SEMANTIC TECHNOLOGIES FOR CULTURAL HERITAGE

VLADIMIR.ALEXIEV@ONTOTEXT.COM

2014-08-21, MALMO, SWEDEN

2D interactive version, pdf, slideshare.

Press O for overview, H for help.

Proudly made in plain text with reveal.js, org-reveal, org-mode and emacs.

TABLE OF CONTENTS

SEMANTIC TECHNOLOGIES

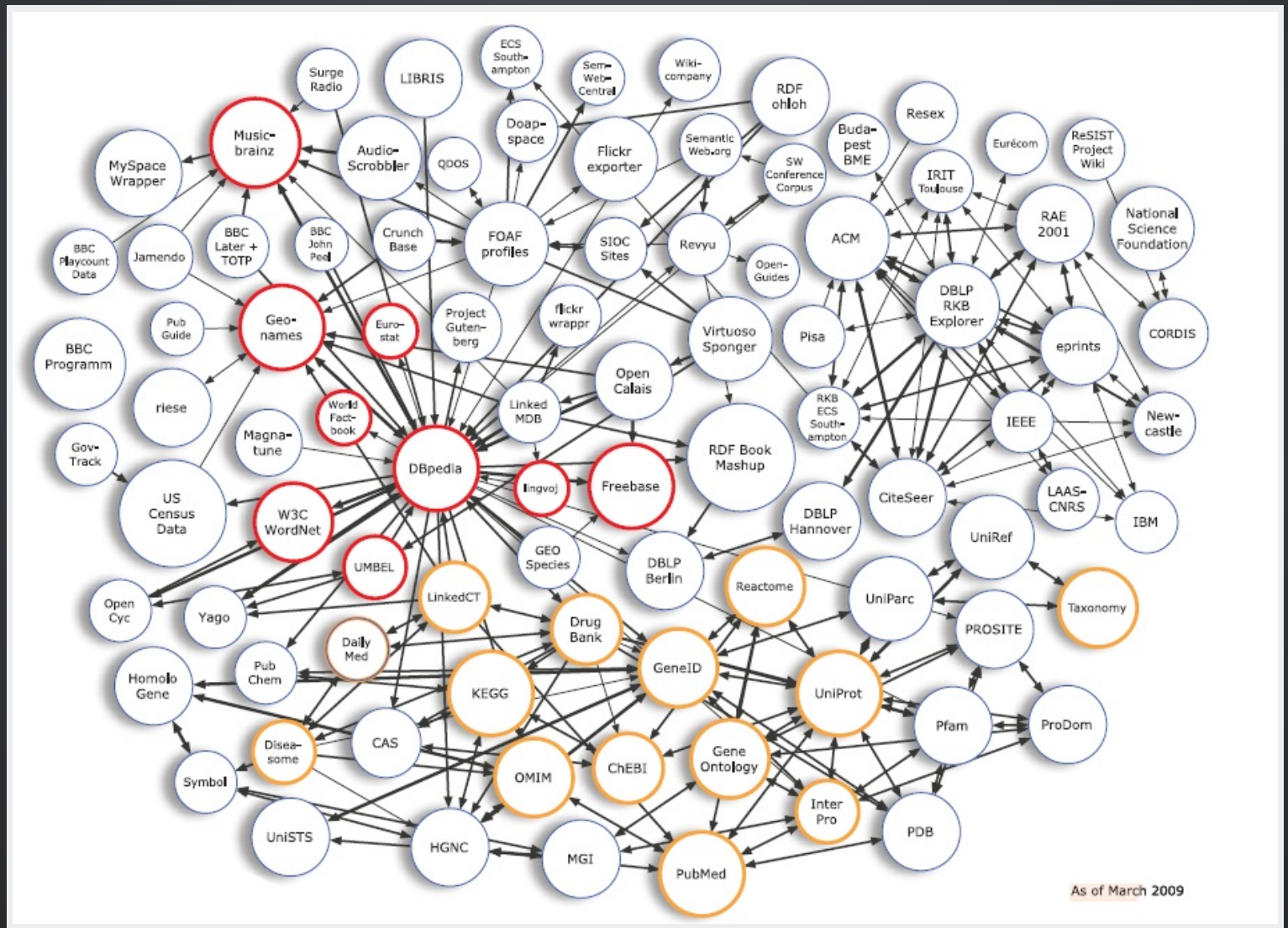
- **Web 1.0**: hyperlinked documents (World Wide Web)
- **Web 2.0**: interactive applications, the Social Web
- **Web 3.0**: interlinked data (Global Giant Graph)

Is this something new?

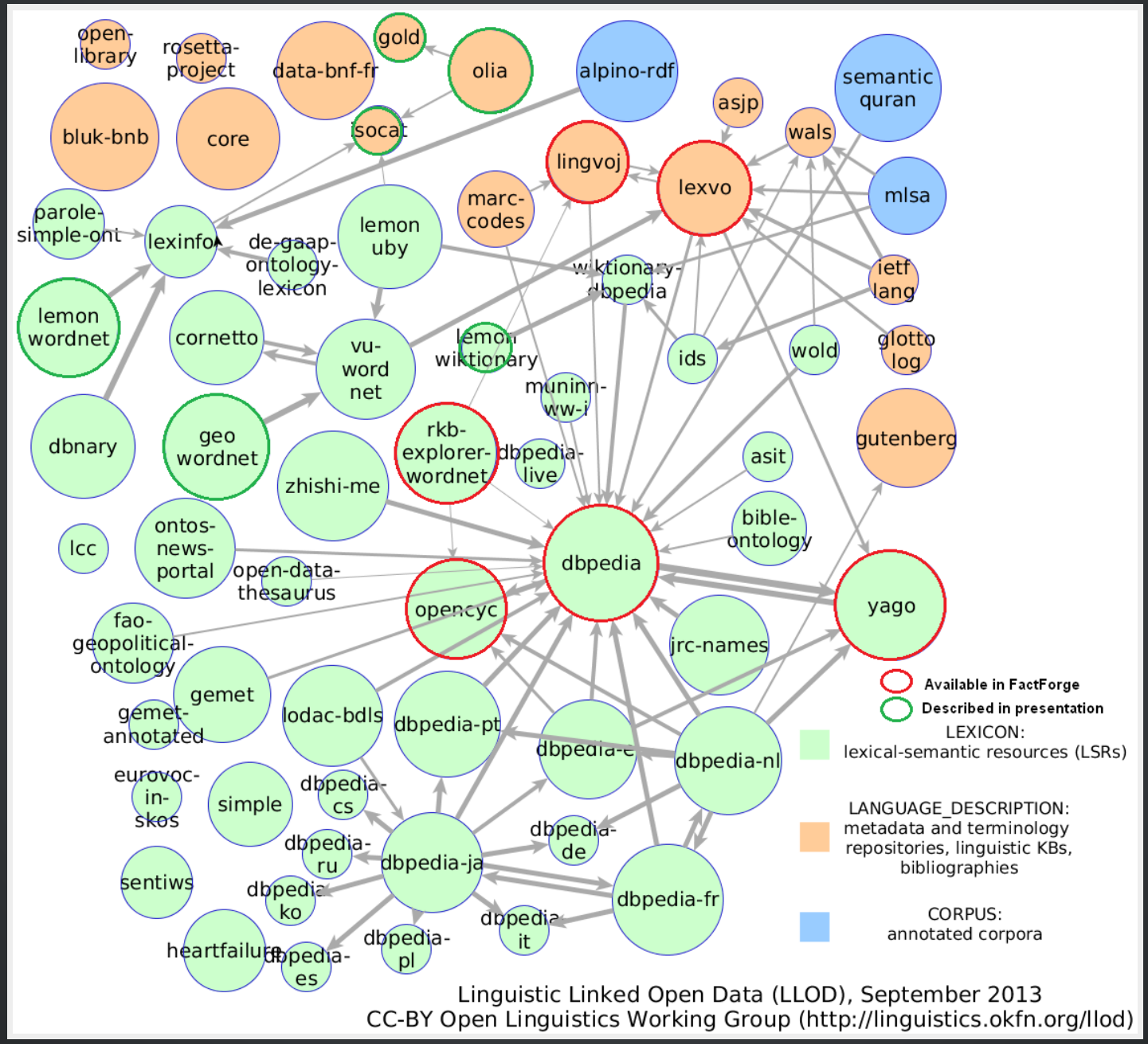
- It was all envisioned by Sir Tim Berners-Lee 25 years ago
- Standardized by W3C: both HTML and sem web standards (RDF, RDFS, OWL, SPARQL…)
- Great flurry of sem tech activity in the last 15 years
- Buzzwords: Big Data, Semantic Analytics, Concept Extraction, Sentiment Analysis…

LINKED OPEN DATA CLOUD

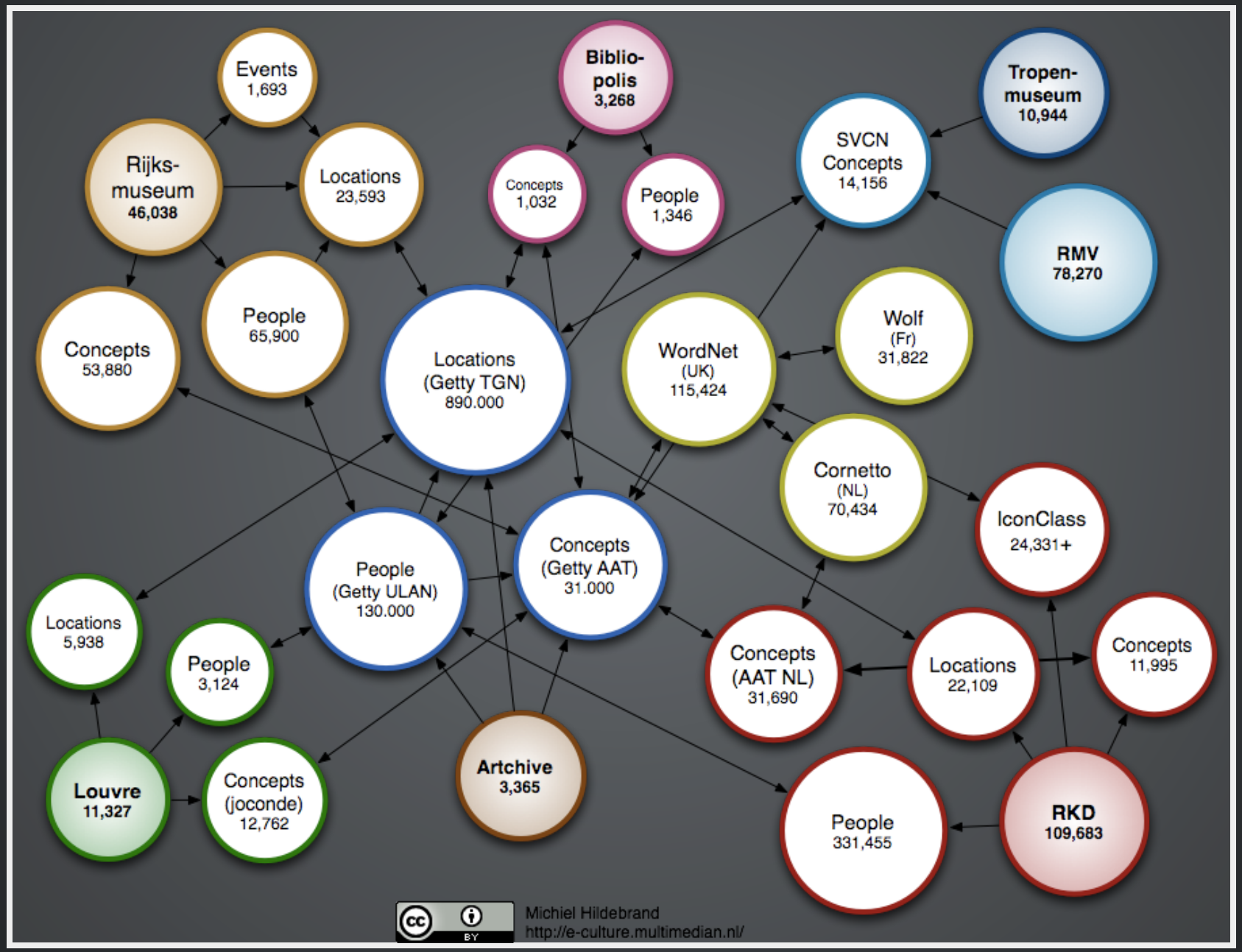
- Parts at factforge.net, linkedlifedata.com. Grown 10x since Mar 2009!



LINGUISTIC LINKED DATA



CULTURAL HERITAGE LINKED DATA



EUROPEANA RECOGNIZES IMPORTANCE OF SEMANTIC TECHNOLOGIES

europeana
think culture

Press Release

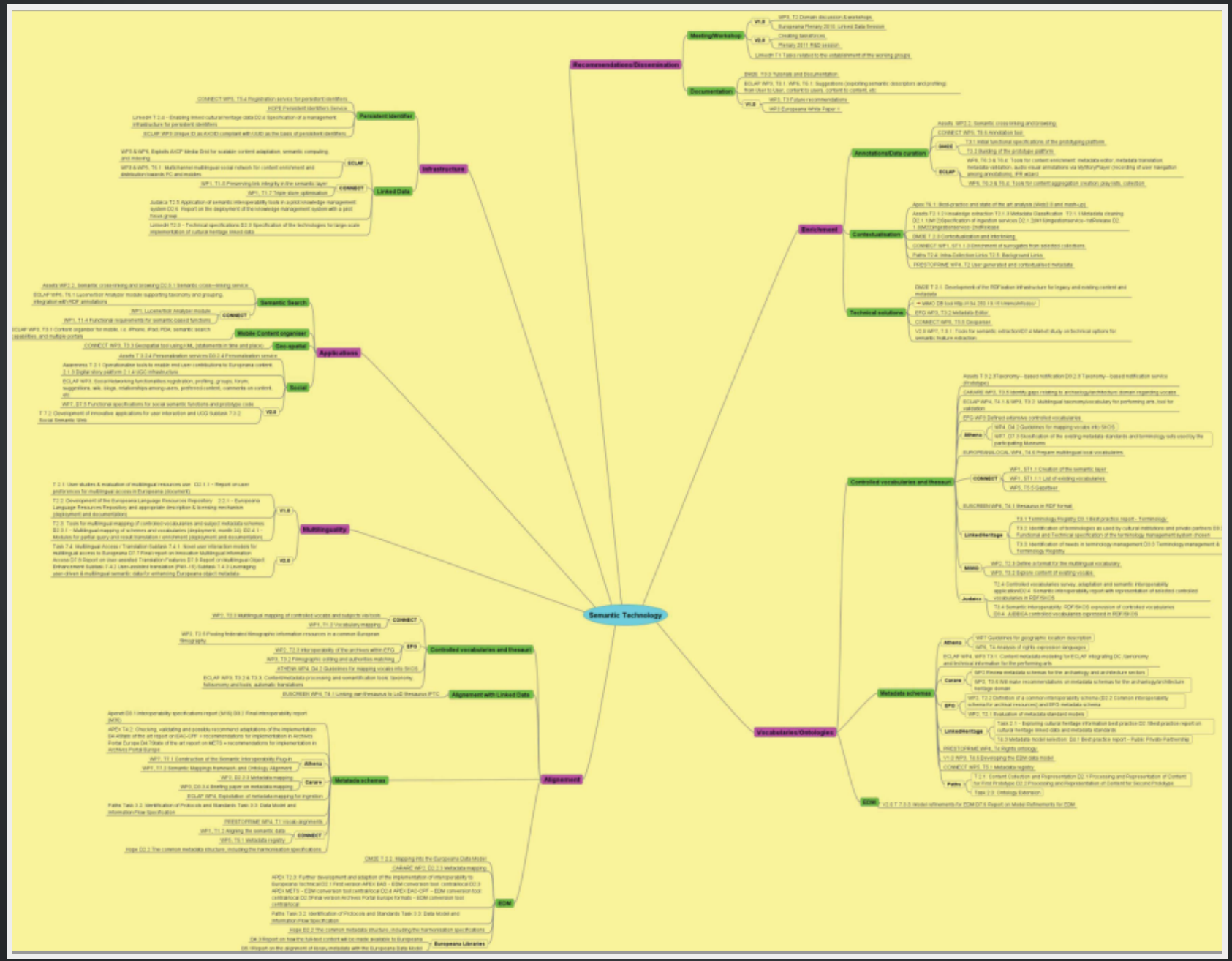
The Hague – June 1st, 2010 For immediate release

Europeana Publishes First White Paper: Knowledge = Information In Context

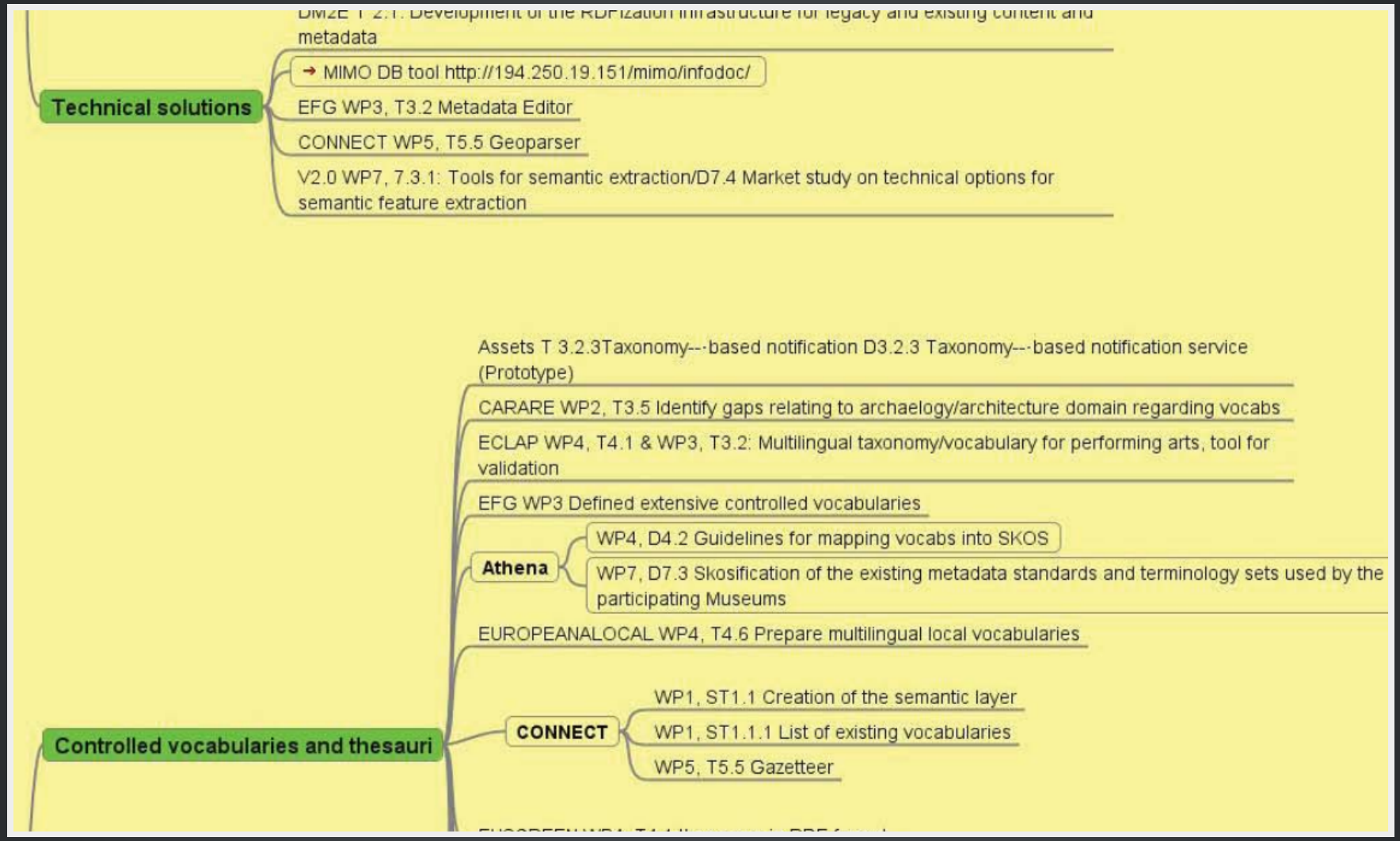
Europeana is proud to announce its first White Paper: a look at the key role linked data will play in Europeana's development and in helping Europe's citizens make connections between existing knowledge to achieve new cultural and scientific advances.

Linked data gives machines the ability to make associations and put search terms into context. Without it, Europeana could be seen as a simple collection of digital objects. With linked data, the potential is far greater, as the author of the white paper, Prof. Stefan Gradmann, explains.

EUROPEANA SEM TECH MINDMAP



EUROPEANA SEM TECH MINDMAP DETAIL



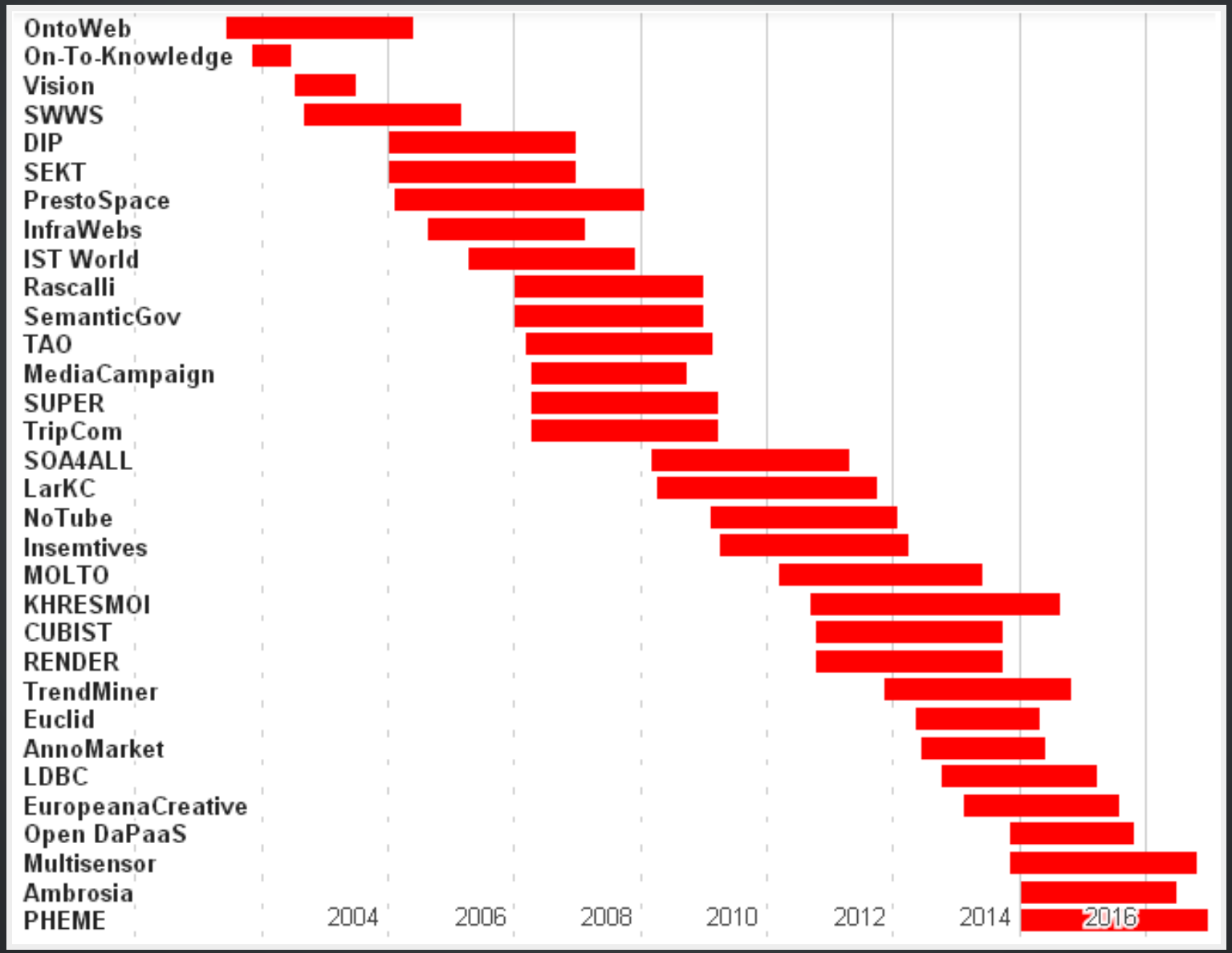
ONTOTEXT CORP

- World Leading semantic technology developer
  - Working in this area since 2000 as part of Sirma Group
  - Spun off in 2008 after venture investment (NEVEQ)
  - 75 employees: Bulgaria (Sofia and Varna), UK, USA (Washington DC)
  - Global leader in semantic databases, semantic annotation and search
- Proven Delivery
  - Highest profile sem web applications
  - BBC: World Cup 2010, London Olympics 2012, all of BBC sport…
  - Dynamic Semantic Publishing: Master Publishing platform
  - Semantic search for multinational pharmaceuticals (eg Astra Zeneca)
- Stable and Growing, both staff and revenue

SOME ONTOTEXT CLIENTS

ONTOTEXT RESEARCH PROJECTS (FP5-FP7)

- Bulgaria's largest participant: over 30 projects



CURRENT RESEARCH PROJECTS

- EUCLID : Educational Curriculum for the usage of Linked Data
  - Professional training curriculum for data practitioners aiming to use Linked Data in their daily work.
  - Strongly relevant to CH metadata specialists and other experts focusing on Linked Open Data
- AnnoMarket : Cloud-Based Text Annotation Marketplace
  - Open marketplace for pay-as-you-go, cloud-based extraction resources and services
  - Multilingual semantic entity extraction from CH text (e.g. museum object descriptions) is important and largely unsolved
- LDBC : Linked Data Benchmark Council
  - NPO for publishing and auditing benchmark results for graph and RDF databases.
  - CH institutions that decide to use repositories require such info, and can provide meaningful use cases

CURRENT RESEARCH PROJECTS (2)

- Europeana Creative : Re-use of cultural heritage metadata and content by the creative industries.
  - Contribution to improving the usefulness and kick-starting the professional use of Europeana data
  - Ontotext plays a core technological role, helping to fulfill 3 Europeana technical KPIs
- Europeana Food and Drink : explore and celebrate European cultural identity through its culinary and social history
  - Ontotext works on culinary culture classification scheme, semantic representation and storage, semantic text analysis, and semantic application

CURRENT RESEARCH PROJECTS (3)

- MultiSensor : Multidimensional content integration
  - Mine heterogeneous content using multilingual technologies with sentiment, social and spatiotemporal competence
  - Application of Linguistic Linked Data
  - Relevant to text and multimedia CH content
- DaPaaS : Data Publishing through the Cloud
  - Data- and Platform-as-a-Service Approach for Efficient Data Publication and Consumption
  - Useful for converting and hosting your Linked Open Data, and implementing Open Data Portals
- Pheme : Computing Veracity Across Media, Languages, and Social Networks

SOME ONTOTEXT PRODUCTS

- GraphDB (OWLIM)
- KIM Semantic Annotation
- Master Publishing Platform
- PROTON Ontology

GRAPHDB (OWLIM)

- High-performance semantic repository created by Ontotext
- Reasoning and query evaluation are performed over a persistent storage layer.
- Loading, reasoning and query evaluation are fast even against complex ontologies and huge knowledge bases
- Can manage billions of statements on desktop hardware, 10s of billions on commodity server hardware
- Pure Java implementation, ensuring ease of deployment and portability
- Compatible with Sesame (OpenRDF), which brings interoperability benefits and support for all major RDF syntaxes and query languages
- Compatible with Jena through a built in adapter layer
- Enterprise-grade
- Used by important commercial clients (see slide above)
- Found a great following in the CH domain (see later)

TODO GRAPHDB FEATURES

- High-performance reasoning: RDFS, OWL-Horst, OWL2 RL, QL
- Custom rule-sets allow tuning for optimal performance and expressivity
- Optimized owl:sameAs handling: dramatic improvements for data integrated from multiple sources
- Clustering: resilience, fail-over and scalable parallel query processing
- Geo-spatial extensions for fast geo queries over WGS84 data
- Full-text search support, based on either Lucene or proprietary search techniques
- High-performance retraction of statements & inferences
- Expressive consistency & integrity constraint checking mechanisms
- Notification mechanism, to allow clients to react to statements in the update stream

NEW GRAPHDB FEATURES

- GraphDB-Workbench with improved management
- JMX-based management and control interfaces
- Cluster deployment and testing tool
- Cluster operational improvements
- Explain Query Plans
- Rule profiling
- Support for external plug-ins. Loaded from the classpath, handle custom functions & predicates
- Connectors that synchronize RDF data to provide extremely fast full-text and facet searches:
  - Elasticsearch GraphDB Connector
  - Lucene GraphDB Connector
  - Solr GraphDB Connector

KIM SEMANTIC ANNOTATION AND SEARCH

- Built on top of GATE
- Ontotext is the largest commercial contributor to GATE
- Used by important commercial clients: BBC, UK Press Association, NDP, Oxford University Press, Financial Times, Euromoney…

Large-scale semantic annotation based on:

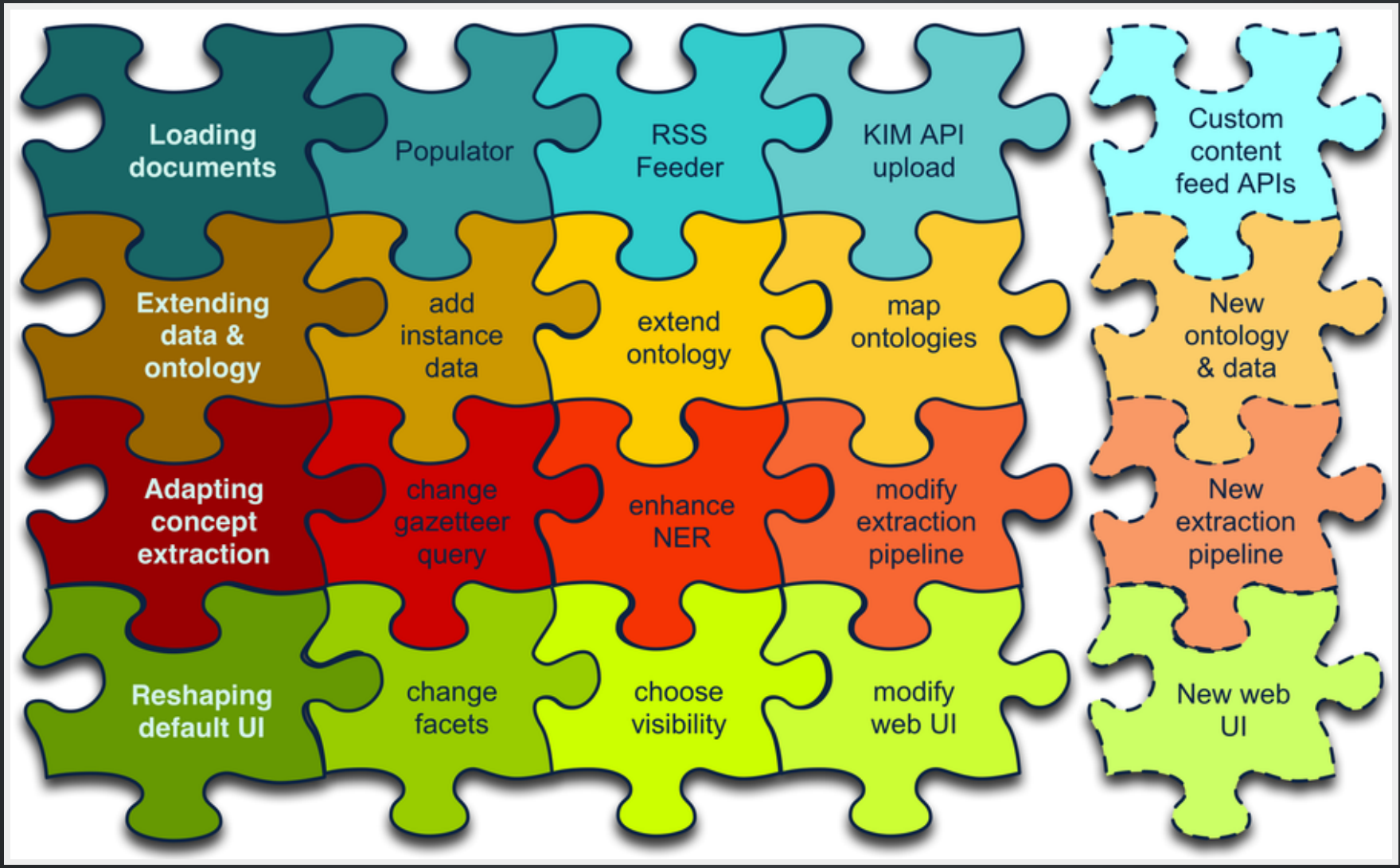
- Assembling a semantic knowledge base of a domain
- Creating annotation guidelines and a Gold Standard Corpus
- Machine learning

Involves:

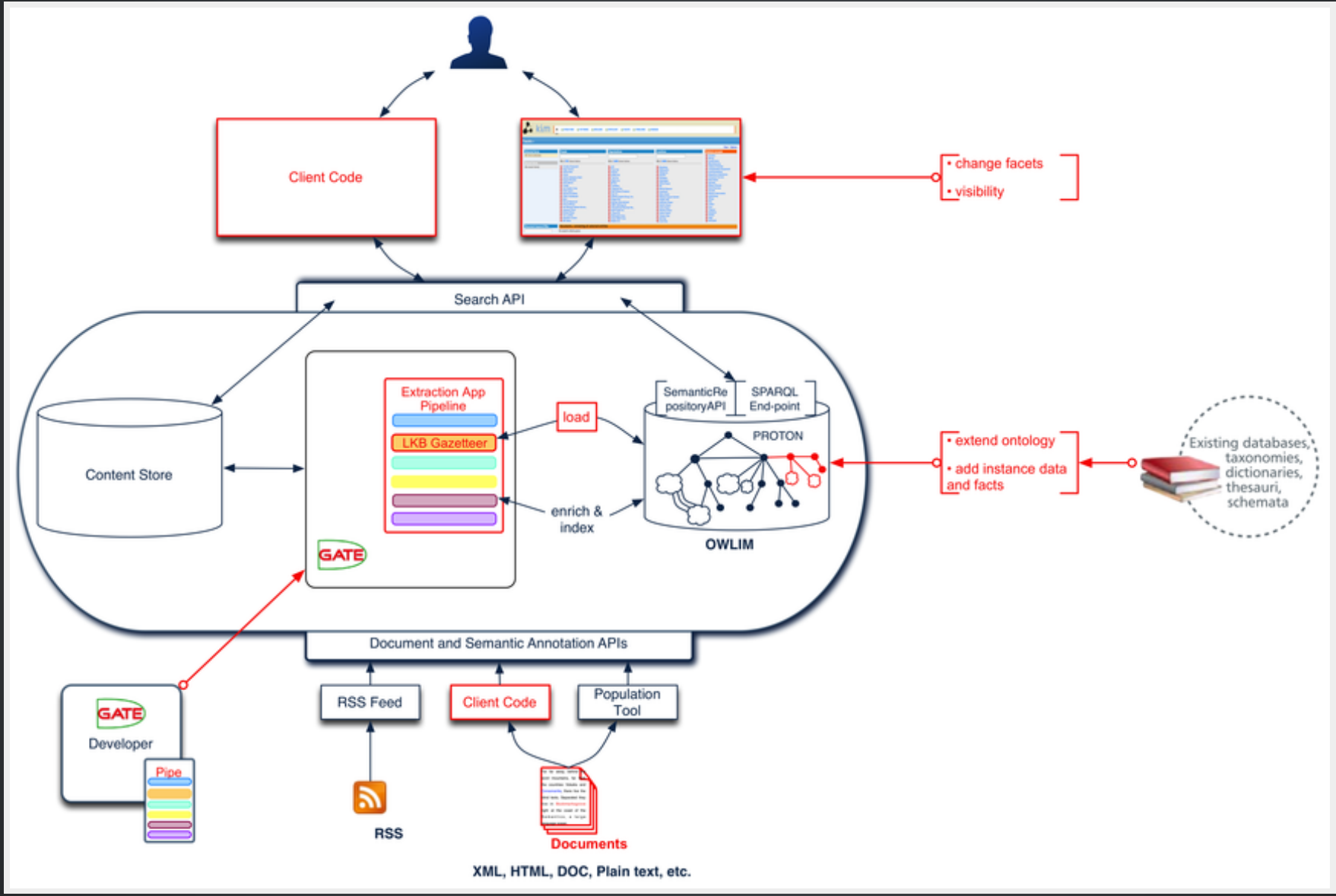
- Named Entity Recognition
- Semantic Disambiguation
- Concept Extraction
- Relation Extraction
- Event Extraction

KIM CUSTOMIZATION

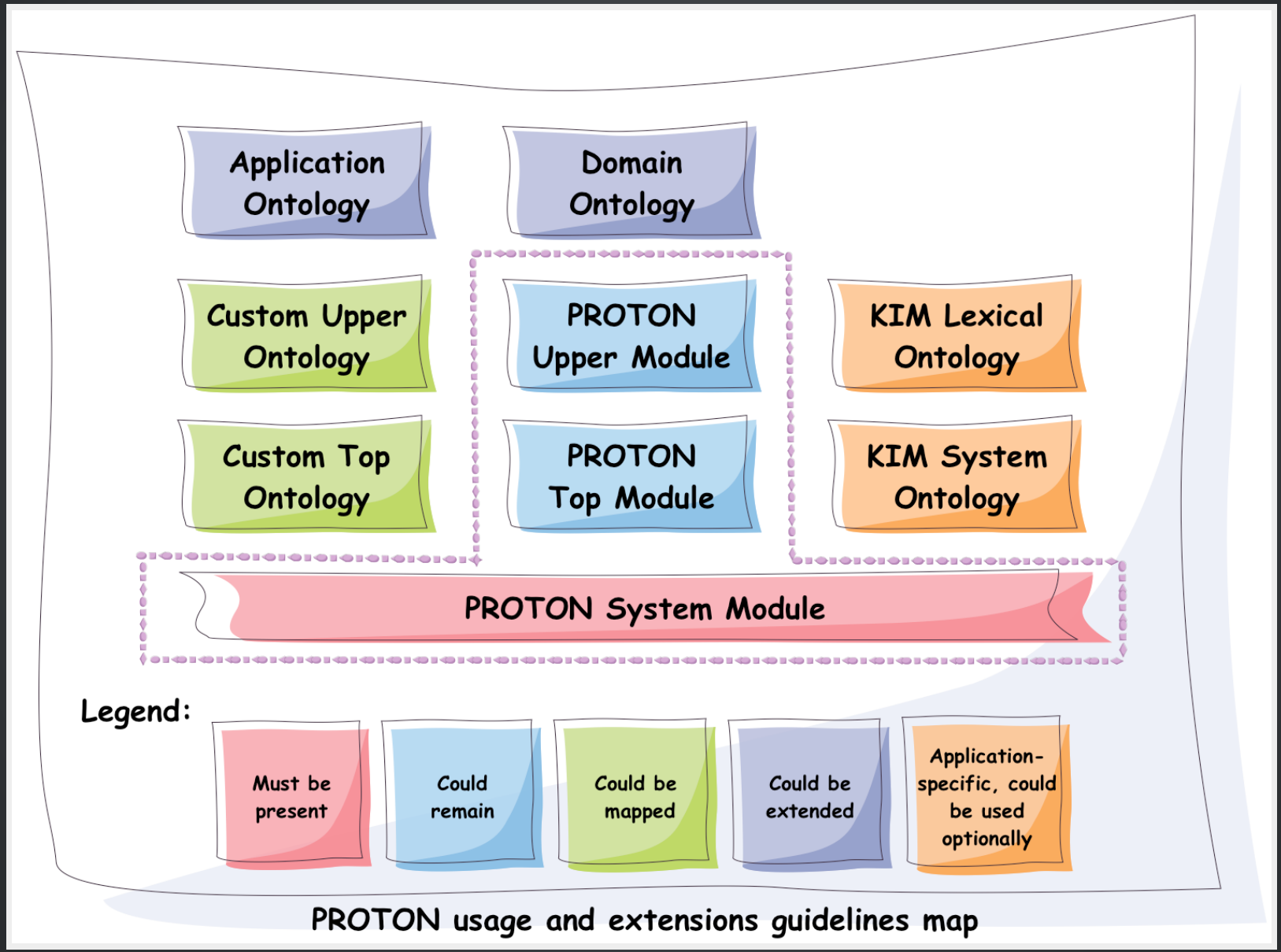
KIM Semantic Solutions describes the various parts of KIM that can be customized



MASTER PUBLISHING FRAMEWORK



PROTON UPPER ONTOLOGY



ONTOTEXT MORE INFO

See more info including brochures, cases etc

For over a decade Ontotext has brought together metadata and content to search, navigate and analyze information in more productive ways. After extensive product development and self-funded R&D, Ontotext products now serve a diverse range of clients. Our clients include world renowned media agencies like the BBC and the Press Association, top pharma companies such as AstraZeneca, important government agencies including the US DoD, The National Archive of UK, US Medicare, and cultural institutions like the British Museum.

Ontotext clients benefit from our sound methodologies, industry standard-compliant products and professional services to address their business objectives and realize their desired solution. We balance our long-term research goals with our current industrial activities to ensure we are useful now and in the long run.

Company Info and Overview (brochures)

- Ontotext Overview, June 2013
- KIM, May 2011
- OWLIM, June 2011
- Life Sciences, June 2011
- Web Mining Framework, June 2011
- Cultural Heritage, June 2013
- UK Job Market Database, Feb 2013

Presentations

- Bringing the Semantic Web Closer to its Tipping Point (Feb 2012)
- Linked Data for the Enterprise - Opportunities and Challenges
- Metadata management for the BBC's 2010 World Cup site using OWLIM
- Semantic Technologies for Big Data
- Practical usage of linked data and semantic annotations
- Ontotext Discovery and Clinical Solutions
- OWLIM@AWS - on-demand RDF data management in the Cloud

Showcases and Demos

- Latest News - a semantic search engine, using text analysis to provide hybrid queries involving structured data and inference.
- Linked Life Data - a platform for semantic data integration trough RDF warehousing and efficient reasoning that helps to resolve conflicts in the data.
- FactForge - a public service that represents a reason-able view to the web of data.
- Lupedia - a Text Enrichment service, which does lookup, but no entity recognition.

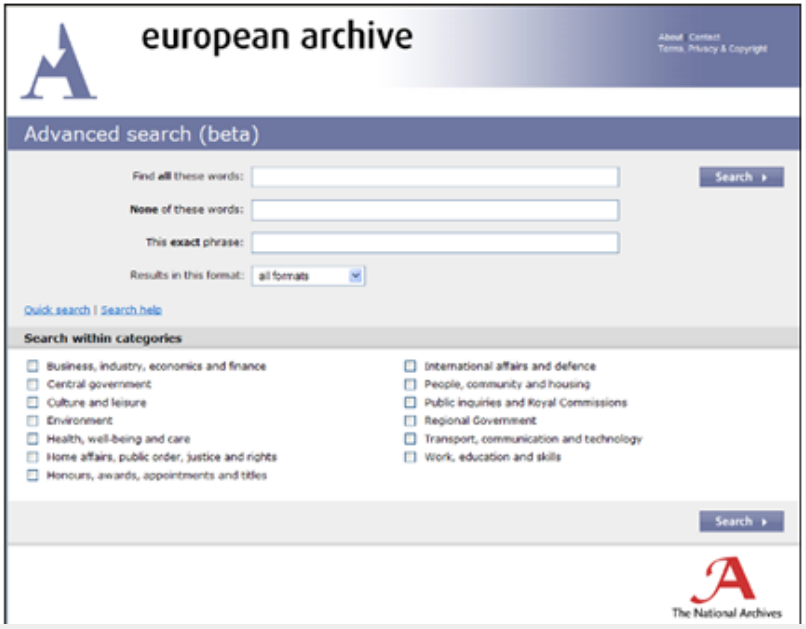
Semantic Solutions Docs Collection

ONTOTEXT GLAM PROJECTS

- UK National Archives: Semantic Knowledge Base
- Europeana Creative
- Europeana Food and Drink
- Bulgariana
- GraphDB CH installations (endpoints)
- ResearchSpace
- Getty LOD

UK NATIONAL ARCHIVES: SEMANTIC KNOWLEDGE BASE

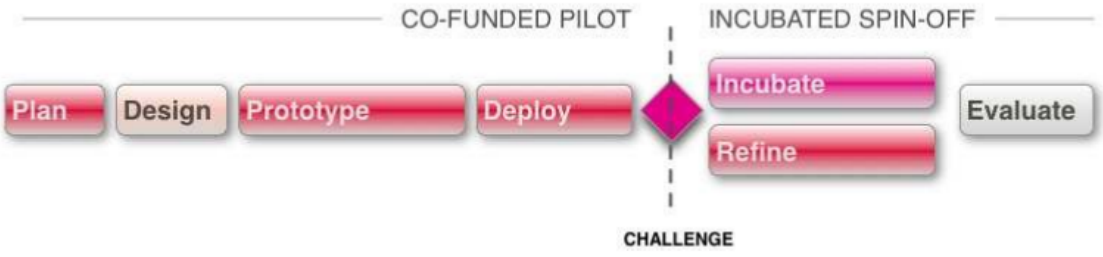
- Semantic index for the entire UK Government Web Archive
- **700M documents**: 42TB, 1.3B files
- 160M unique documents after de-duplication
- Background knowledge (UK Government Ontology): **5B facts**
- Automatic text analysis: **extracted 3B facts** of metadata
- Faceted semantic search in KIM
- 33K hours of cloud processing; up to 500 servers
- www.ontotext.com/case/nationalArchives-skb



EUROPEANA CREATIVE

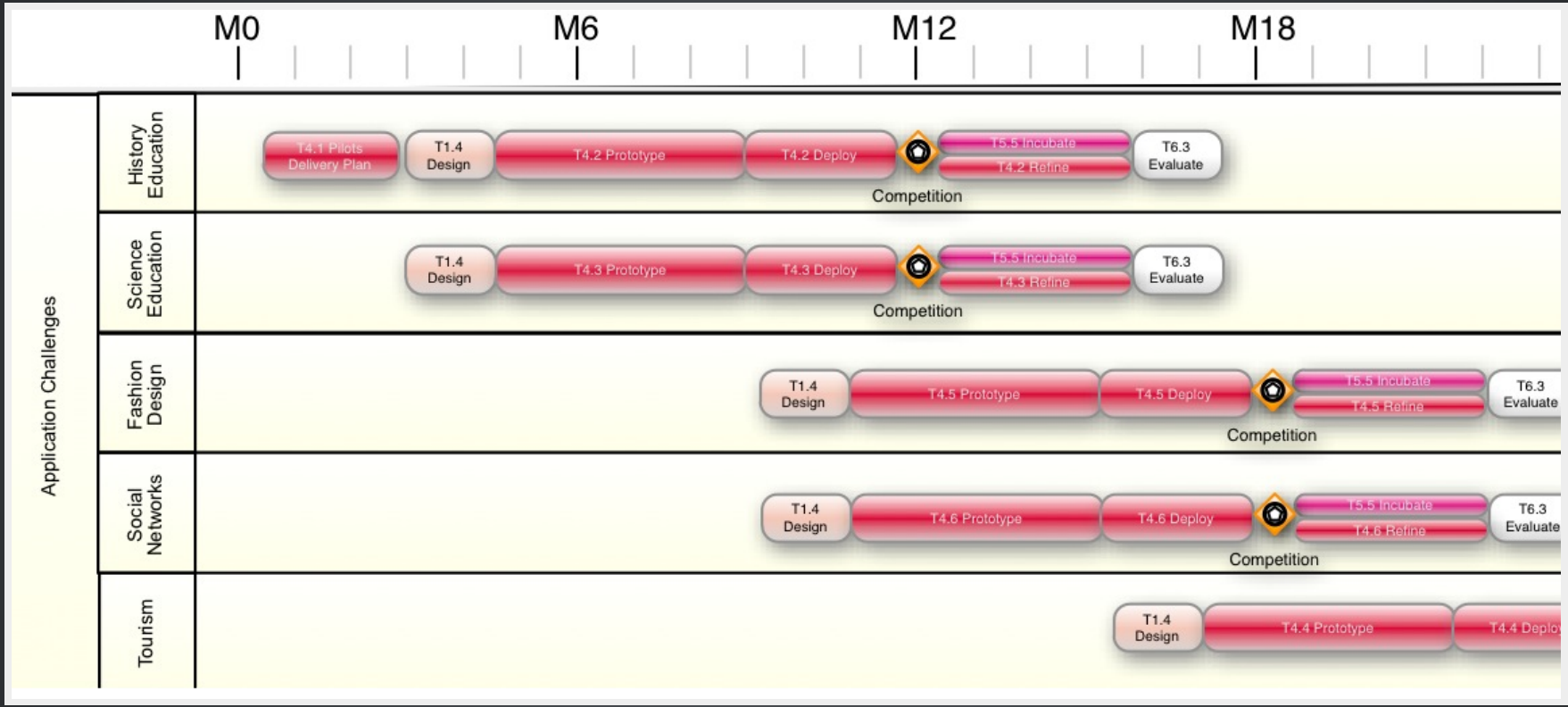
- Enabling Creatives to Work with CH Data
- Pilots by eCreative partners
- Open challenges, growing to incubation support
- Help with collection data, content reuse, Europeana APIs, creative workshop ideas…

Pilot → Challenge → Spin-Off



1. Create pilot project in each thematic area
   - Cooperate on technical, business and content issues
2. Bring each pilot to mature state as rapidly as possible
3. Launch creative industries Challenge in each area
   - Demonstrate APIs & infrastructure along with pilot
4. Incubate most viable project of each Challenge
   - Decrease time-to-market by providing spin-off with business, technical and design coaching & marketing support
5. Facilitate additional projects through outreach campaign

IN 5 PILOT AREAS: TOURISM, SOCIAL NETWORKS, DESIGN, NATURE, HISTORY



ONTOTEXT IN EUROPEANA / EUROPEANA CREATIVE

Ontotext works on fundamental backend technologies important for tech KPIs

Ontotext contributes to 3 KPIs of Europeana in 2014

March 28, 2014

The Europeana Business Plan 2014 was published on 12 March 2014.

Three of the Key Performance Indicators in the important Product Development section (see p.30) are being implemented by Ontotext within the frame of the Europeana Creative project:

- An OAI-PMH service to allow others to use EDM data in bulk
- Semantic repository (triple store), synchronized with the EDM data
- A SPARQL semantic querying endpoint

OAI and SPARQL access will be key new interface services in the upcoming Europeana Labs, complementing the existing Europeana API.

Additional notes about these key services:

- The Open Archives Initiative Protocol for Metadata Harvesting (OAI PMH) is the fundamental protocol for aggregating data from repositories. It is used by Europeana to ingest data from aggregators and providers. However, Europeana itself does not provide an OAI server until now, which is an important omission. The Europeana OAI server will be useful to keep the semantic repository up to date with new Europeana data; and to other large-scale consumers and aggregators
- Ontotext developed and hosts the Europeana semantic repository on OWLIM at http://europeana.ontotext.com. It allows display and downloading of RDF triples and custom views for each Cultural Heritage Object: portal view and graph view. We are currently working on the task of synchronizing through the OAI protocol.
- The repository provides complete SPARQL 1.1 support and a Full Text Search interface. While more Europeana objects are being semantically enriched, this querying interface will be useful to pose graph (relation-oriented) queries that cannot be answered through the existing Europeana API.

An example of the EDM graph view:

EUROPEANA OAI AND SPARQL

Ontotext creates OAI PMH server for Europeana

- So we or others can download objects in bulk

Ontotext hosts the Europeana semantic data (EDM) in OWLIM

- http://europeana.ontotext.com/sparql
  20M objects, obsolete: 1.5 years old
- http://europeana-test.ontotext.com/sparql
  20M objects, incomplete: working with Europeana to update it
- Provides SPARQL querying

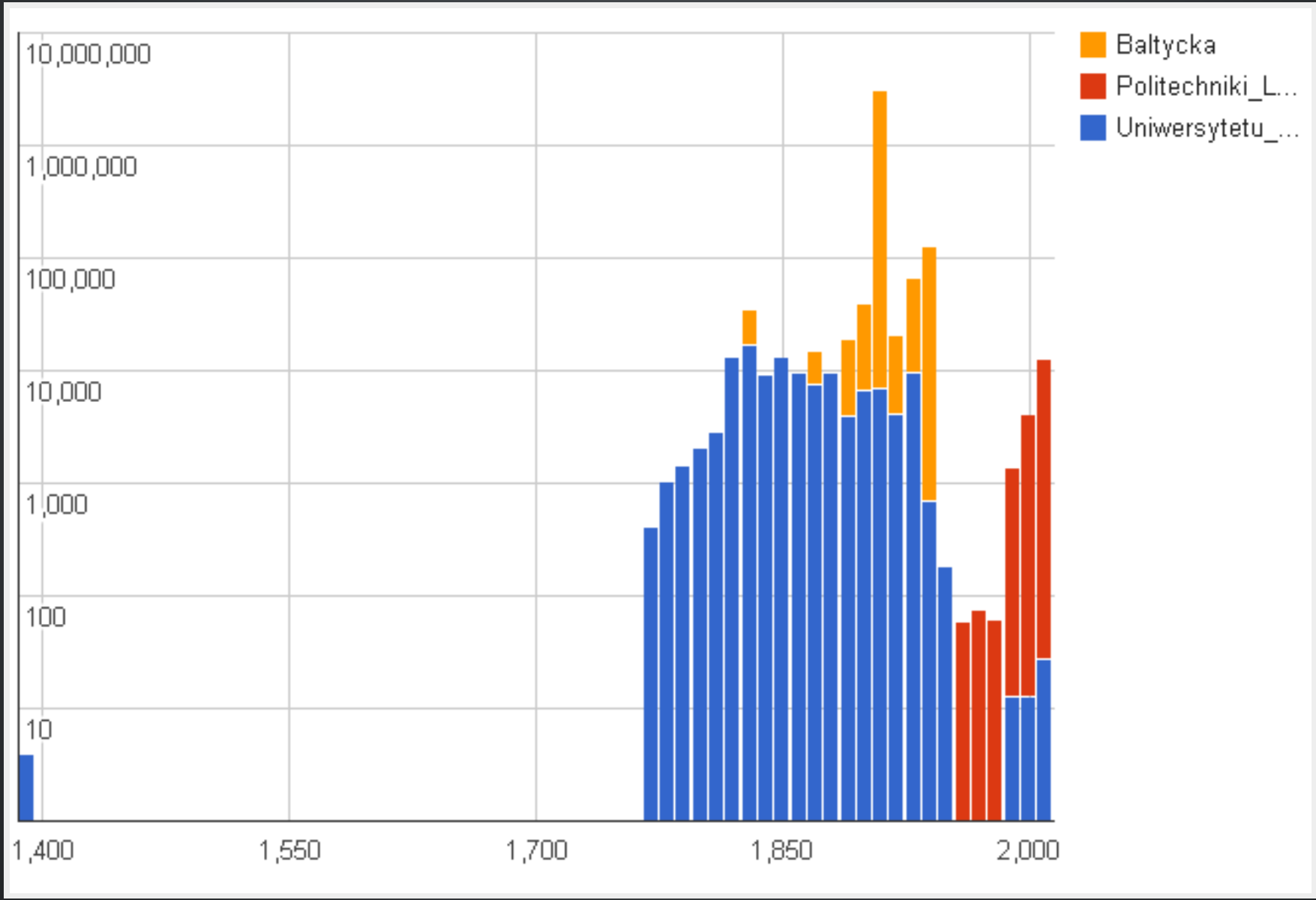
SPARQL 1.1 QUERIES

Eg Polish Periodicals by library and decade
http://europeana-test.ontotext.com/sparql

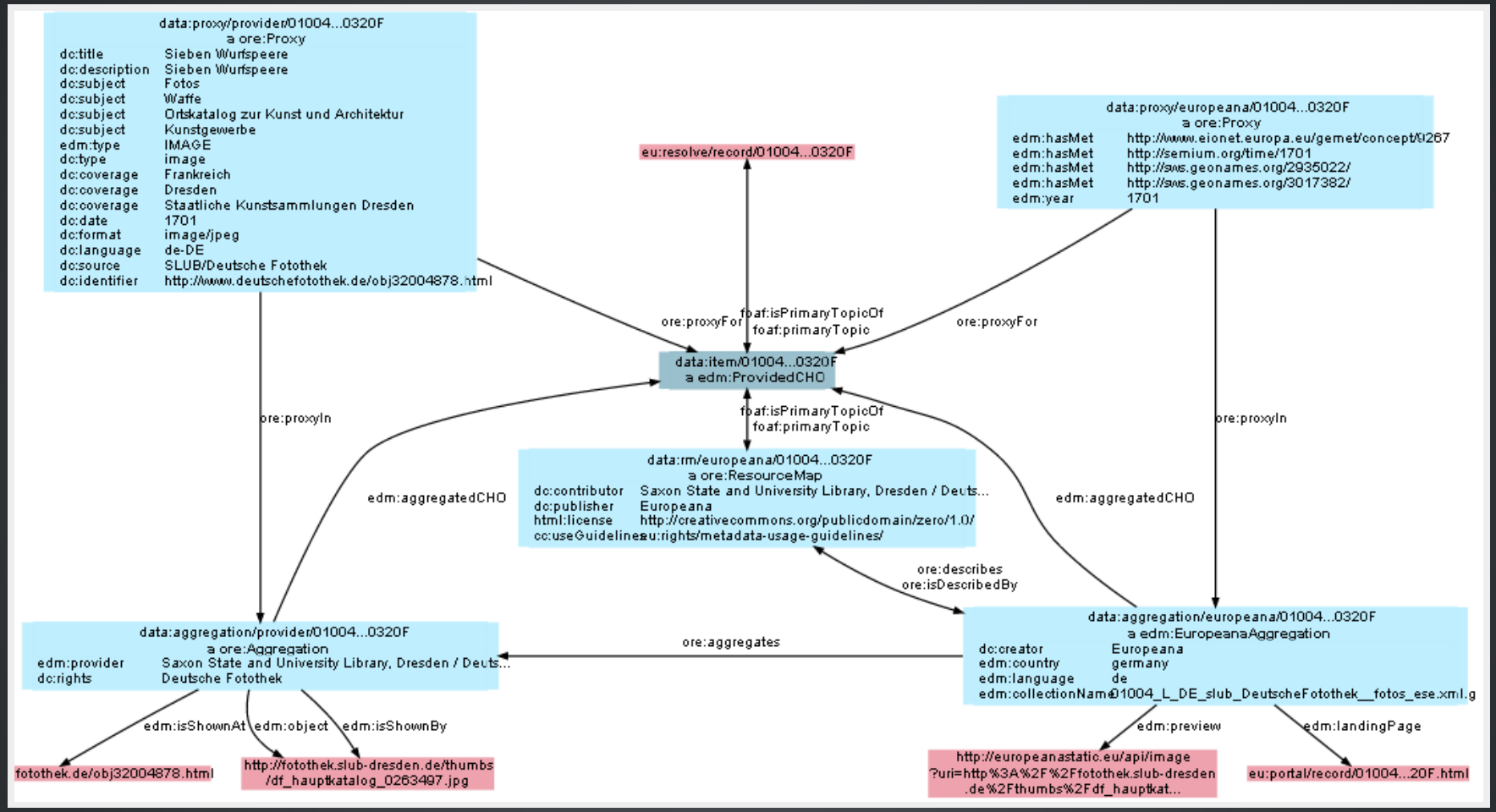
```
select
  ?date
  (sum(?n1) as ?Uniwersytetu_Warszawskiego)
  (sum(?n2) as ?Politechniki_Lubelskiej)
  (sum(?n3) as ?Baltycka)
{
  ?x dc:type 'periodical'@en.
  ?x ore:proxyIn/edm:dataProvider ?dataProvider.
  ?x dc:date ?date2.
  bind (xsd:integer(concat(substr(?date2,1,3),'0')) as ?date)
  bind (if(?dataProvider='e-biblioteka Uniwersytetu Warszawskiego',1,0) as ?n1
)
  bind (if(?dataProvider='Biblioteka Cyfrowa Politechniki Lubelskiej',1,0) as
?n2)
  bind (if(?dataProvider='Bałtycka Biblioteka Cyfrowa',1,0) as ?n3)
} group by ?date order by ?date
```

SPARQL ANALYTICS

Eg Polish Periodicals by library & decade (you can jsfiddle with it)



EDM OBJECT GRAPH



EUROPEANA FOOD AND DRINK

Europeana Food and Drink:

- Explore and celebrate European cultural identity through its culinary and social history
- 29 partners, of which perhaps 20 are content providers

Ontotext works on:

- culinary culture classification scheme
- semantic representation and storage
- semantic text analysis
- semantic application (pilot)

EDAMAM RECIPE/FOOD KNOWLEDGE BASE

Crawled 1.5M recipes, extracted ingredients, matched to SR23 enabling semantic search

Asian Noodle Toss Recipe

"Clara Coulston of Washington Court House, Ohio combines yesterday's chicken with mandarin oranges, crisp veggies and spaghetti for a colorful medley. TIP: For a change of pace Clara used citrus-flavored stir-fry sauce."

8.0 ounces spaghetti dry enriched

1.0 package carrot raw

8.0 ounces sugar granulated

2.0 cups chicken ground raw

1.0 cans tangerine (mandarin oranges) raw

20 min
5 servings
343 kcal/serving

[LOW_FAT, LOW_SODIUM]
[DAIRY_FREE]

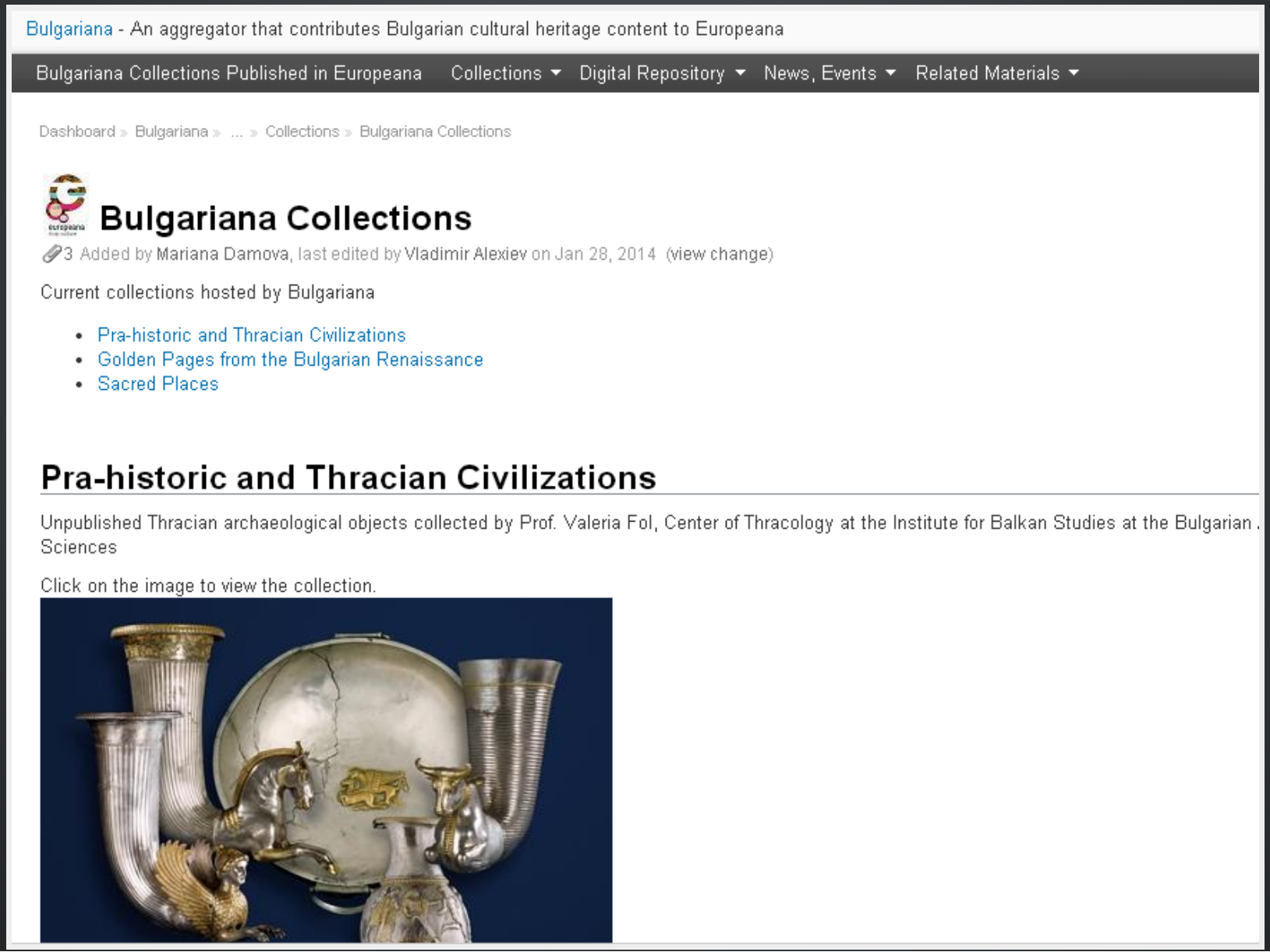
Carb value: 1570 kcal (91%)

allrecipes.com

Protein	29.575 *g*
Total lipid (fat)	3.425 *g*
Energy from fat	30.822 *kcal*
Carbohydrate, by difference	396.106 *g*
Ash	2.019 *g*
Energy	1719.144 *kcal*
Starch	141.637 *g*
Sucrose	227.435 *g*
Glucose (dextrose)	0.272 *g*
Fructose	0.249 *g*
Maltose	4.445 *g*
Water	22.499 *g*
Energy	7191.828 *kJ*
Sugars, total	232.402 *g*

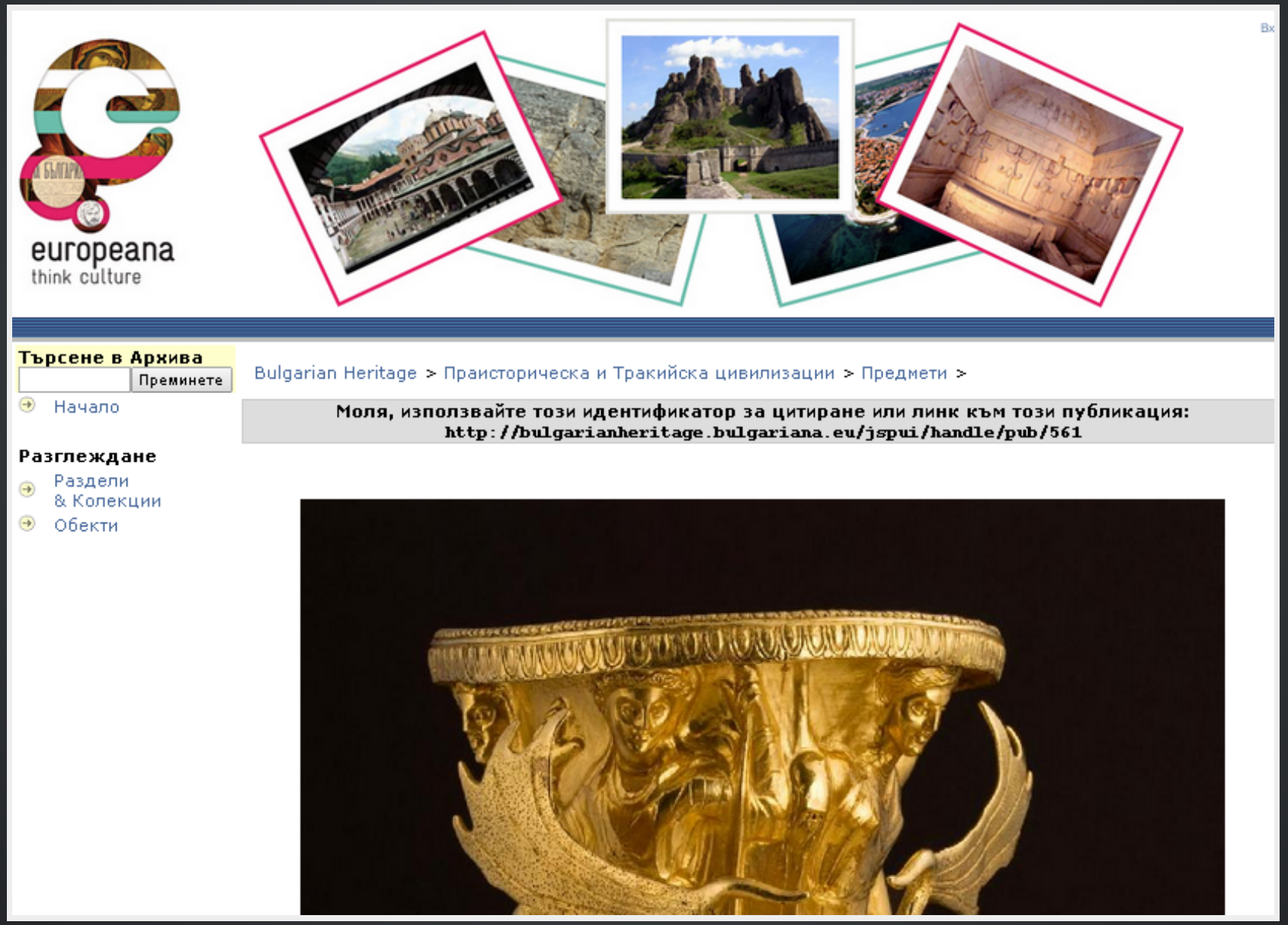
BULGARIANA

A Bulgarian aggregator to Europeana



BULGARIANA COLLECTION: THRACIAN GOLD

World-famous Bulgarian treasures:



RHYTON AT EUROPEANA

Now any European citizen can find it!

europeana
think culture

Search | europeana_collectionName:2021502* @8B>= | Search | Help

Return to search results

CC BY

View item at
Институт за балканистика с Център по тракология

Share

Ритон, страна Б

Description: Троянският принц Парис (Александър), седнал между Атина и Хера, е облечен в т. н. фригийско или „тракийско" облекло – къс, препасан хитон, плътно прилепнали към краката панталони и островърхи обувки. На главата има кожена шапка с дълги, спуснати надолу краища, а в лявата ръка държи жезъл. Вдясно от Парис е седналата на трон Хера, облечена в дълъг, двойно препасан хитон и мантия. Елка Пенкова

Contributor: Институт за балканистика с Център по тракология

Geographic coverage: София ; София ; http://sws.geonames.org/727011/

Date of creation: края на IV - нач. на III в. пр. Хр.

Format: 674,6 г.(тегло) ; 13,5 см.(дължина) ; 8,8 см.(диаметър)

Identifier: 3197, РАМ Пловдив

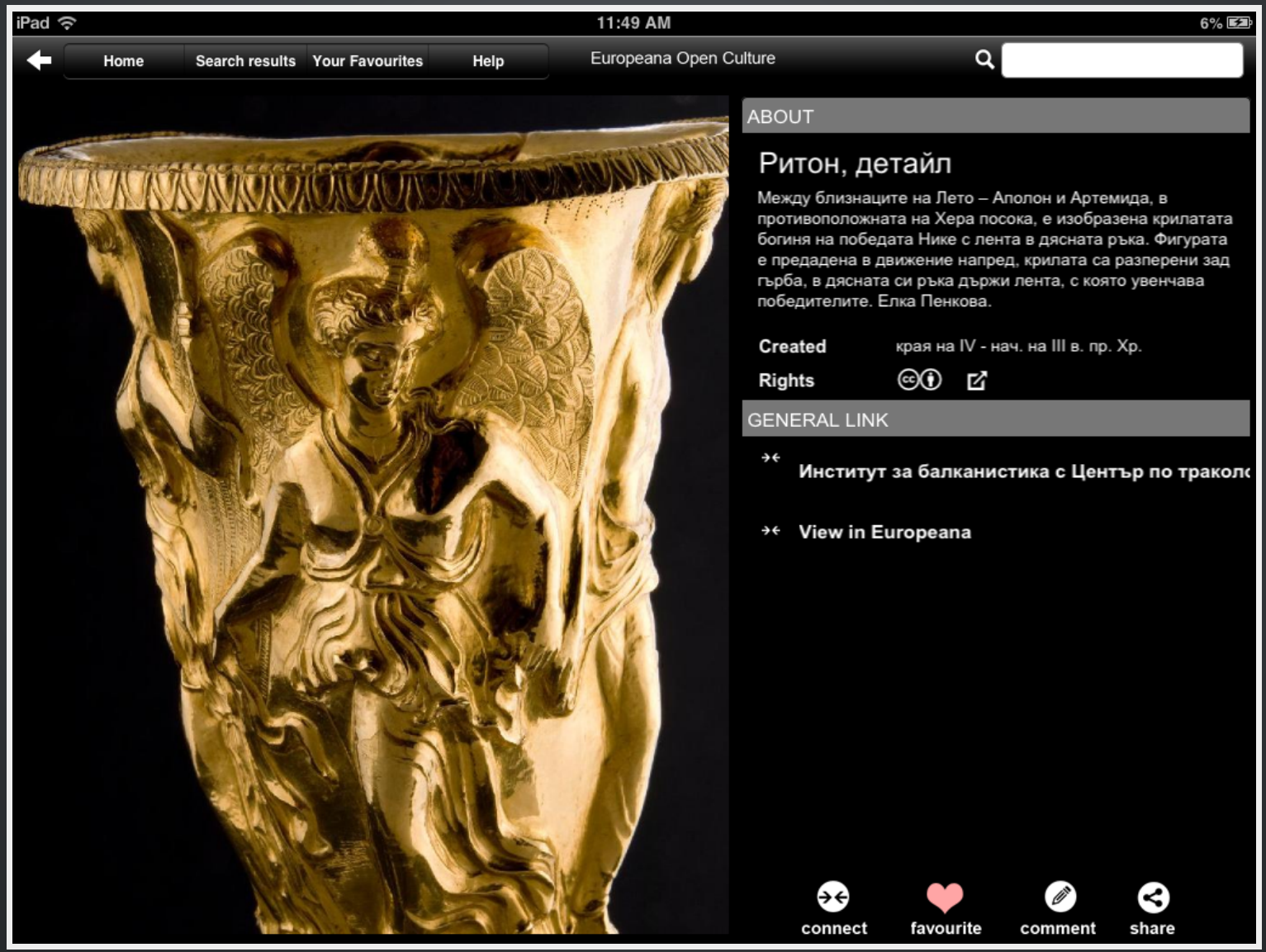
Relation: Венедиков Ив. 1961. Панагюрското златно съкровище, 7 - 9, фиг. 2 - 5. ; Маразов Ив. 1978. Ритоните в Древна Тракия, София, с. 76 - 80. ; Фол Ал. 2002. Тракийският Дионис. Книга трета. Назоваване и вяра. София, 224. ; Стоянов Т. 1999. Съкровището от Борово в археологически и исторически контекст. Първи академични четения в памет на академик Гаврил Кацаров. Seminarium Thracicum 3, с.77-91. ; Китов Г. 2006. Панагюрското съкровище, Варна, 18 - 25.

Rights: Снимка Николай Генов

Provenance: Регионален археологически музей Пловдив

RHYTON AT EUROPEANA OPEN CULTURE

Others make beautiful apps with your data! Bulgariana Collection Featured in Open Culture



ONTOTEXT / GRAPHDB IN CH

Ontotext helped create some of the significant CH LOD datasets, hosted on GraphDB:

- British Museum (CRM): http://collection.britishmuseum.org/sparql
- PSNC Polish Digital Library (CRM/FRBRoo): http://dl.psnc.pl
- Europeana (EDM): http://europeana.ontotext.com
- Getty AAT & TGN (SKOS, SKOS-XL…): http://vocab.getty.edu
- JP LOD.AC: Japanese LOD initiative: http://lod.ac
- FP7 CHARISMA: art database portal: http://archives-charisma-portal.eu/
- FP7 3D COFORM: architectural and archaeological objects
- ConservationSpace: system for conservation specialists

Comparing to:

- FactForge (9 general LOD): http://www.factforge.net
- LinkedLifeData (13 bio LOD): http://linkedlifedata.com

GRAPHDB REPO SIZES

Millions: objects, explicit statements, ex.st per object, total statements; expansion ratio

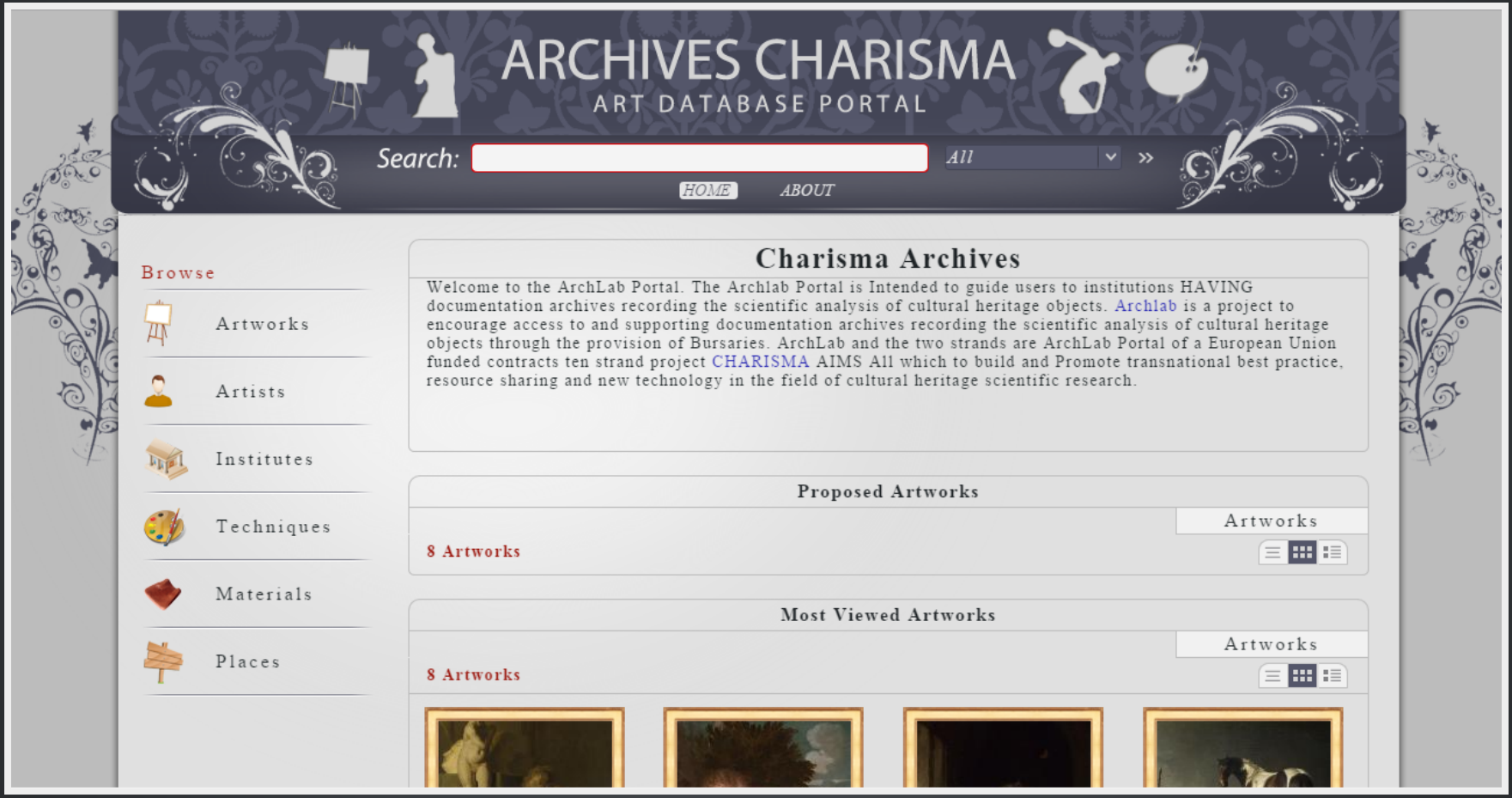
Repo	Ontology	Obj	Ex.st	Ex.st/obj	Tot.st	Exp.	Nodes	Density	Reasonin
BM	CRM	2.0	195	90	916	4.7	54	17.0	rdfs+tran
PSNC	CRM/FRBRoo	3.1	234	75	535	2.3	60	8.9	rdfs-subC
Europeana	EDM	20.3	998	50	3798	3.8	266	14.3	owl-horst
Getty	SKOS etc	1.3	103	79	163	1.6	28	5.8	owl-horst
FF	DC, DBP		1673		3211	1.9	456	7.0	owl-horst
LLD			6706		10192	1.5	1554	6.6	rdfs+tran

References (Partial):

- Large-scale Reasoning with a Complex Cultural Heritage Ontology (CIDOC CRM), CRMEX 2013
- OWLIM Reasoning over FactForge, ORE 2012
- Transforming a Flat Metadata Schema to a Semantic Web Ontology: The Polish Digital Libraries Federation and CIDOC CRM Case Study. Studies in Computational Intelligence 2012

EXAMPLE GRAPHDB USE: CHARISMA PORTAL

http://archives-charisma-portal.eu/



RESEARCHSPACE

- A Virtual Research Environment for art research
- Funded by the Andrew Mellon Foundation
- Executed by the British Museum
- Software developed by Ontotext
- Uses Ontotext's semantic database (GraphDB)

Papers:

- Types and annotations for CIDOC CRM properties, DiPP 2012
- Implementing CIDOC CRM search based on fundamental relations and OWLIM rules, SDA 2012
- Large-scale Reasoning with a Complex Cultural Heritage Ontology (CIDOC CRM), CRMEX 2013
- RDF data and image annotations in ResearchSpace, DH-CASE 2013

RESEARCHSPACE PRESENTATIONS AND VIDEOS

- ResearchSpace website
- News & Files, including presentations & papers
- Videos by Dominic Oldman (British Museum)
- Presentations by Barry Norton (BM, former Ontotext)

For example:

- Oxford Summer School slides, Jul 2014
- GLAMorous LOD and ResearchSpace introduction, Rijksmuseum, May 2014
- GLAMorous LOD, NGA, Washington DC, Apr 2014
- ResearchSpace, CIDOC CRM and Ethical data, UC London
- ResearchSpace CIDOC CRM Search System, Apr 2013
- CIDOC CRM Cultural Semantic Search using Fundamental Relationships, May 2013
- Book of the Dead Project: using CIDOC-CRM, FRBRoo and RDFa
- Querying Cultural Heritage Data

2M BRITISH MUSEUM OBJECTS AS LOD

Eg http://collection.britishmuseum.org/id/object/EOC3130

Hoa Hakananai'a RDF Rank

Source: http://collection.britishmuseum.org/id/object/EOC3130

Subject (89) | Predicate | Object | All

Download in: JSON | RDF | N3/Turtle | N-Triples

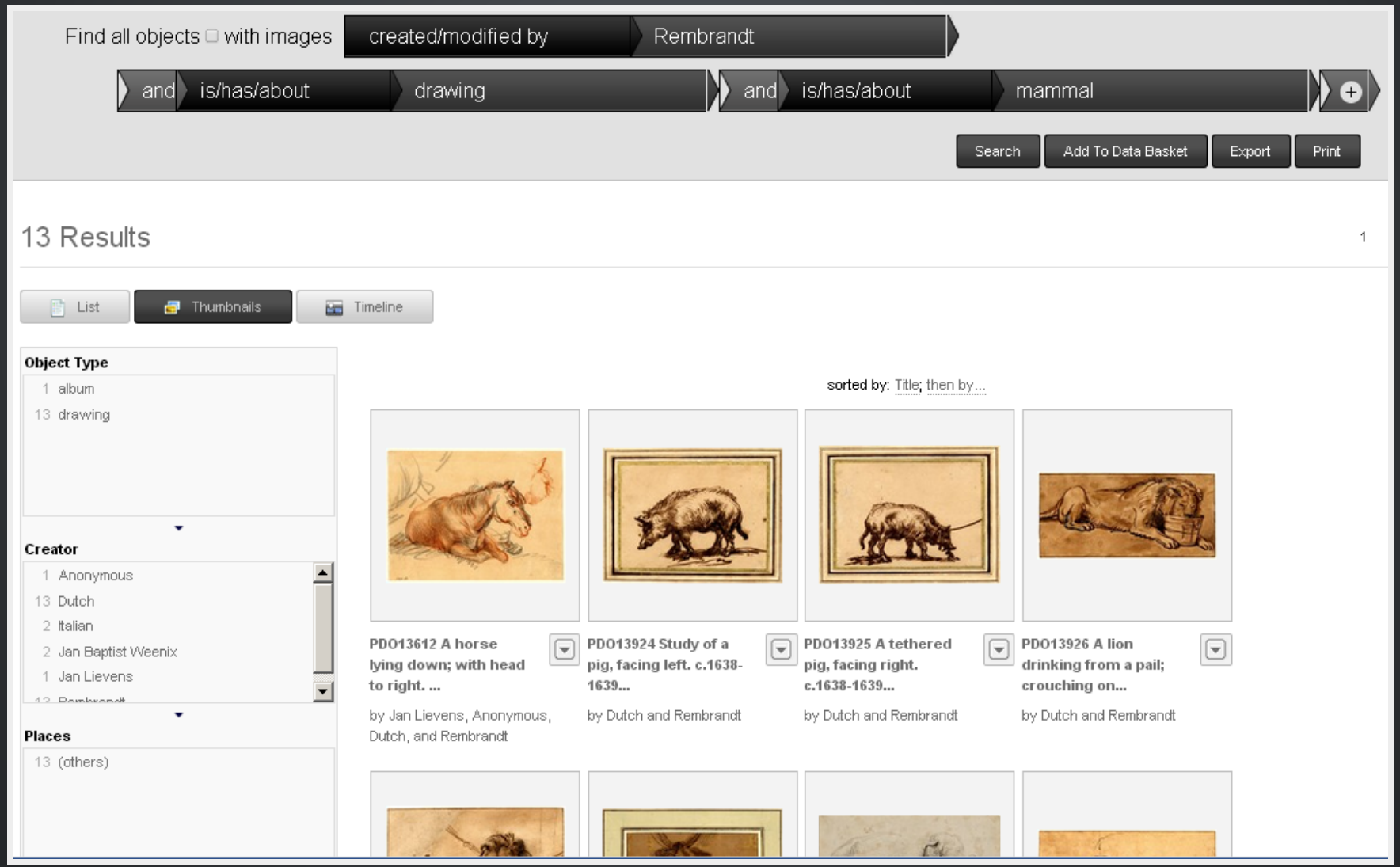
Named Graph: All ▾ Language: English ▾ Inference: Explicit only ▾

Statements in which the resource exists as a subject.

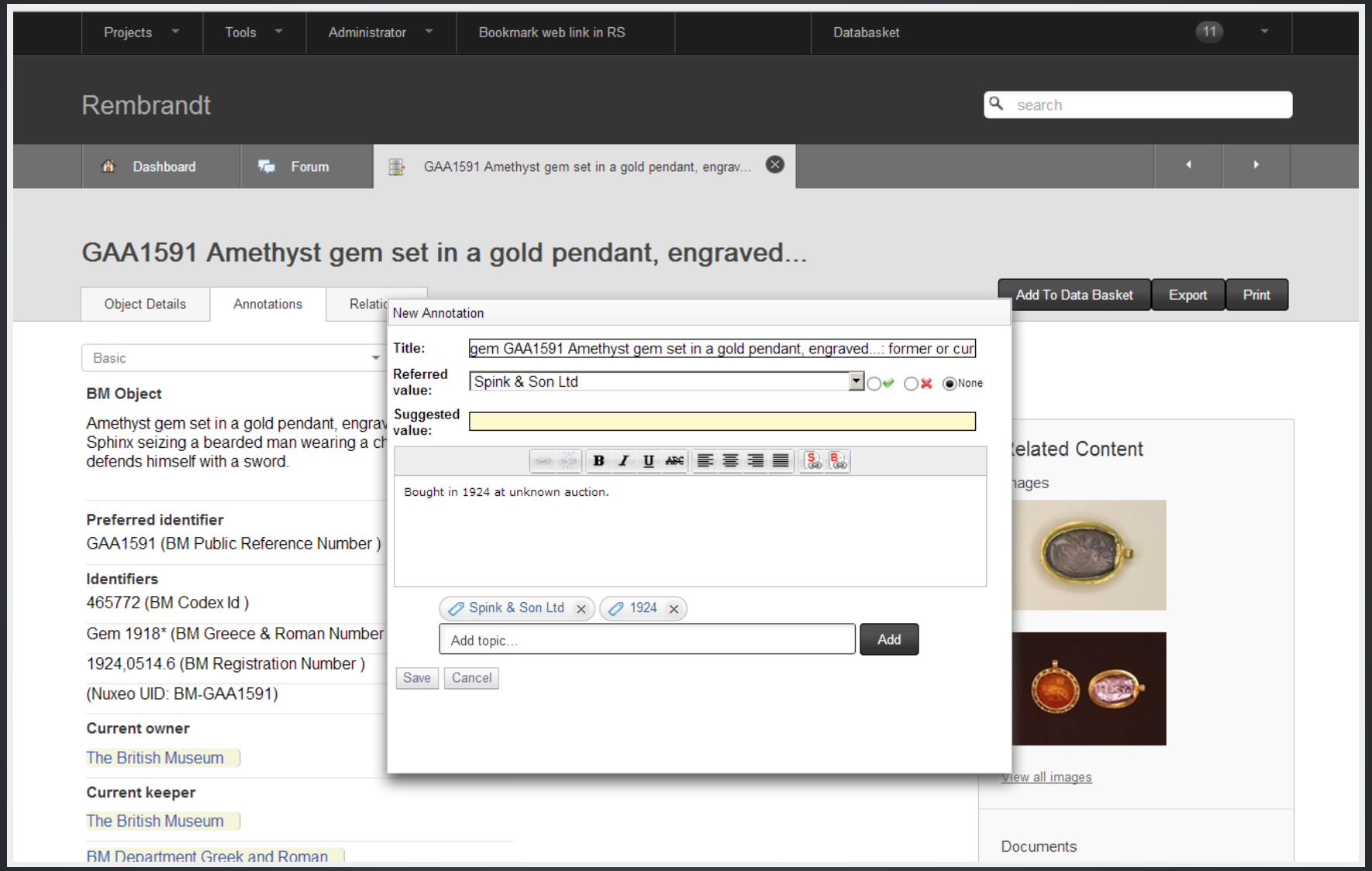
Predicate	Object
rdf:type	ecrm:E22_Man-Made_Object
ecrm:P45_consists_of	thes:x10325, , thes:x10631 thes:x11794
ecrm:P51_has_former_or_curr...	http://collection.britishmuseum.org/id/person-institution/120561, http://collection.britishmuseum.org/id/person-institution/49731
rdfs:label	Hoa Hakananai'a, Moai

RESEARCHSPACE SEMANTIC SEARCH

Also works across collections, eg BM and Yale Center for British Art

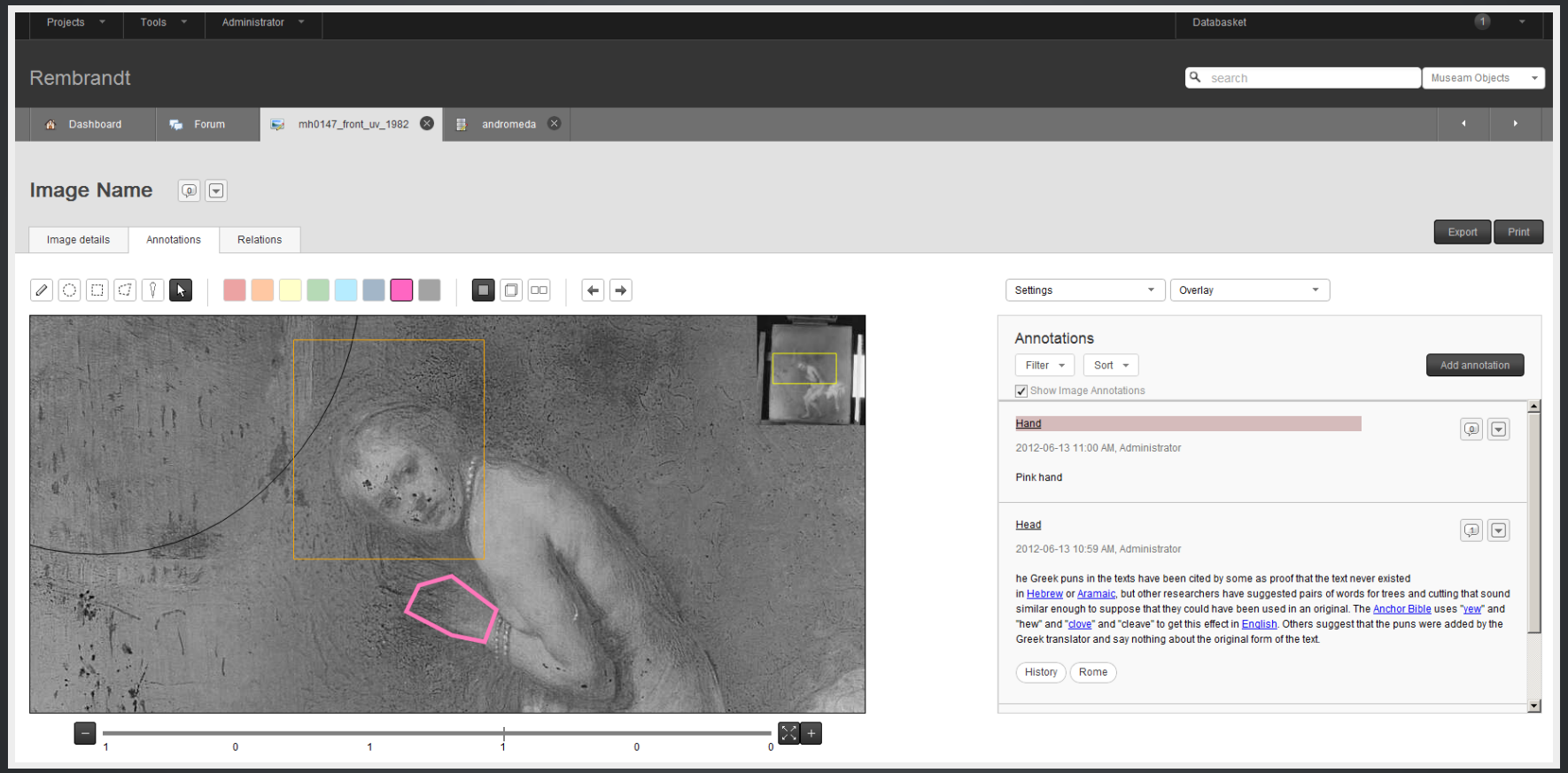


RESEARCHSPACE: SEMANTIC DATA ANNOTATION



RESEARCHSPACE: SEMANTIC IMAGE ANNOTATION

Allows arbitrary shapes using SvgEdit, supports deep zoom, relates to semantic facts, or free discussion



CRM SEARCH (FUNDAMENTAL RELATIONS)

CRM data comprises complex graphs of nodes and properties.

- How can a user search through such complex graphs?
- The number of possible combinations is staggering

FC/FR Approach:

- New Framework for Querying Semantic Networks (FORTH TR419, 2011)
- Fundamental Categories and Relationships for intuitive querying CIDOC-CRM based repositories (FORTH TR-429, Apr 2012, 153 pages)
- "Compresses" the semantic network by mapping networks of CRM properties to single FRs
- FRs serve as a "search index" over the CRM semantic web
- Allow the user to use a simpler query vocabulary

CRM FUNDAMENTAL RELATIONS MATRIX

- 114 FRs over all combinations of FCs; 18 "specialization FRs"

Domain (select)	Range(query parameter)				
	Thing	Actor	Place	Event	Time
Thing	2.is part of 3.is similar or the same with 4. has met 5. from 6. is origin of 8. refers to 9.is referred by	4.has met 5.from 8.refers to 9.is referred by	4.from 8.refers to 9.is referred to at	4.from 8.refers to	4.from
Actor	4.has met 6.is creator or provider of 8. refers to 9.is referred by	2.is member of 4. has met 5.has parent or founder 6.is parent or founder of 8.refers to 9.is referred by	4.has met 5.from 8.refers to 9.is referred to at	4.has met 8.refers to 6.has met	8.refers to 6.has met 4.from
Place	5.is origin of 8.refers to or is about 9.isreferred by	5.is origin of 8.refers to or is about 9.is referred by	2.is part of 5.is origin of	9.is referred by 5.is origin of	7.at
Event	5.is origin of 9.is referred by 8.refers to or is about	4.from 9.is referred by 8.refers to or is about 6.has met	8.refers to or is about 7.at	8.refers to or is about 2.is part of	8.refers to or is about 7.at
Time	5.is origin of	5.is origin of	5.is origin of	5.is origin of	2.is part of

EXAMPLE: THING FROM PLACE

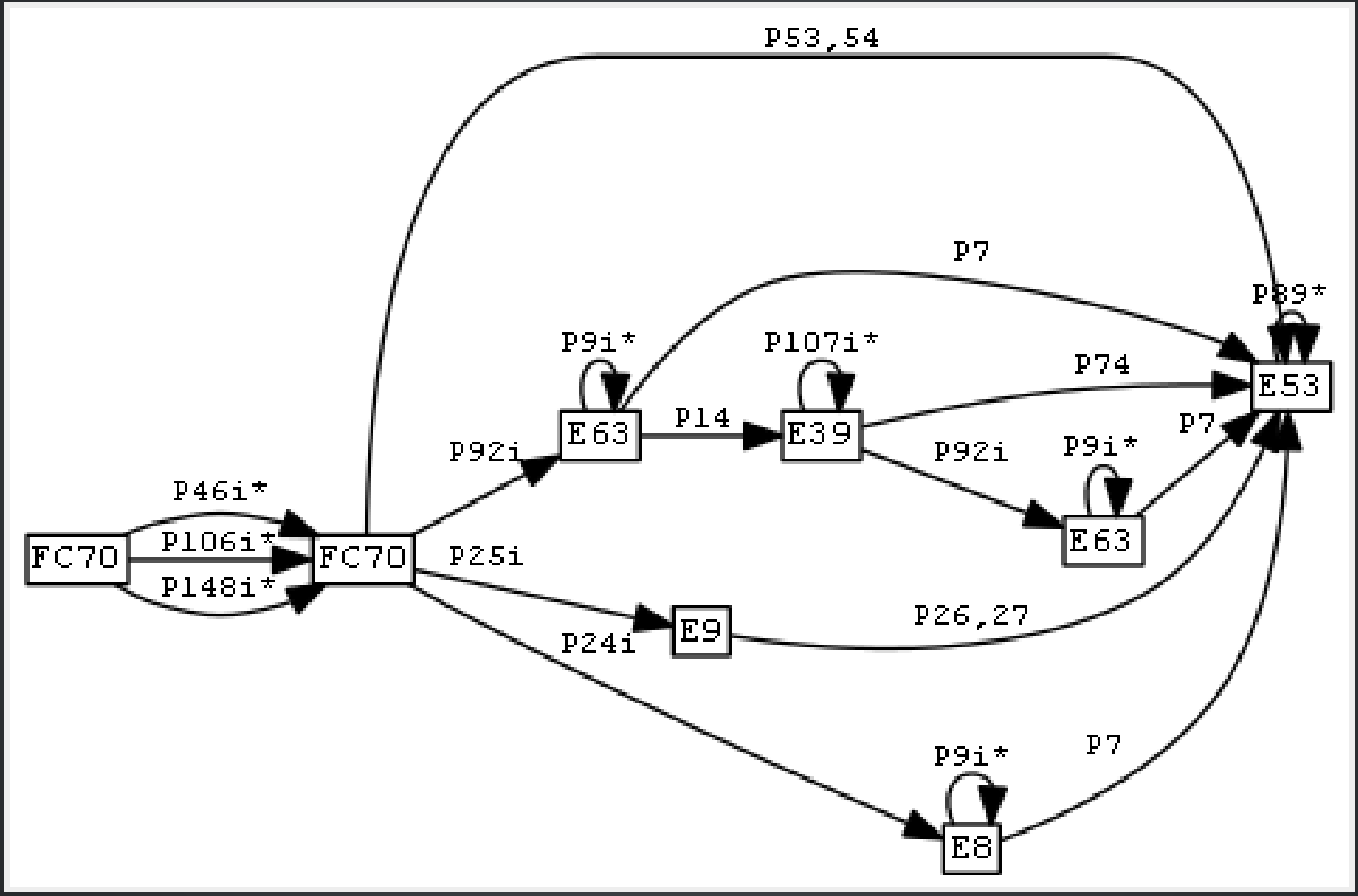
How a Thing's **origin** can be related to Place (* = recursion)

- Thing (part of another)* considered to be "from" Place if:
- is formerly or currently located at Place (falling in another)*
- or was brought into existence (produced/created) by an Event (part of another)*
  - that happened at Place (falling in another)*
  - or was carried out by an Actor (who is member of a Group)*
    - who formerly or currently has residence at Place (falling in another)*
    - or was brought into existence (born/formed) by an Event (part of another)* that happened at Place (falling in another)*
- or was Moved to/from a Place (falling in another)*
- or changed ownership through an Acquisition (part of another)*
  - that happened at Place (falling in another)*

THING FROM PLACE: DEFINITION (CRM CLASSES & PROPERTIES)

```
FC70_Thing --(P46i_forms_part_of* | P106i_forms_part_of* | P148i_is_component_
of*)-> FC70_Thing:
  {FC70_Thing --(P53_has_former_or_current_location | P54_has_current_permanen
t_location)-> E53_Place:
    {E53_Place --P89_falls_within*-> E53_Place}
  OR FC70_Thing --P92i_was_brought_into_existence_by-> E63_Beginning_of_Existe
nce:
    {E63_Beginning_of_Existence --P9i_forms_part_of*-> E5_Event:
      {E5_Event --P7_took_place_at-> E53_Place:
        {E53_Place --P89_falls_within*-> E53_Place}
      OR E7_Activity --P14_carried_out_by-> E39_Actor:
        {E39_Actor --P107i_is_current_or_former_member_of* -> E39_Actor:
          {E39_Actor --P74_has_current_or_former_residence  -> E53_Place:
            {E53_Place --P89_falls_within*-> E53_Place}
          OR E39_Actor --P92i_was_brought_into_existence_by-> E63_Beginning_of
_Existence:
            {E63_Beginning_of_Existence --P9i_forms_part_of*-> E5_Event:
              {E5_Event --P7_took_place_at-> E53_Place:
                {E53_Place --P89_falls_within* -> E53_Place}}}}}}}
  OR E19_Physical_Thing  --P25i_moved_by-> E9_Move:
    {E9_Move --(P26_moved_to | P27_moved_from)-> E53_Place:
      {E53_Place  --P89_falls_within*-> E53_Place}}
  OR E19_Physical_Object --P24i_changed_ownership_through-> E8_Acquisition:
    {E8_Acquisition --P9i_forms_part_of*-> E5_Event:
      {E5_Event --P7_took_place_at-> E53_Place:
        {E53_Place --P89_falls_within*-> E53_Place}}}}
```

THING FROM PLACE: GRAPHICAL REPRESENTATION

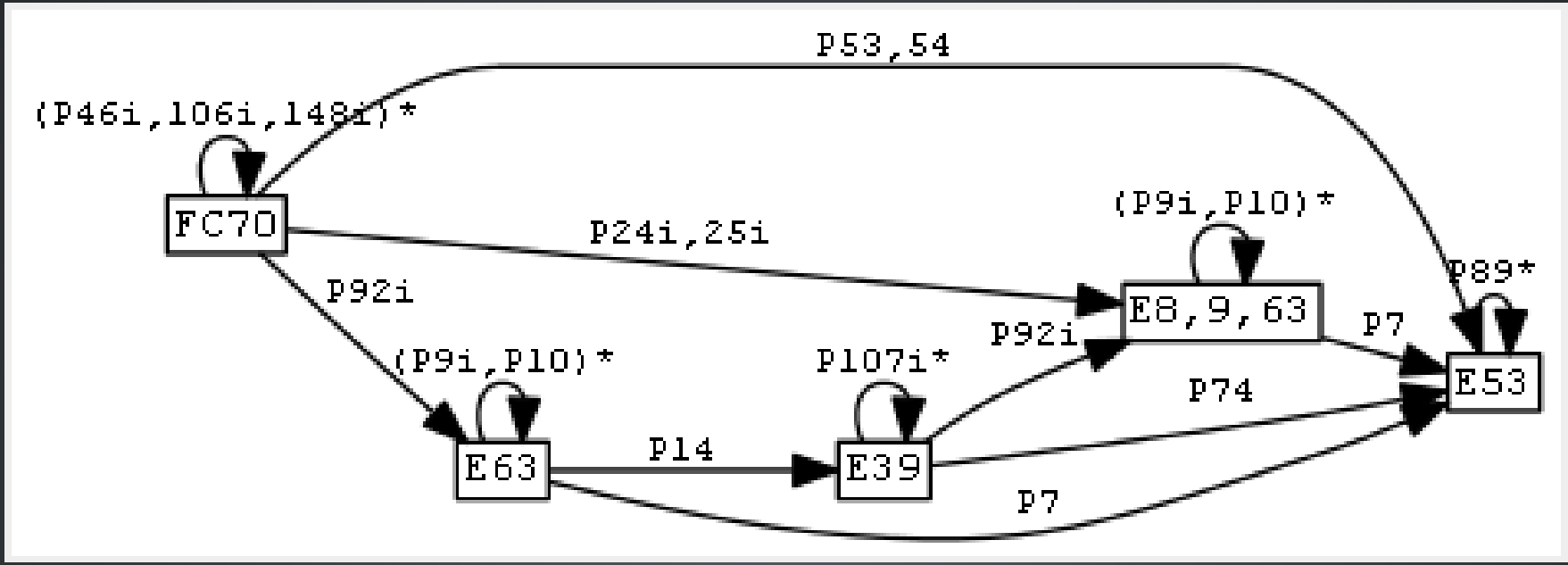


THING FROM PLACE: SPARQL QUERY

```
select ?t ?p2 {
?t a FC70_Thing. ?t (P46i_forms_part_of* | P106i_forms_part_of* | P148i_is_com
ponent_of*) ?t1.
  {?t1 (P53_has_former_or_current_location | P54_has_current_permanent_locatio
n) ?p1}
  UNION
  {?t1 P92i_was_brought_into_existence_by ?e1. ?e1 P9i_forms_part_of* ?e2.
      {?e2 P7_took_place_at ?p1}
      UNION
      {?e2 P14_carried_out_by ?a1.
        ?a1 P107i_is_current_or_former_member_of* ?a2.
          {?a2 P74_has_current_or_former_residence ?p1}
          UNION
          {?a2 P92i_was_brought_into_existence_by ?e3. ?e3 P9i_forms_part_of*
?e4.
           ?e4 P7_took_place_at ?p1}}}
  UNION
  {?t2 P25i_moved_by ?e5. ?e5 (P26_moved_to | P27_moved_from) ?p1}
  UNION
  {?t2 P24i_changed_ownership_through ?e6.
    ?e6 P9i_forms_part_of ?e7. ?e7 P7_took_place_at ?p1}.
?p1 P89_falls_within* ?p2}
```

- Very complex and expensive, especially when you need to combine with other FRs into composite queries
- Tried in 3D COFORM, just doesn't work

THING FROM PLACE: CORRECTED/RATIONALIZED DEFINITION



THING FROM PLACE: DECOMPOSING INTO SUB-FRS

- "Sub-FRs" are auxiliary relations used to build up the final FR
- The numbering comes from CRM property and entity names
- Prefixes: FR: final result, FRT: transitive, FRX: non-transitive, FC70=Thing or E: from/to that class

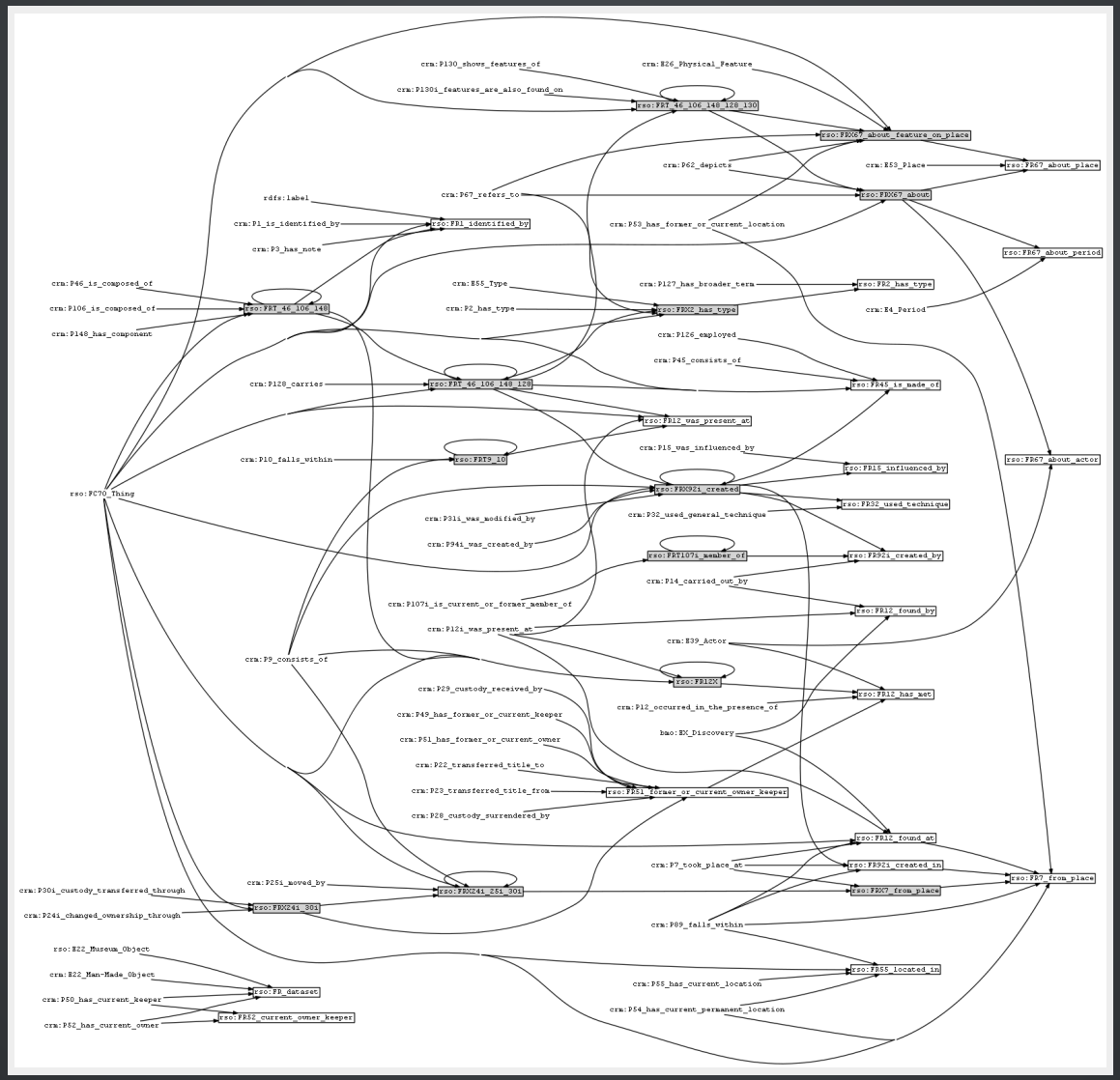
```
        # self-loops and simple disjunctions
FRT_46i_106i_148i := (P46i|P106i|P148i)+
FRT_9i_10 := (P9|P10)+
FRT_107i := P107i+
FRT_89 := P89+
FRX_53_54 := (P53|P54)
FRX_24i_25i := (P24i|P25i)
         # growing fragments
FRX_92i := P92i | P92i/FRT_9i_10
FRX_92i_14 := FRX_92i/P14 | FRX_92i/P14/FRT_107i
FRX_FC70_E8_9_63 := FRX_92i_14/P92i | FRX_24i_25i
FRX_FC70_E8_9_63_P7 := FRX_FC70_E8_9_63/P7 | FRX_FC70_E8_9_63/FRT_9i_10/P7
FRX7 := FRX_53_54 | FRX_FC70_E8_9_63_P7 | FRX_92i_14/P74 | FRX_92i/P7
FRX7_P89 := FRX7 | FRX7/FRT_89
FR7 := FRX7_P89 | FRT_46i_106i_148i/FRX7_P89
```

FR IMPLEMENTATION AS OWLIM RULES

- OWL2 doesn't have conjunctive properties
- So we implemented with OWLIM rules, using the parallel/sequential decompositions above
- Details: FR Implementation
- Implemented 19 FRs of Thing (see FR Names):
  - refers to or is about Place; from Place; is/was located in Place
  - has met Actor; by Actor
  - refers to or is about Event; has met Event
  - is made of Material; is/has Type; used technique; identified by Identifier
- Use 44 CRM properties. Took 86 rules, 10 axioms, 26 sub-FRs (gray on next slide)
- Refactoring idea: http://vladimiralexiev.github.io/pres/extending-owl2/index.html

FR DEPENDENCY DIAGRAM

Used to check no disconnected props, no misspelling in rules



GETTY VOCABULARIES LOD

Well-known and important cultural heritage thesauri:

- Art and Architecture Thesaurus (AAT)
- Thesaurus of Geographic Names (TGN)
- Unified List of Artist Names (ULAN)
- Cultural Object Names Authority (CONA)

Ontotext helps Getty publish them as LOD:
http://vocab.getty.edu

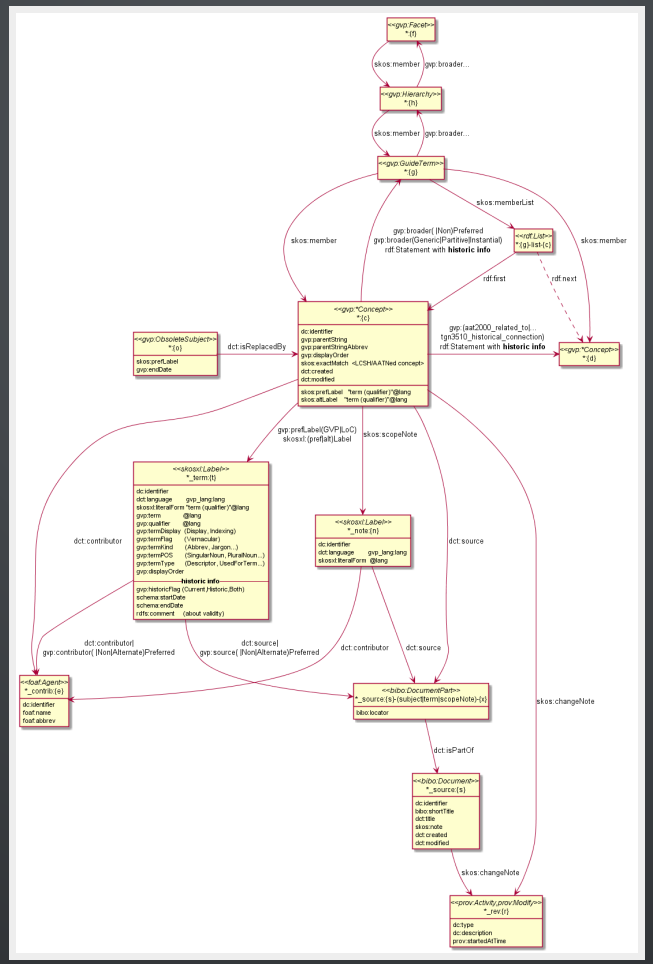
- AAT published Feb 2014, already sees numerous use cases
- TGN published Aug 2014
- Continuing with ULAN, CONA; AATA (bibliography), Getty Museum data
- Special session at CIDOC Congress (Dresden, Sep 2014)

GETTY EXTERNAL ONTOLOGIES

- SKOS, SKOSXL, ISO 25964 for representing thesaurus info;
- DC, DCT for common properties;
- BIBO, FOAF for sources and contributors;
- WGS, Schema for geographic information;
- PROV for revision history;
- RDF, RDFS, OWL, XSD for system properties;
- R2RML for implementing the conversion.

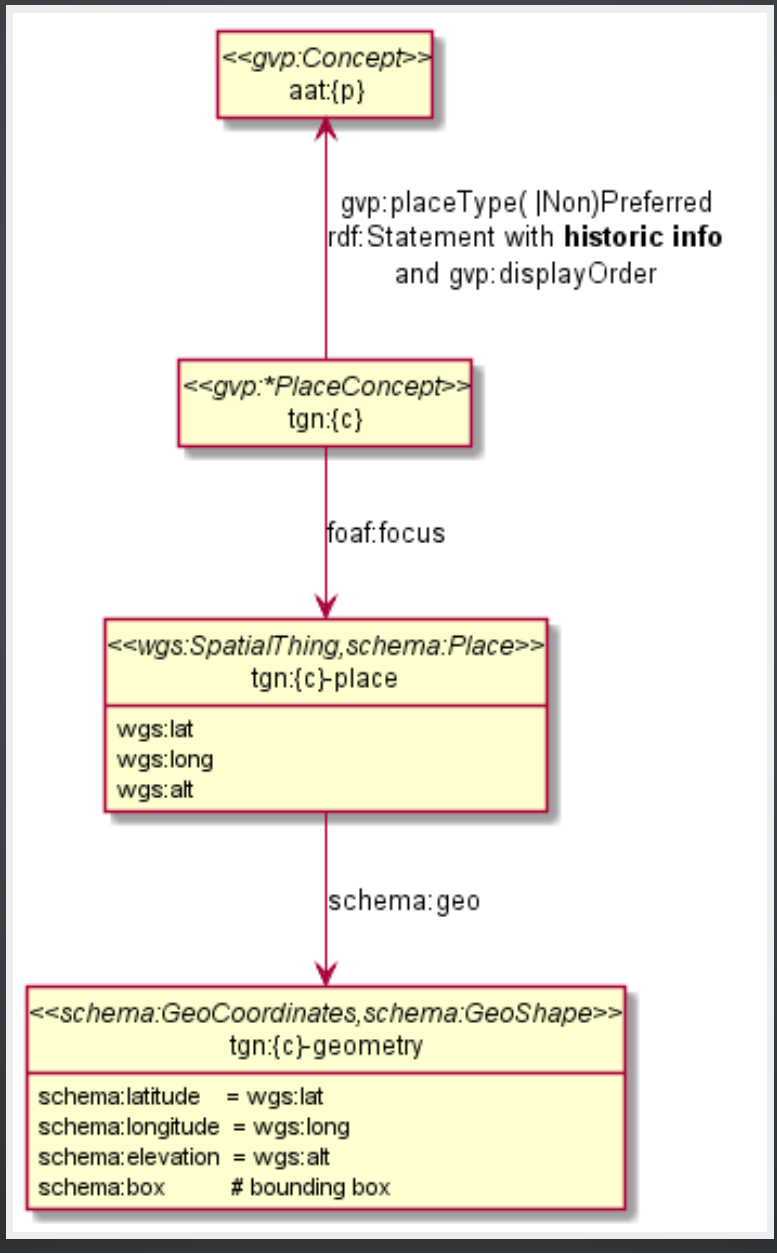
GETTY SEMANTIC REPRESENTATION

- Covers subjects (concepts, guide terms…), hierarchical rels, associative rels, historic info, labels, sources, contributors, revision history, languages …
- Doc (100 pages!): below is Semantic Overview



TGN SEMANTIC REPRESENTATION

Duality Concept-Place (ala VIAF, UK BL, FR BnF, SE KB..)



GVP ONTOLOGY

Custom ontology: http://vocab.getty.edu/ontology

Table of contents

Ontologies (1):

Getty Vocabulary Program ontology

Classes (10):

AdminPlaceConcept , Concept , Facet , GuideTerm , Hierarchy , ObsoleteSubject , PhysAdminPlaceConcept , PhysPlaceConcept , Scope Note , Subject

Properties (177):

aat2000_related_to , aat2001_formerly_referred_to , aat2100_distinguished_from , aat2110_meaning-usage_overlaps_with , aat2203_associated_with , aat2205_causes-is_required , aat2206_caused_by-requires , aat2208_locus-setting_for , aat2209_used-located_in , aat2211_produce , aat2212_produced_by , aat2215_required_for , aat2216_require , aat2218_used-function_as , aat2219_have_form , aat2222_act_upon , aat2223_are_acted_upon , aat2281_have_counterpart , aat2285_practiced-studied_by , aat2286_practice-study , aat2291_locus-setting_for , aat2292_work-live_in , aat2294_locus-setting_for , aat2295_located_in , aat2311_performed_by , aat2312_perform , aat2315_used_by , aat2316_use , aat2318_involved_in , aat2319_involves , aat2321_used_by , aat2322_use , aat2325_created_by , aat2326_create , aat2328_involved_with , aat2329_involves , aat2332_affiliated_with , aat2333_have_affiliates , aat2335_associated_with , aat2336_has_associates , aat2397_focus_of , aat2398_focuses_on , aat2408_locus-setting_for , aat2409_takes_place_in , aat2411_involved_in , aat2412_involves , aat2415_required_for , aat2416_requires , aat2418_uses , aat2419_used_for , aat2421_locus-setting_for , aat2422_takes_place_in , aat2424_produced_by , aat2425_produces , aat2427_produced_by , aat2428_produces , aat2431_required_for , aat2432_requires , aat2434_contextualized_in , aat2435_context_for , aat2501_made_of-require , aat2502_material_for , aat2504_used-located_in , aat2505_locus-setting_for , aat2507_produce-process , aat2508_produced-processed_by , aat2551_reflect-produced_by , aat2552_reflected_in-produces , aat2554_reflects , aat2555_reflected_in , aat2557_reflects , aat2558_reflected_in , aat2562_locus-setting_for , aat2563_located_in , aat2601_possessed_by-existing_in , aat2602_possess-exist_in , aat2604_possessed_by-existing_in , aat2605_possess-exist_in , aat2607_caused_by-requires , aat2608_causes-required_for , aat2612_possessed_by-existing_in , aat2613_possess-exist_in , aat2801_conjuncted_with , aat2802_exemplified_by , aat2803_example_of , aat2805_contextualized_in , aat2806_provide_context , aat2807_derived_from_common_source , aat2809_coexisted_with , aat2811_preceded , aat2812_followed , aat2814_constituent_of , aat2815_composed_of , aat2817_derived-made_from , aat2818_source_for , aat2821_based_on , aat2822_basis_of , aat2824_has_parallels_with , aat2826_used_with , aat2828_use-require , aat2829_used-required_for , aat2831_associated_with , aat2833_ancestor_of , aat2834_decendant_of , aat2836_derived-made_from , aat2837_source_for , aat2841_derived-made_from , aat2842_source_for , aat2845_used-located_in , aat2846_locus-setting_for , aat2848_involved_in , aat2849_involves , aat2852_involved_with , aat2853_involve , aat2875_characteristic_of , aat2876_characterized_by , aat2878_preceded-source_for , aat2879_followed-developed_from , aat2881_reflected_in , aat2882_reflect , aat2884_involved_with , aat2885_involves , aat2891_exemplified_by , aat2892_example_of , aat2894_exemplified_by , aat2895_example_of , aat2900_miscellaneous_relationship , broader , broaderExtended , broaderGeneric , broaderGenericExtended , broaderInstantial , broaderInstantialExtended , broaderNonPreferred , broaderPartitive , broaderPartitiveExtended , broaderPreferred , broaderPreferredExtended , contributorAlternatePreferred , contributorNonPreferred , contributorPreferred , displayOrder , historicFlag , narrower , narrowerExtended , parentString , parentStringAbbrev , placeTypeNonPreferred , placeTypePreferred , prefLabelGVP , prefLabelLoC , qualifier , sourceAlternatePreferred , sourceNonPreferred , sourcePreferred , term , termDisplay , termFlag , termKind , termPOS , termType , tgn3000_related_to , tgn3001_distinguished_from , tgn3005_possibly_identified_as , tgn3101_adjacent_to , tgn3102_coextensive_with , tgn3110_meaning-usage_overlaps_with , tgn3201_capital_of , tgn3202_capital_is , tgn3301_ally_of , tgn3317_member_of , tgn3318_member_is , tgn3401_moved_from , tgn3402_moved_to , tgn3411_successor_of , tgn3412_predecessor_of , tgn3510_historical_connection

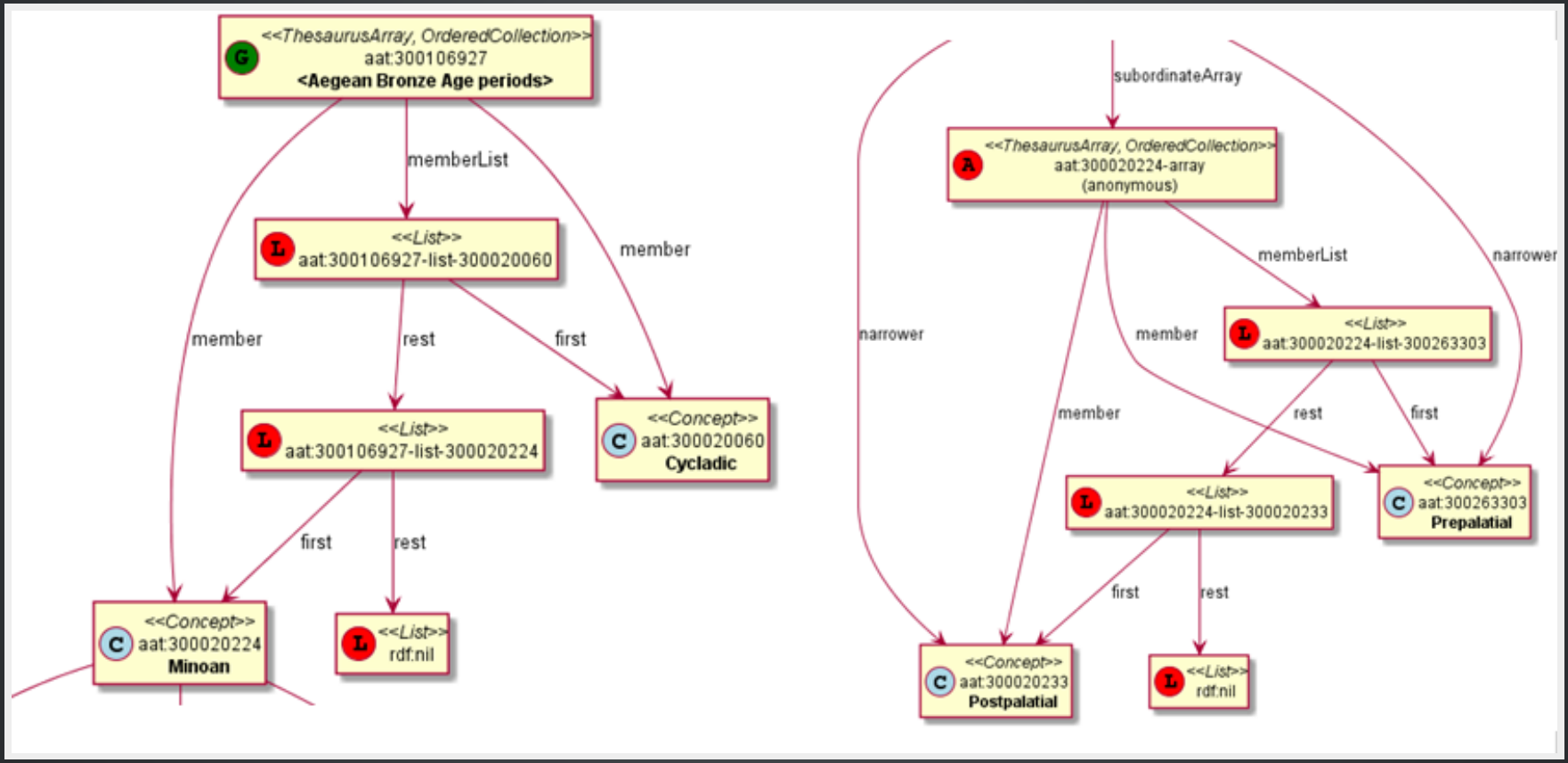
Individuals (44):

Abbreviation , Adjectival , Alternate Descriptor , Art and Architecture Thesaurus , Both , Both Singular and Plural , Chemical Name , Common term , Current , Descriptor , FIPS Code , Full term , Getty Research Institute , Historic , Historic Flag concept scheme , ISO alpha-2 code , ISO alpha-3 code , ISO numeric-2 code , ISO numeric-3 code , Jargon or Slang , Loan Term , Neologism , Noun , Official Name , Ontotext Corp , Part of Speech concept scheme , Past Participle , Plural Noun , Provisional Name , Pseudonym , Scientific or Technical term , Singular Noun , Site Name , Term Display concept scheme , Term Flag concept scheme , Term Kind concept scheme , Term Type concept scheme , Thesaurus of Geographic Names , US Postal Service Code , Use for Display , Use in Indexes/lists , Used for Term , Verbal Noun/Gerund , Vernacular

Go to top | Ontologies | Classes | Properties | Individuals | Expand/Collapse all | Help | Back to form

USE OF ISO 25946 IN GETTY LOD

Latest standard on thesauri: ISO 25946. Use Thesaurus Array for ordered children

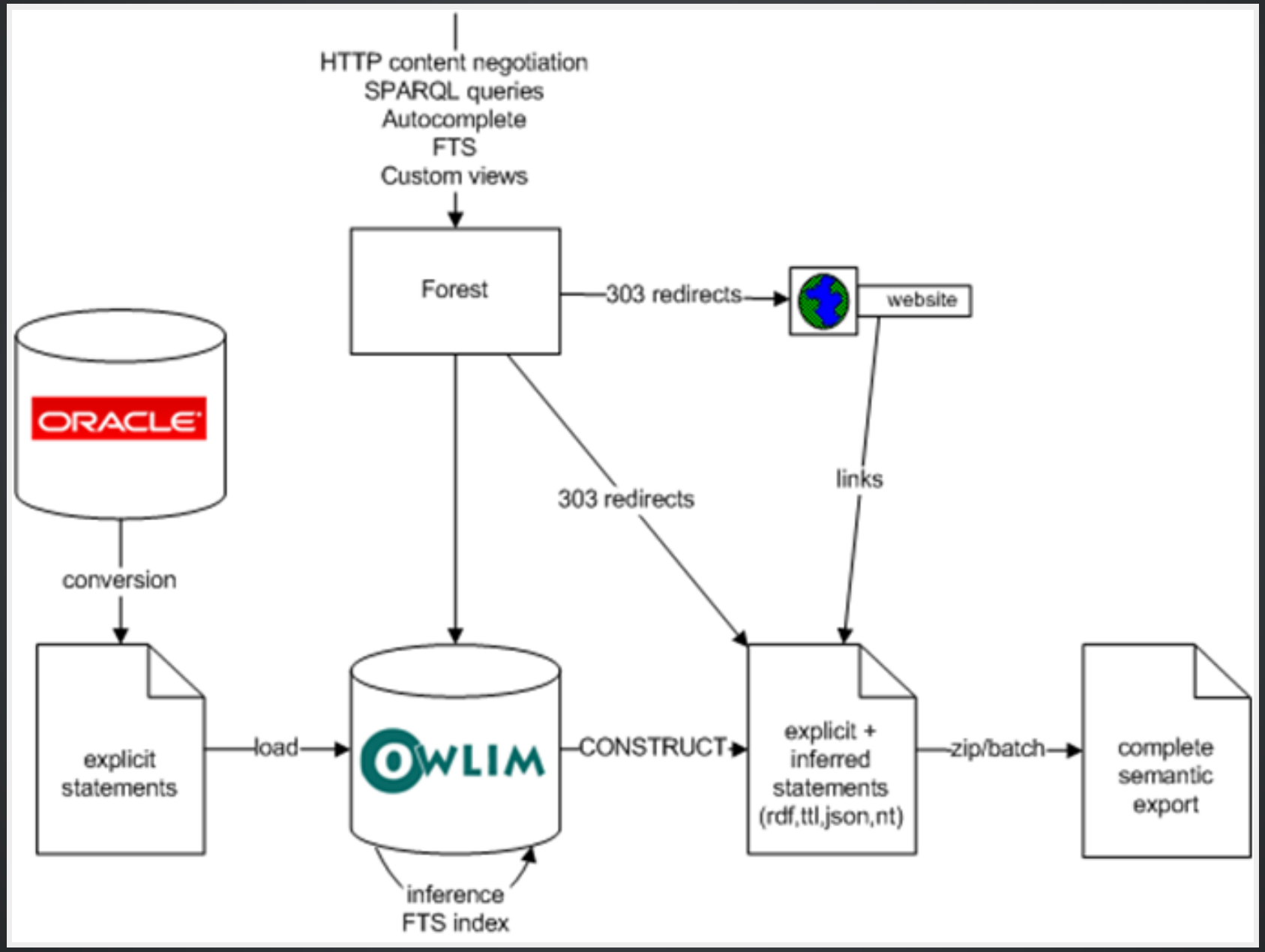


CONTRIBUTION TO ISO 25946

- Contributed to ISO 25946 ontology: http://purl.org/iso25964/skos-thes
- See Linked Open Vocabularies entry
- First industrial use of ISO 25946
- Defined appropriate combinations of BTG, BTP, BTI relations (first formally defined in ISO).

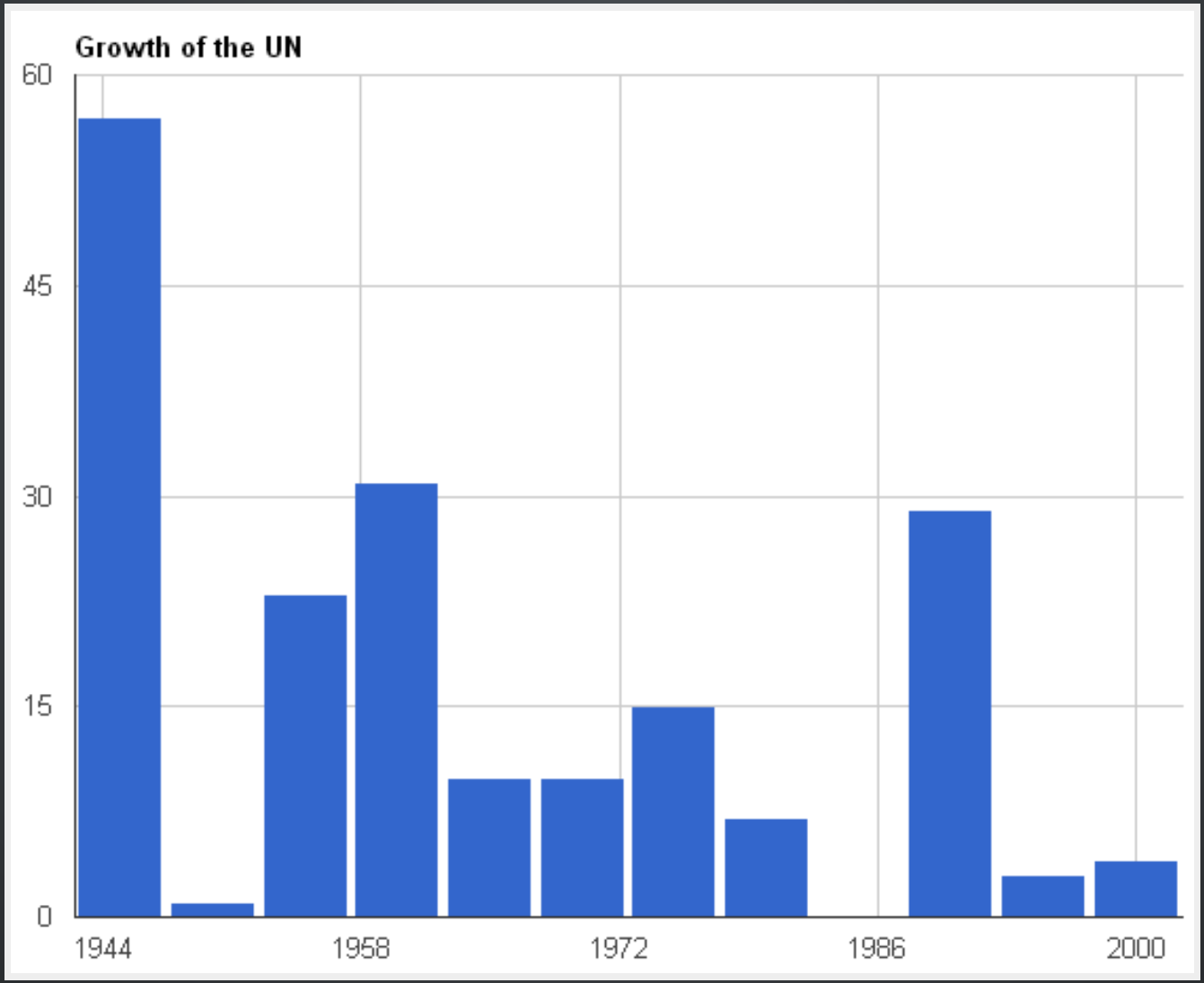
On Compositionality of ISO 25964 Hierarchical Relations (BTG, BTP, BTI), V.Alexiev, J.Lindenthal, A.Isaac. Networked Knowledge Organization Systems (NKOS 2014) Workshop at DL2014, London, 11-12 Sep 2014

GVP LOD ARCHITECTURE



TGN CHARTING WITH SPARQL

Number of members of the UN per year. See doc or jsfiddle with it



POSSIBLE FUTURE TOPICS

- Deploying thesaurus management system (VocBench) based on SKOS, SKOS-XL and semantic repository
- Text analytics and semantic annotation of CH records
- Linguistic Linked Data
- Manuscripts: semantic integration, semantic search, semantic annotation
- Research Infrastructures