Research Infrastructure: semantic archive integration. 4+4 years, heading towards ERIC
- Canadian Heritage Information Network consulting (national aggregator moving to LOD)
- Wikidata: frequent contributions (authority control)
- DBpedia: contributions, association member, data quality and ontology committee
- CLADA BG: key participant in both CLARIN (NLP) and DARIAH (CH/DH)

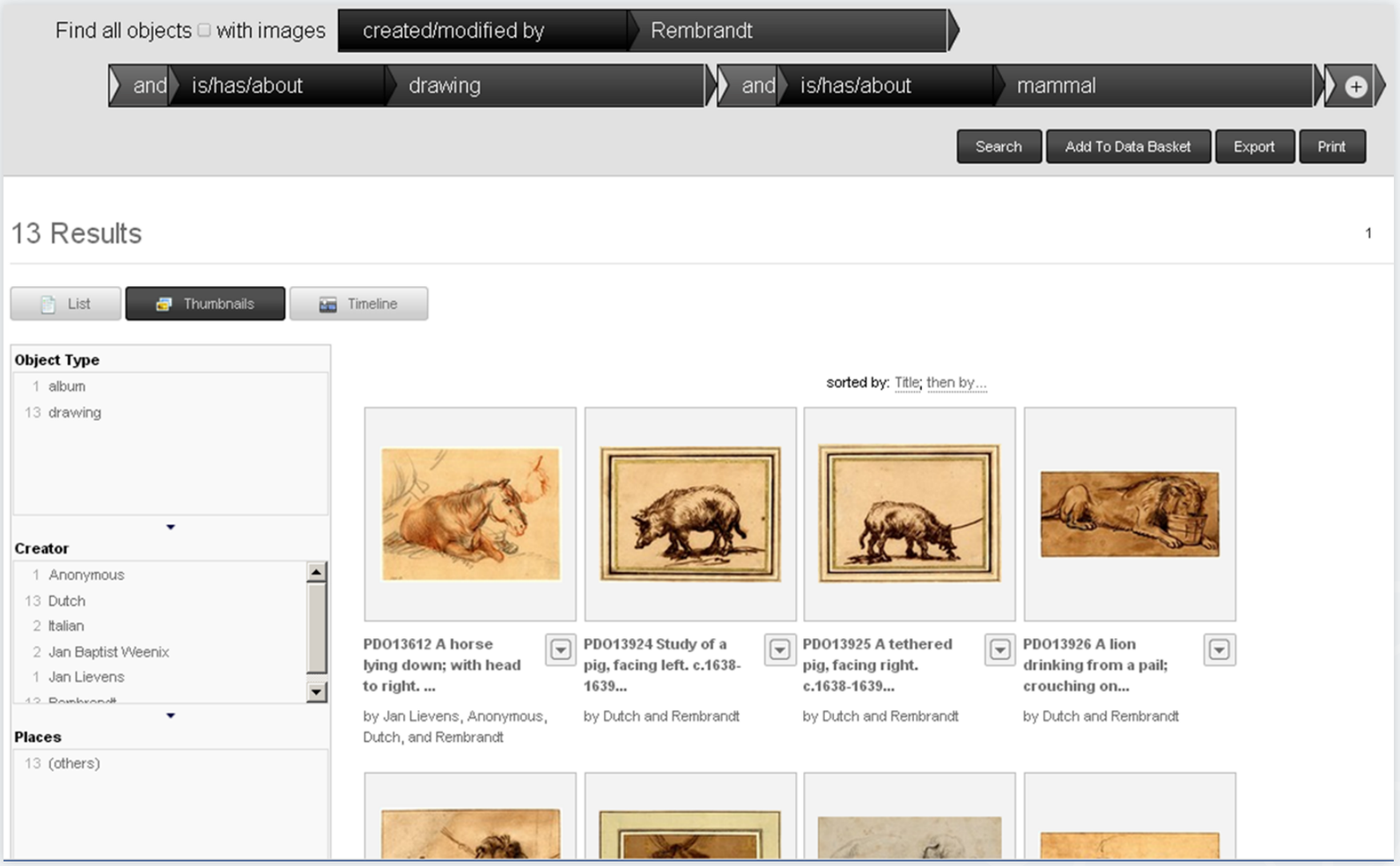
Knowledge Graphs

Knowledge Graph	Year	M obj	B triples
British Museum	2013	2	0.92
Polish Digital Library	2013	3.1	0.53
Europeana	2014	20.3	3.8
FactForge	2006-now	~14	3.2
LinkedLifeData	2008-now	~12	10.2
Company Graph	2017-now	6	3
Dun & Bradstreet	2017	210	30

Details about the first 5 are in V.Alexiev et al, Large-scale Reasoning with a Complex Cultural Heritage Ontology (CIDOC CRM), Workshop Practical Experiences with CIDOC CRM and its Extensions (CRMEX), TPDL 2013, slide 17

ResearchSpace

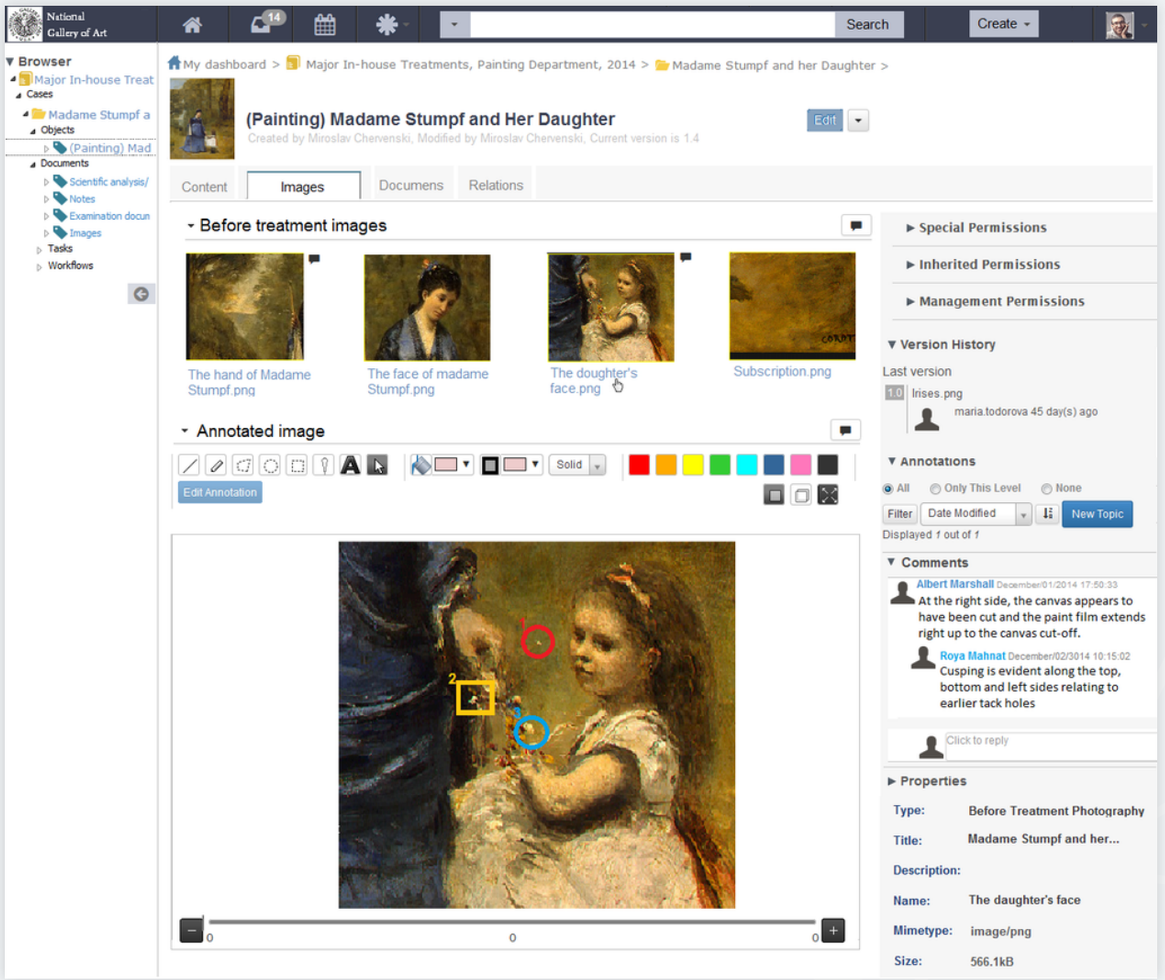
- Semantic integration based on CIDOC CRM, search (first implementation of Fundamental Relations search), data & image annotation, data basket, etc



ontotext

ConservationSpace

- Line-of-business application for conservation specialists. International consortium (US NGA, DK SMK, UK Courtauld etc). Based on the Sirma Enterprise Platform and Ontotext GraphDB, Ontotext helped with the ontologies.



ontotext

MuseumSpace

- Based on the Sirma Enterprise Platform and ConservationSpace experience. Collections, exhibitions, curation…

Conservation Management is the commercially supported and cloud-based version of the open-source Conservation Space project of the National Gallery of Art (USA) with the support of The Andrew W. Mellon Foundation....

Collection Management simplifies your work, such as creating libraries with cultural objects, organized by different criteria. Perfect for small and medium organizations with basic collection management requirements. All cultural objects are...

Image Annotation feature, uses Mirador's image tool. Mirador is an open-source, configurable, extensible, and easy-to-integrate image viewer, which enables image annotation and comparison of images from repositories dispersed around the world....

Exhibition management helps organizations to prepare for an upcoming exhibition. This feature allows you to manage all budgeting, marketing activity, floor plans, contracting and other processes related to exhibitions process. [gallery...

Loans Management allows institutions to manage all incoming and outgoing loans regarding all cultural objects within the institution. Incoming/Outgoing Loans Prepare contract agreements or Document special requirements Schedule/Prepare for Arrival/Approval...

Location and Movements feature, allows institutions to monitor the location of each cultural object and the status of every room/building within the facility. Manage Locations Record information for all buildings/rooms, Log...

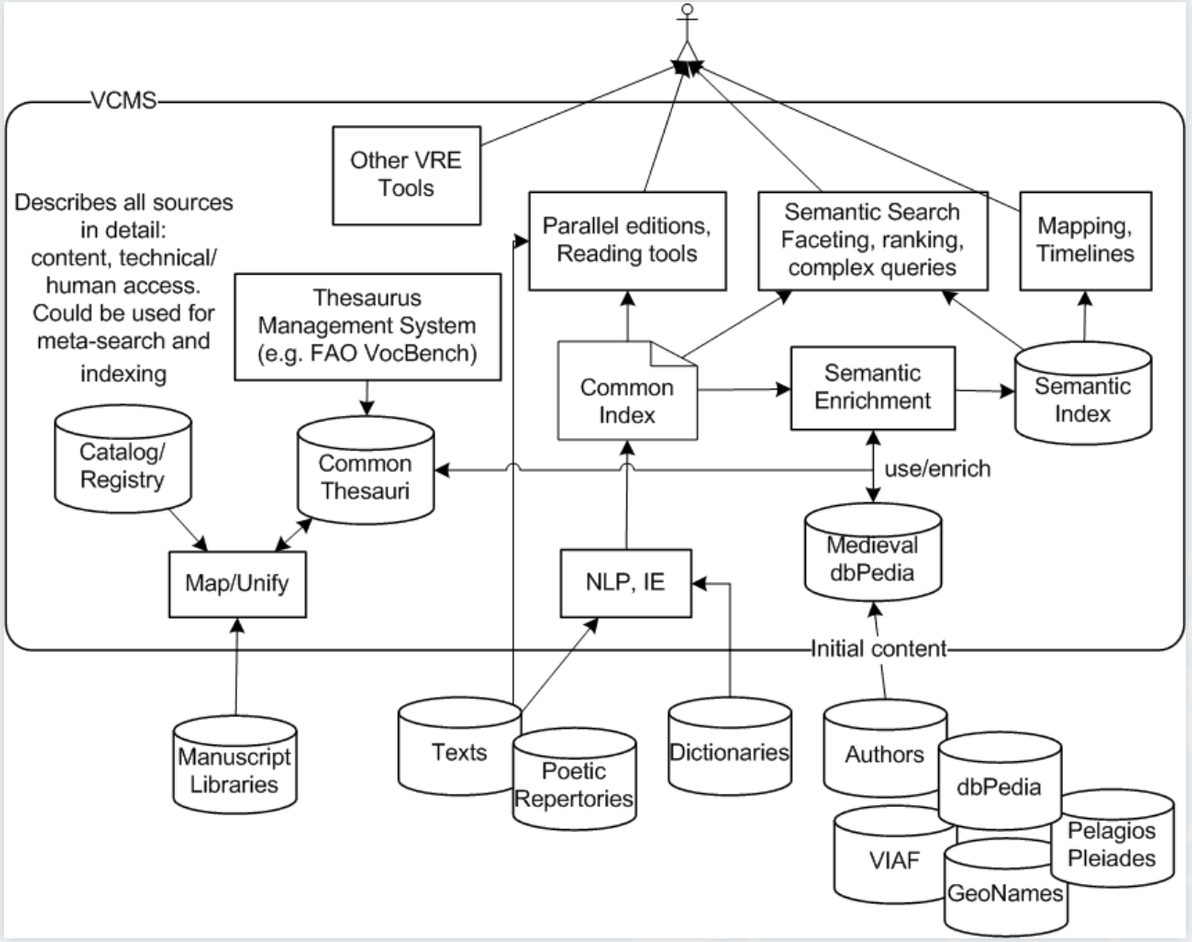
Acquisition Management allows institutions to track all current, pending, and historic acquisitions of the institution, creating an audit trail of each cultural object that enters the institution. Case for New Acquisition...

Deaccessions Management gives managers a bird's eye view of all deaccessions, past and present to easily manage and maintain records of deaccessions for the institution. Donor Records Log information relevant to...

ontotext

Virtual Center for Medieval Studies

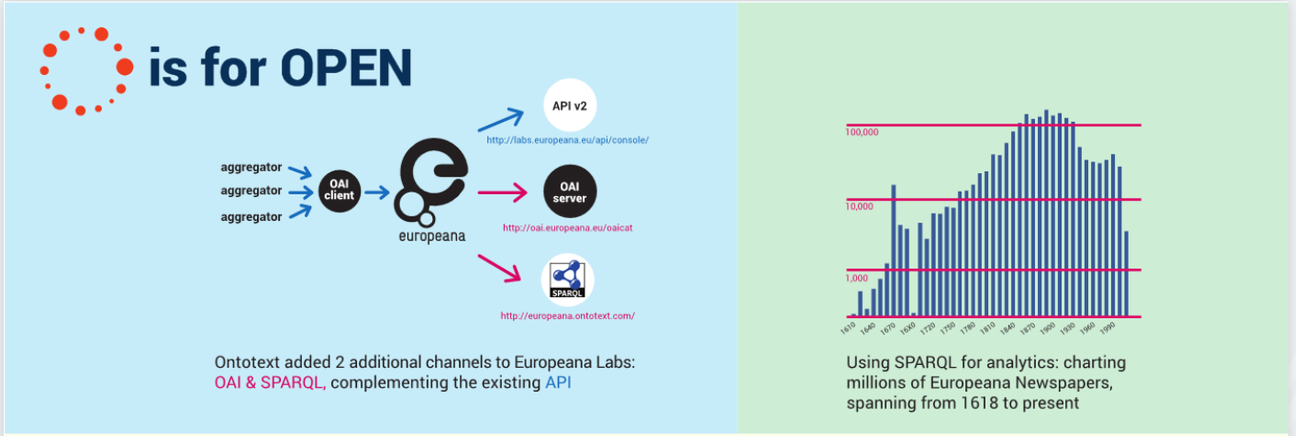
- Medieval Cultures and Technological Resources (VCMS) COST action. FET proposals for medieval lexicography, historic research, Virtual Research Environments



ontotext

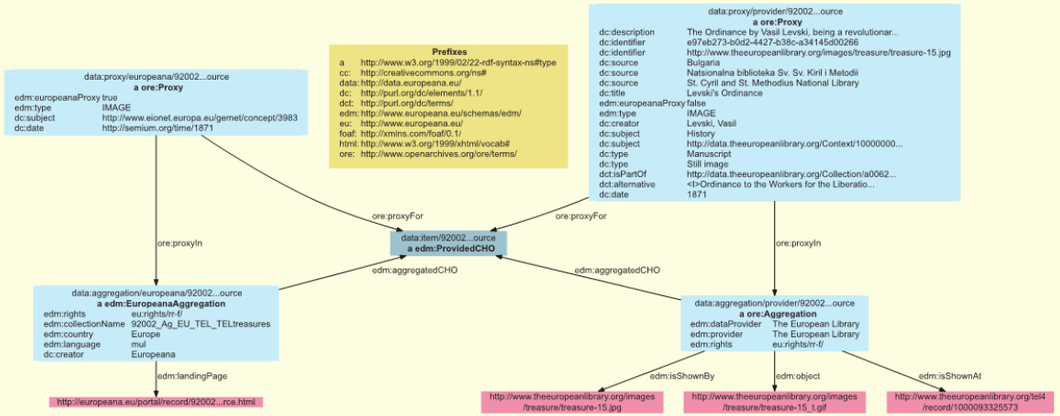
Europeana Creative

OAI PMH, and SPARQL servers for Europeana (part of Europeana Labs).



Ontotext added 2 additional channels to Europeana Labs:
OAI & SPARQL, complementing the existing API

Using SPARQL for analytics: charting
millions of Europeana Newspapers,
spanning from 1618 to present

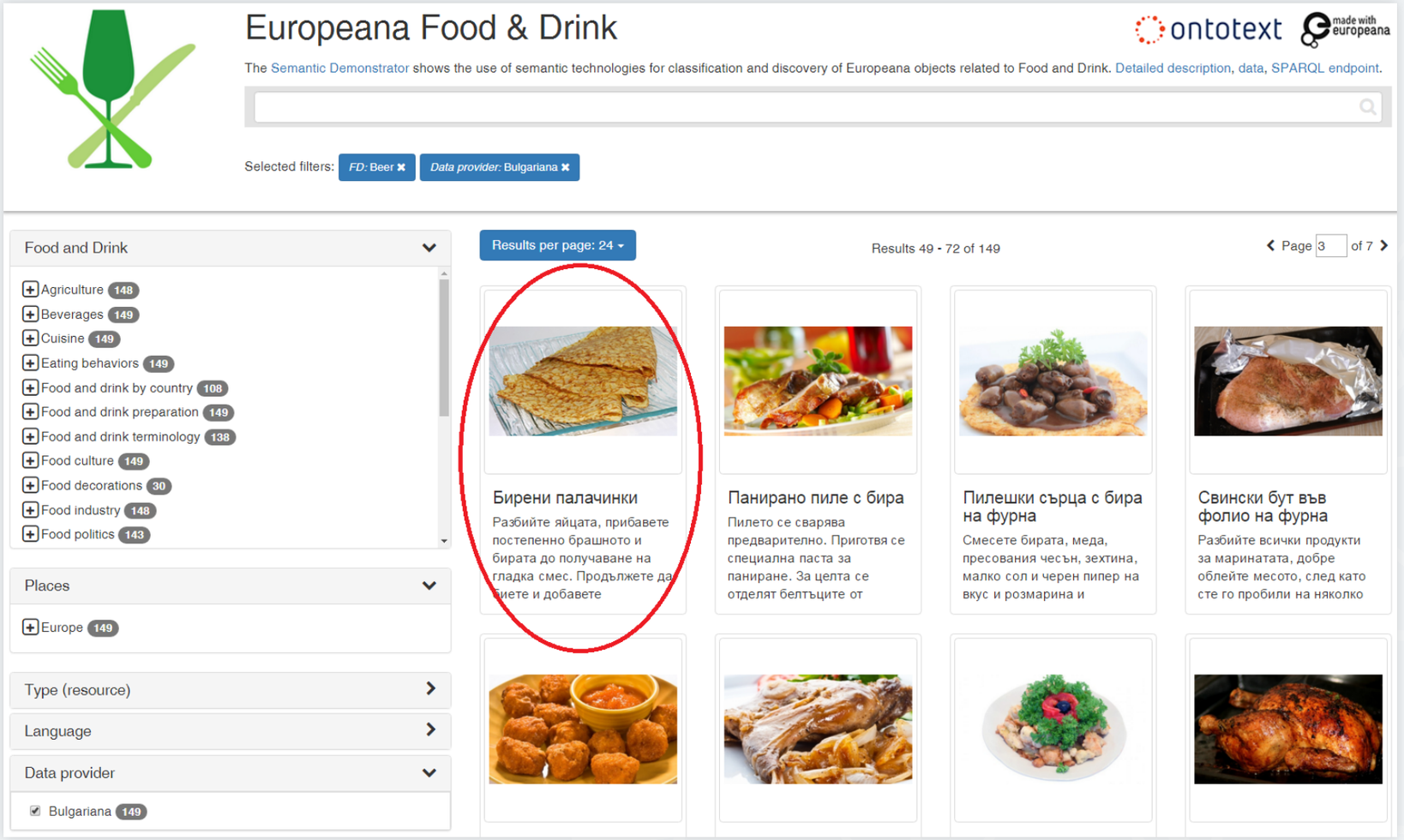


EDM (Europeana Data Model) is a semantic web format based on RDF and DC/DCT.
Fields representing the CHO (cultural heritage object) are split into 5 nodes

ontotext

Europeana Food and Drink

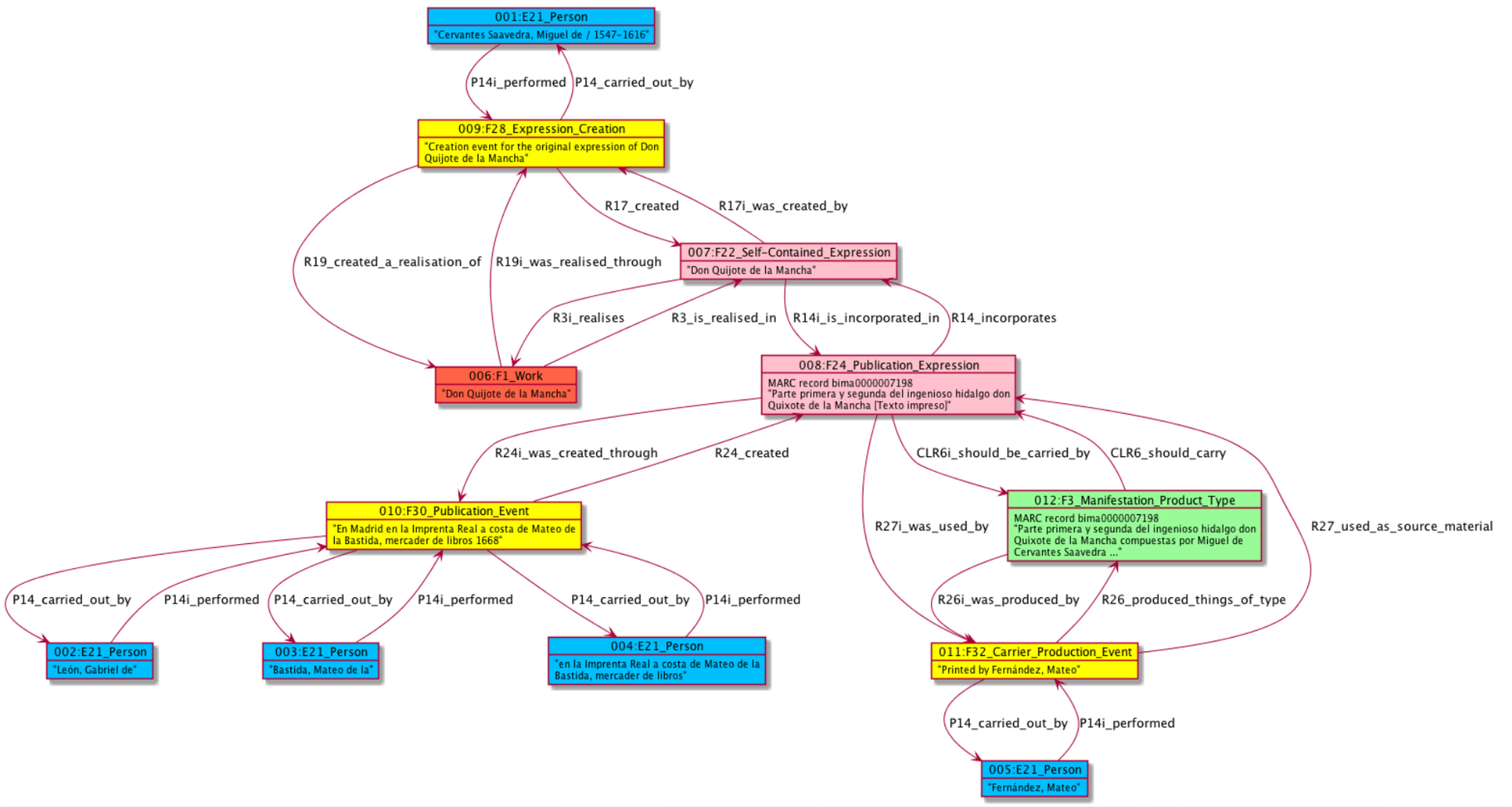
- sem app: enrichment (EN, FR, manual BG), hierarchical semantic facets, contributed BG recipes



ontotext

Europeana

- Members council, 5 task forces (e.g. below: working on FRBRoo-EDM profile), Data Quality Committee



ontotext

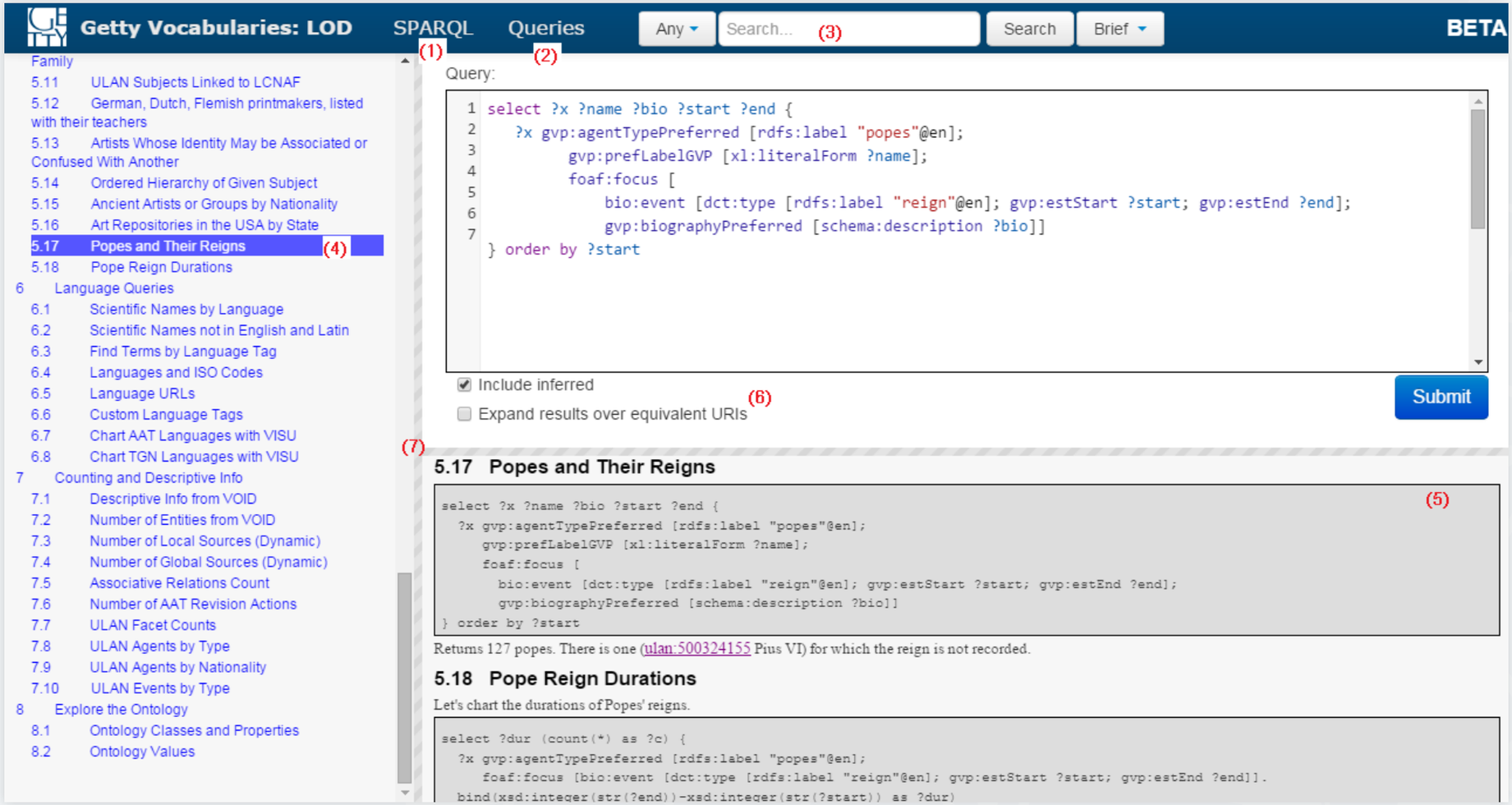
Bulgariana

- National aggegator: initiator

ontotext

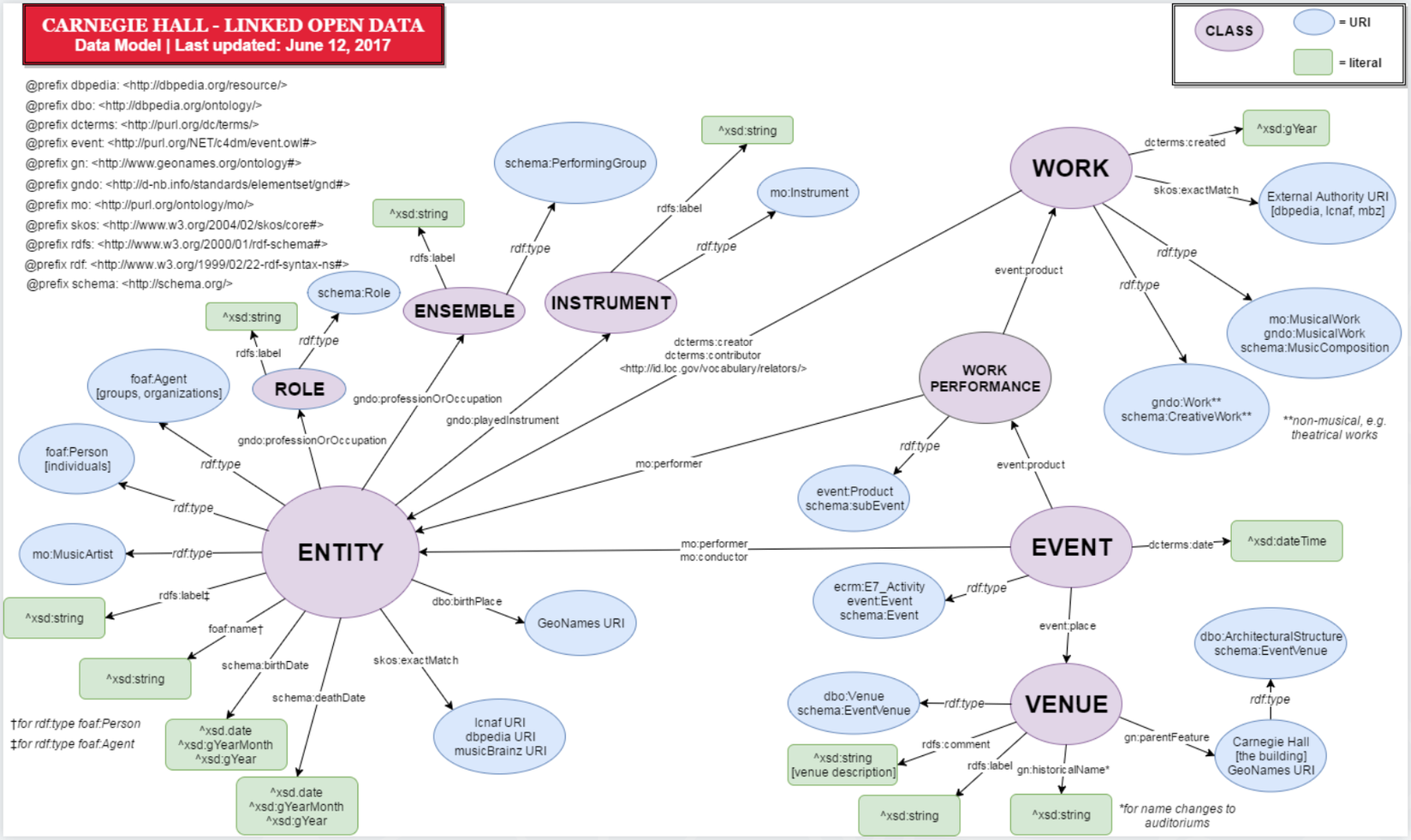
Getty Research Institute: Vocabularies LOD

- Complete services: GraphDB, ontology design, mapping, documentation, support… Helping on Getty Museum LOD



ontotext

Carnegie Hall LOD



ontotext

American Art Collaborative

Consulting: 14 US museums integrating data using CIDOC CRM. Mapping Review app

Classification
The type of object the work is.

Style
The style or period of the work.

Subject
What the work depicts.

Concept
What the work is about.

Technique
How the object was made.

Materials
What the work is made of.

Medium Text
A medium descriptive text.

'hysical Object

Current Owner
Who owns the work.

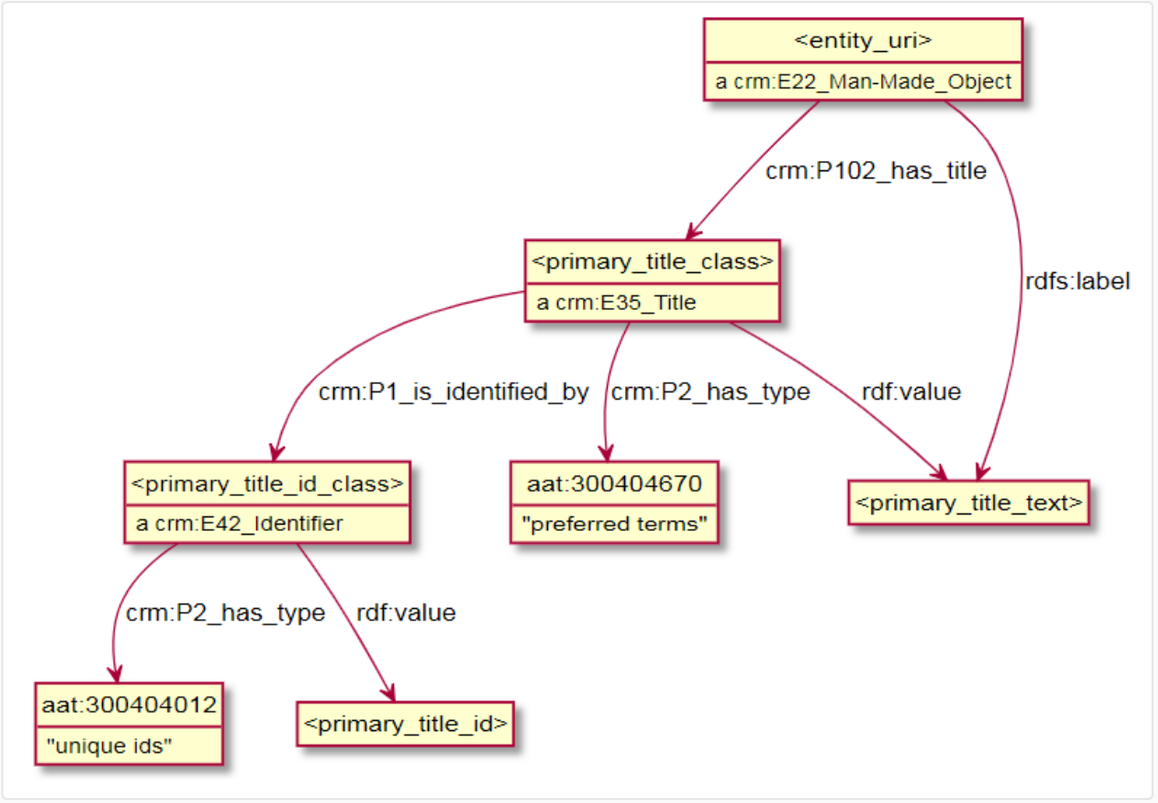
Current Location
Where the work is located.

Responsible Department
Institution Department responsible for the work.

Dimensions
Physical size of the work.

Dimensions (Part)

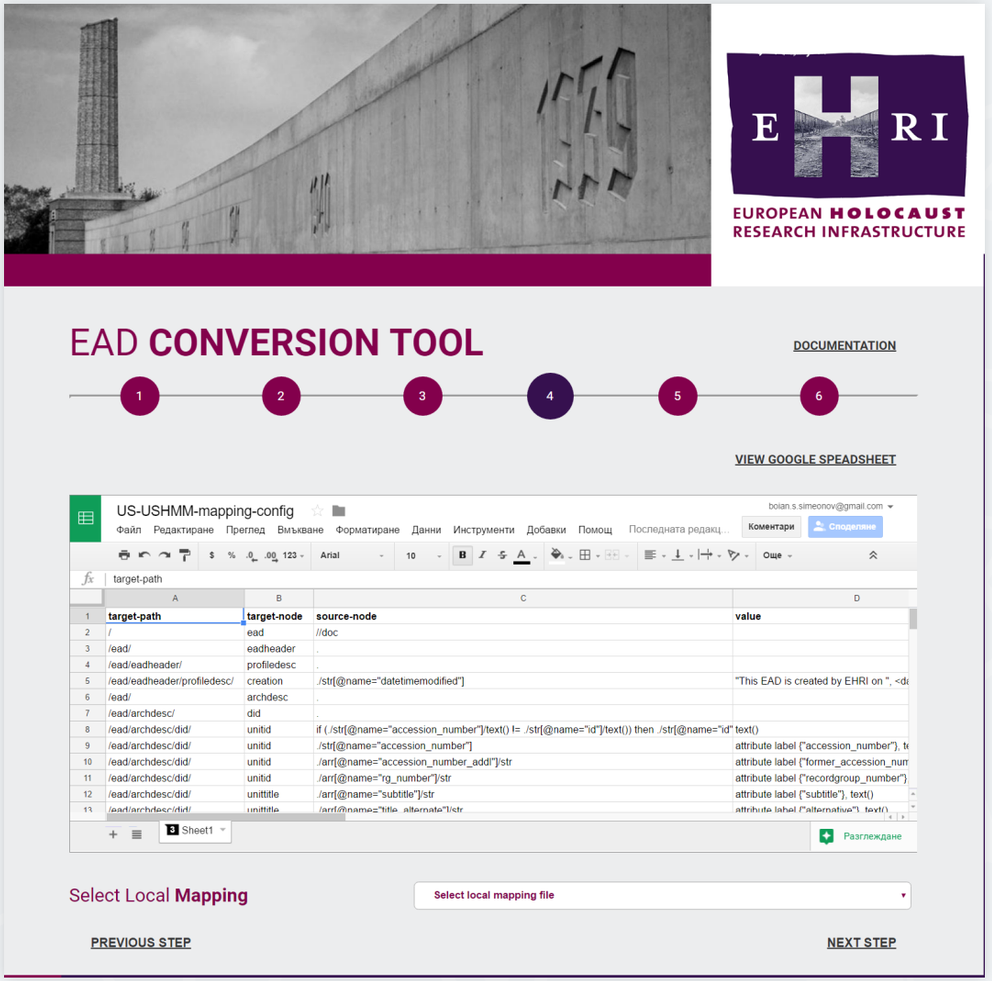
AAC Target Mapping For **Primary Title**



ontotext

European Holocaust Research Infrastructure

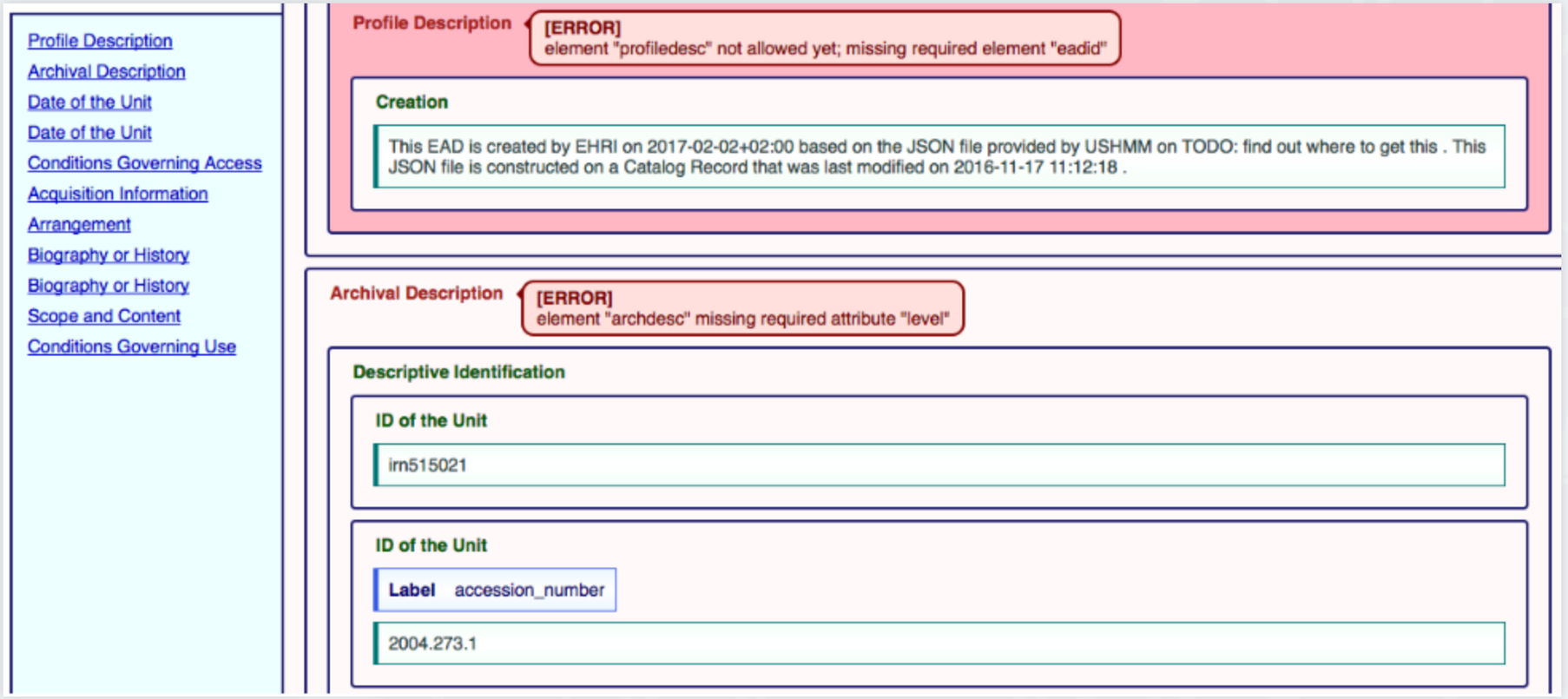
- Semantic archive integration. 4+4 years, heading towards ERIC. EAD conversion



ontotext

EHRI EAD Validation

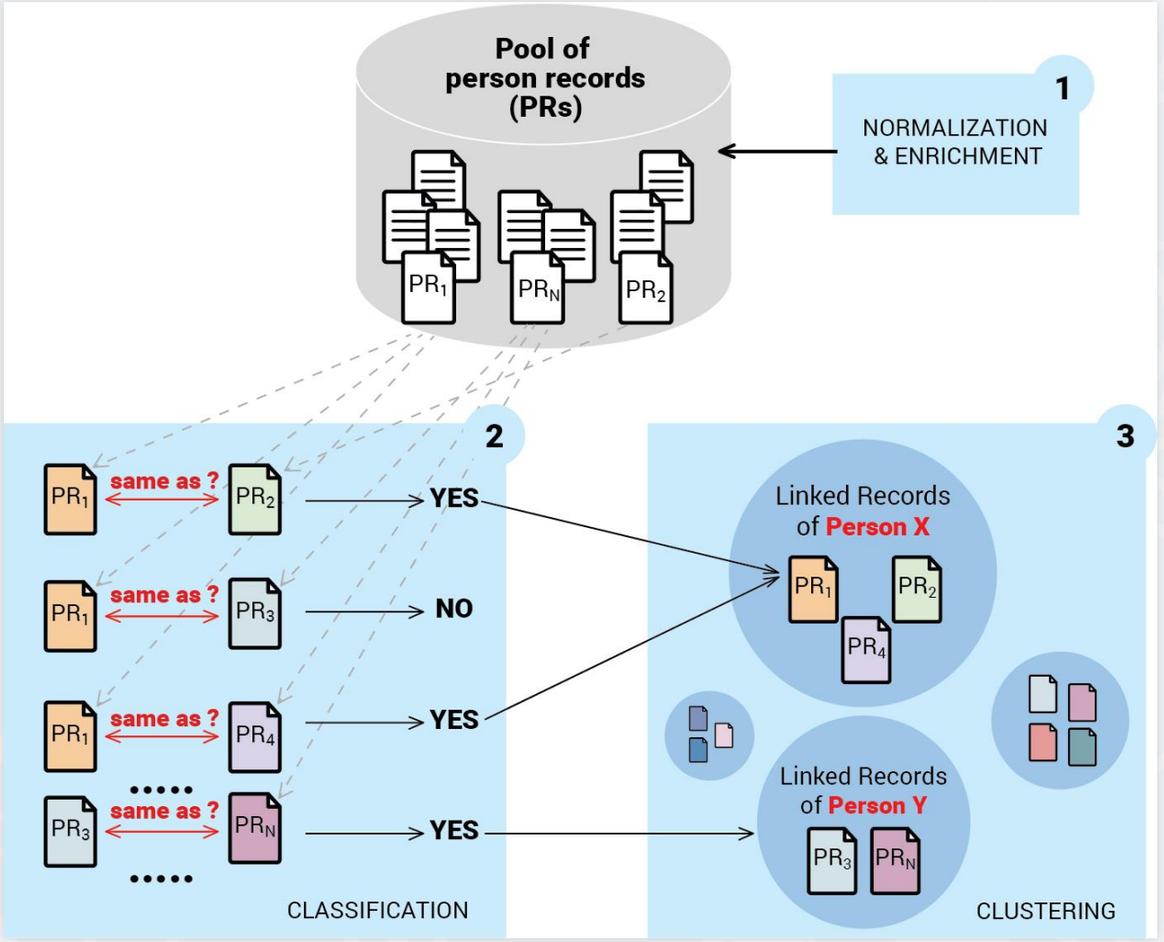
EAD validation, HTML preview and integrated error display; publishing/transport and ingest to Portal



ontotext

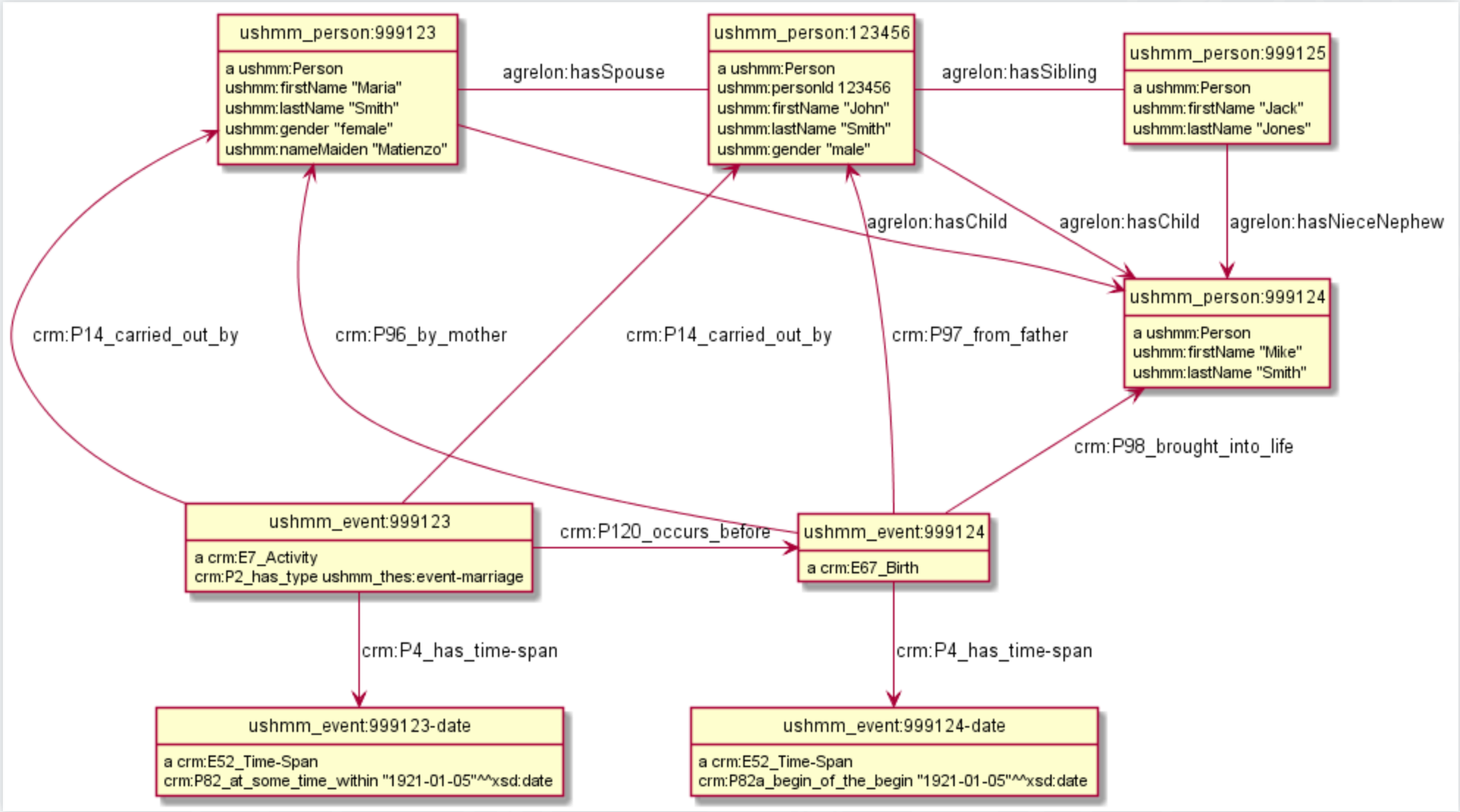
EHRI Semantic Services

- Geonames Coreferencing service (from EAD access points etc), EHRI Thesaurus (VocBench), Person Deduplication (record linking)



ontotext

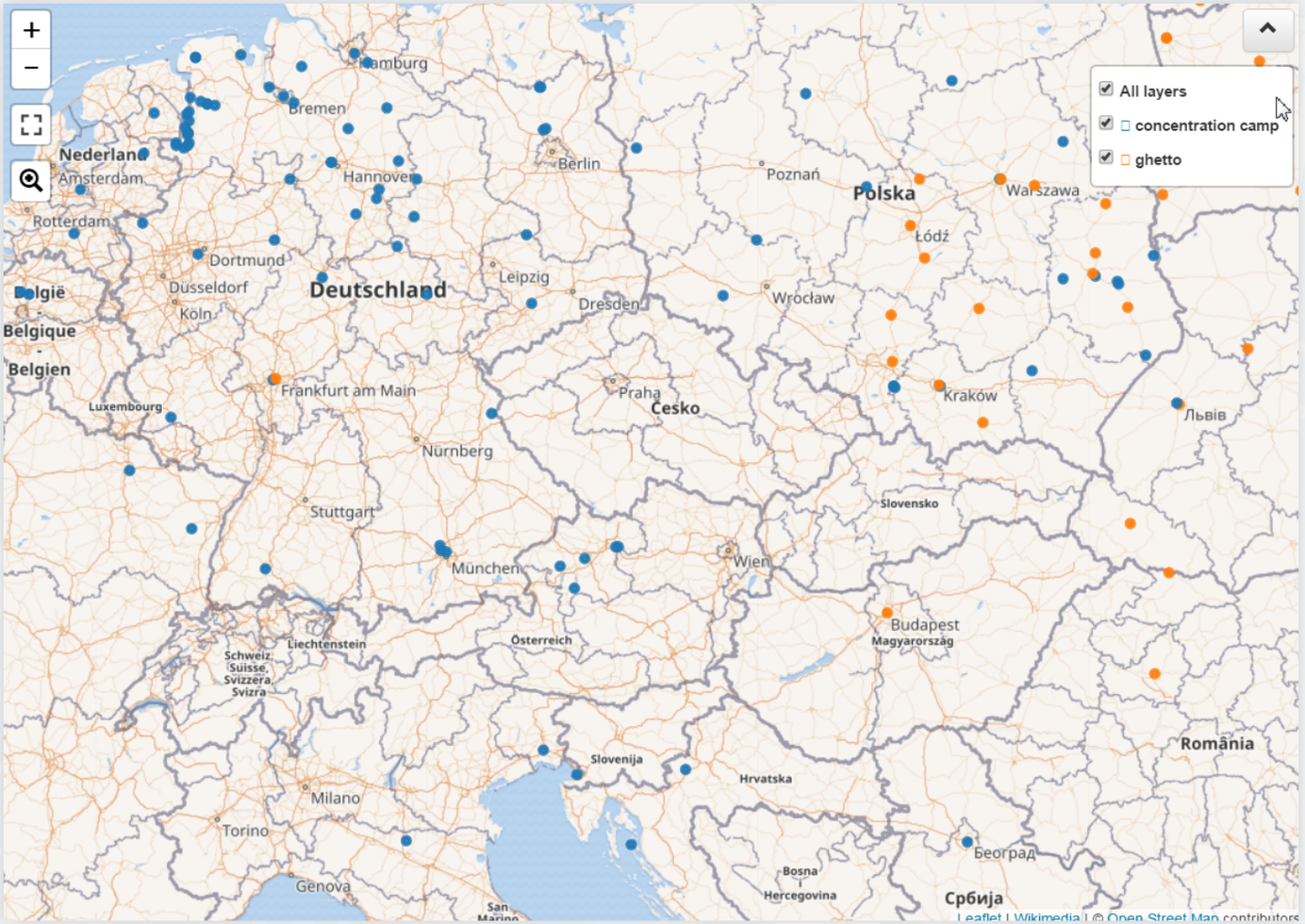
EHRI Person networks



ontotext

EHRI Camps and Ghettos

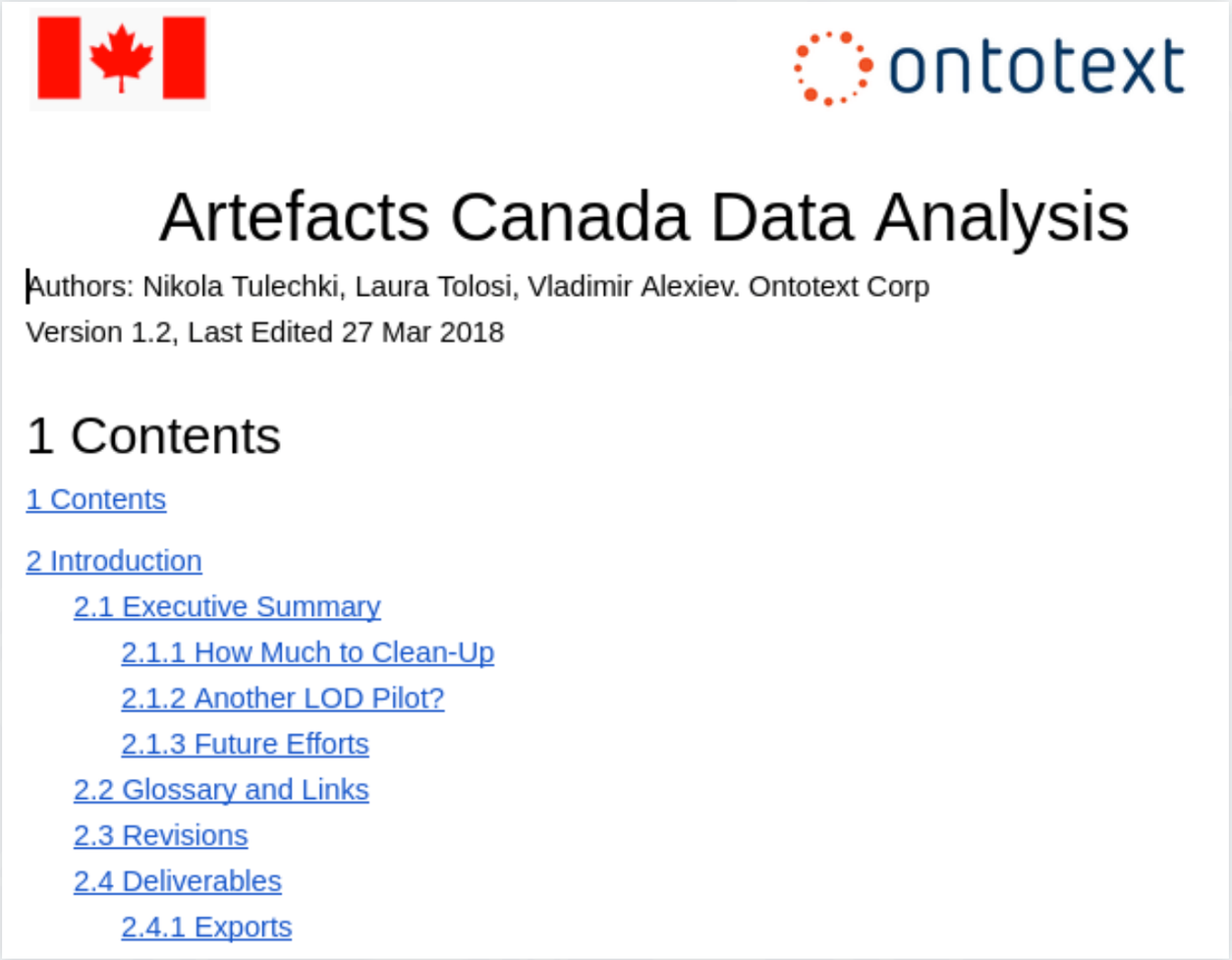
- Integrating Camps and Ghettos info between EHRI and Wikidata



ontotext

Canadian Heritage Information Network

- CA national aggregator is transitioning to LOD. 4 Consulting projects: environment scan, strategy, Artefacts Canada data analysis, national authorities



ontotext

Artefacts Canada Data Analysis

Authors: Nikola Tulechki, Laura Tolosi, Vladimir Alexiev. Ontotext Corp

Version 1.2, Last Edited 27 Mar 2018

1 Contents

Wikidata Authority Control

1-50 | 51-100 | ☐ Show unmatched | ☑ Show auto-matched | ☐ Show user-matched | ☐ Show NoWD | ☐ Show N/A | Site stats

Title/Q	Description	Actions
pyrolusite	mineral, inorganic material, <materials by composition>, materials (matter), Materials (Hierarchy Name), Materials Facet	Matched by Vladimir Alexiev
pyrolusite Q413293	Rutile mineral group, named after fire and washing; oxide mineral	Remove
sodium chlorite	sodium compounds, sodium, inorganic material, <materials by composition>, materials (matter), Materials (Hierarchy Name), Materials Facet	Matched by Vladimir Alexiev
sodium chlorite Q411294	Chemical compound; chemical compound	Remove
silanol	compounds (materials), <materials by chemical form>, <materials by form>, materials (matter), Materials (Hierarchy Name), Materials Facet	Matched by Vladimir Alexiev
Silanol Q420482	Chemical substance	Remove
gellan gum	gel, colloid (particulate material), <materials by physical form>, <materials by form>, materials (matter), Materials (Hierarchy Name), Materials Facet	Matched by Vladimir Alexiev
Gellan gum Q416694	Chemical compound, thickening agent, and polysaccharide; chemical compound	Remove
stonemasonry	<processes and techniques by material>, <processes and techniques by specific type>, <processes and techniques (processes and techniques)>, Processes and Techniques (Hierarchy Name), Activities Facet	*Not matched*
Search Wikidata | Search en.wikipedia | Google-search Wikipedias | Google-search Wikidata | Create Wikidata item		Set Q | New item | N/A

ontotext

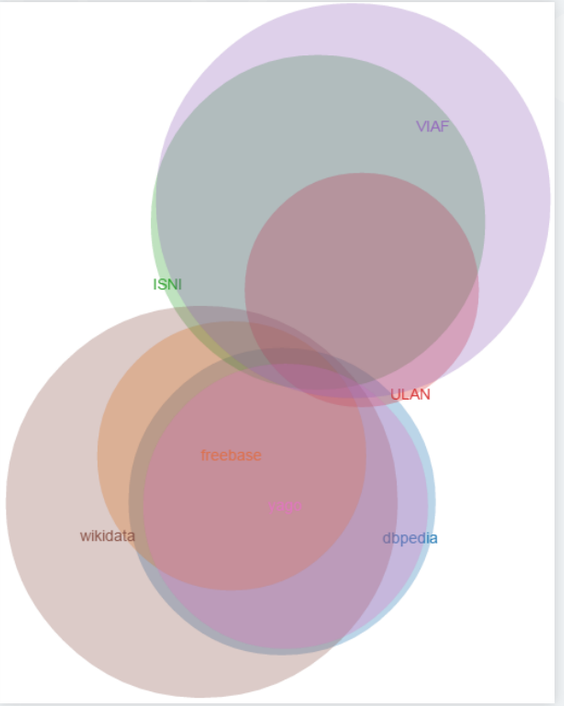
Wikidata/DBpedia vs VIAF/GND; and Europeana

Name Data Sources for Semantic Enrichment (Europeana Creative D2.4). Wikidata, a Target for Europeana's Semantic Strategy (GlamWiki 2015)

VIAF-Wikidata Coreferences for Lucas Cranach

VIAF	id in VIAF	Wikidata	id in Wikidata
viafID	49268177	VIAF	49268177
BAV	ADV10197613		
BNC	.a10853637		
BNE	XX907273		
BNF	cb12176451h	BNF	12176451h
DNB	118522582	GND	118522582
ISNI	0000000121319721	ISNI	0000 0001 2131 9721
JPG	500115364	ULAN	500115364
LC	n50020861	LCCN	n50020861
LNB	LNC10-000002573		
NDL	00436834		
NKC	jn20000700335		
NLA	000035031951		
NLI	000035532,001445575,001448179		
NLP	a16828161		
NTA	068435312	NTA PPN	068435312
NUKAT	vtls000190728		
SELIBR	182422		
SUDOC	028710010		
WKP	Lucas_Cranach_the_Elder	Many Wikipedias	
IMAGINE	T7238,T267474	Cantic	a10853637
		Commons Creator	Lucas Cranach (I)
		Commons category	Lucas Cranach d. Ä.
		Freebase	/m/0kqp0
		RKDartists	18978
		SIMBAD	CRANACH, Lucas the Elder
		Your Paintings	lucas-the-elder-cranach

→ Can be leveraged to **fill the gaps**, e.g. bring RKDartists into VIAF



ontotext