

SEMANTIC TECHNOLOGIES FOR CULTURAL HERITAGE

VLADIMIR ALEXIEV, ONTOTEXT CORP

VLADIMIR.ALEXIEV@ONTOTEXT.COM

2014-06-11

TABLE OF CONTENTS

- Ontotext Corp
- Semantic Technologies
- Cultural Heritage LOD Cloud
- ResearchSpace
- Bulgariana
- Europeana Creative
- Getty Vocabularies

SMARTCULTURE CONFERENCE (THANKS FOR INVITING ME!)



ROUND TABLE 1

13h-14h

SMARTMUSEUMS
MUSEUMS OF THE FUTURE

ONTOTEXT CORP

- World Leading semantic technology developer
 - Established in 2000 as part of Sirma Group
 - Spun off in 2008 after venture investment (NEVEQ)
 - 75 employees
 - Offices in Bulgaria (Sofia and Varna), UK (London), USA (New York)
 - Global leader in semantic databases and search
- Proven Delivery
 - More high-profile show cases than competitors
 - Highest profile sem web applications
 - BBC: World Cup 2010, London Olympics 2012
 - Semantic search for multinational pharmaceuticals (Astra Zeneca)
- Stable and Growing
 - Both staff and revenue growing for 12th year in a row

SOME ONTOTEXT CLIENTS



ONTOTEXT RESEARCH PROJECTS (FP5-FP7)

Bulgaria's largest participant: over 30. Domains:
semantic databases, reasoning, semantic search, eGovernment,
life sciences, cultural heritage...

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

- This is from Mar 2009, now has grown 10x!
- Just the part hosted on <http://www.factforge.net>

LINGUISTIC LINKED DATA

FIRST ONTOTEXT GLAM PROJECT: UK NATIONAL ARCHIVES

- 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



CULTURAL HERITAGE LOD CLOUD

EUROPEANA RECOGNIZES IMPORTANCE OF SEMANTIC TECHNOLOGIES



europeana
think culture

Press Release

The Hague – June 1st, 2010

For immediate release

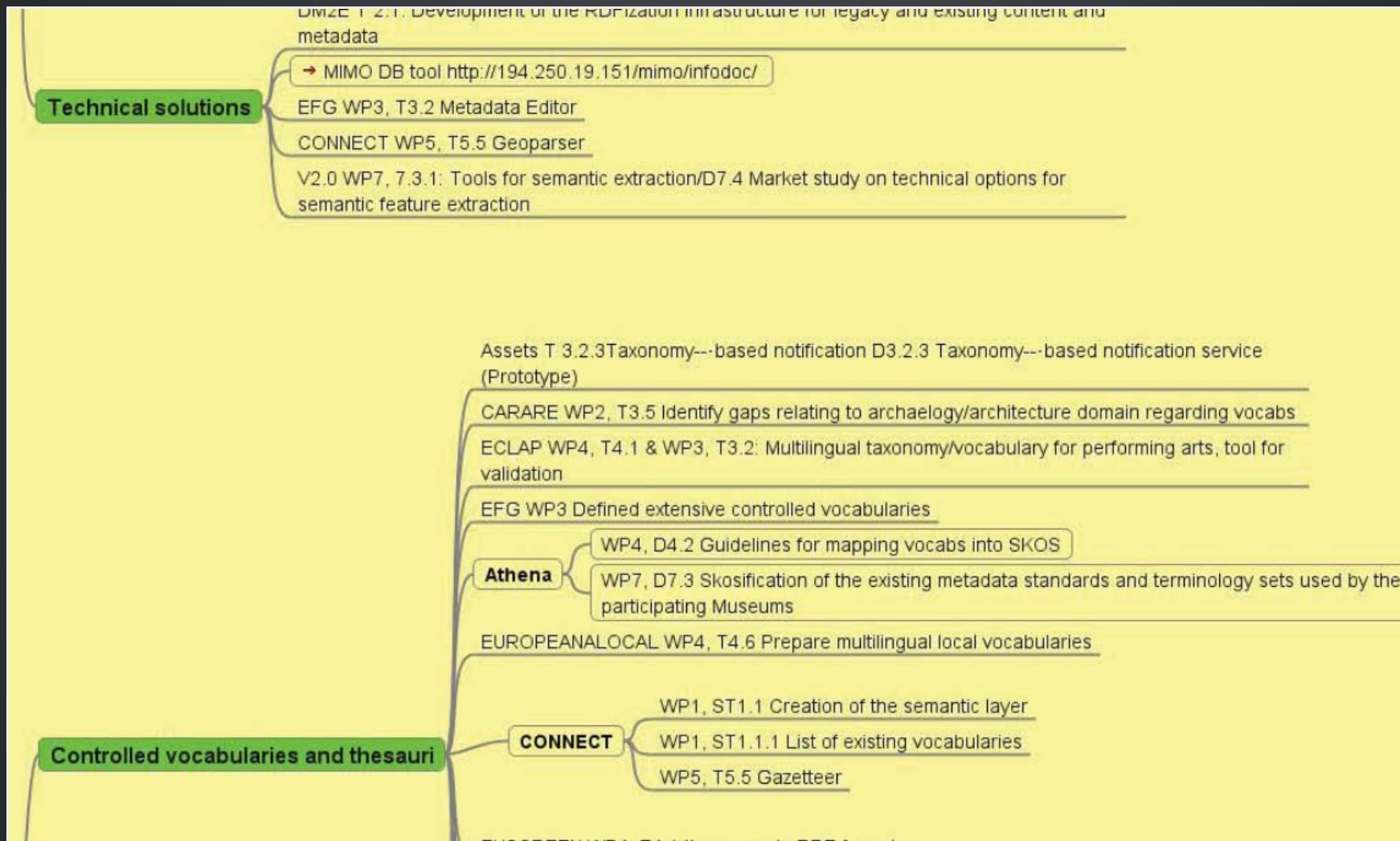
Europeana Publishes First White Paper: Knowledge = Information In Context

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



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 (OWLIM)
- Provided all 2.5M objects of the British Museum as Linked Open Data (LOD)

2.5M BRITISH MUSEUM OBJECTS AS LOD

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_current_institution	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

Find all objects with images created/modified by Rembrandt

and is/has/about drawing and is/has/about mammal +

Search Add To Data Basket Export Print

13 Results 1

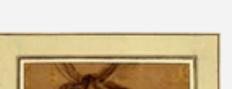
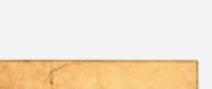
List Thumbnails (selected) Timeline

Object Type
1 album
13 drawing

Creator
1 Anonymous
13 Dutch
2 Italian
2 Jan Baptist Weenix
1 Jan Lievens
12 Rembrandt

Places
13 (others)

sorted by: Title; then by...

			
PDO13612 A horse lying down; with head to right. ... by Jan Lievens, Anonymous, Dutch, and Rembrandt	PDO13924 Study of a pig, facing left. c.1638-1639... by Dutch and Rembrandt	PDO13925 A tethered pig, facing right. c.1638-1639... by Dutch and Rembrandt	PDO13926 A lion drinking from a pail; crouching on... by Dutch and Rembrandt
			

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

RESEARCHSPACE: SEMANTIC DATA ANNOTATION

Projects Tools Administrator Bookmark web link in RS Databasket 11 ▾

Rembrandt search

Dashboard Forum GAA1591 Amethyst gem set in a gold pendant, engraved... ×

GAA1591 Amethyst gem set in a gold pendant, engraved...

Object Details Annotations Relations New Annotation

Add To Data Basket Export Print

Basic

BM Object
Amethyst gem set in a gold pendant, engraved Sphinx seizing a bearded man wearing a ch defends himself with a sword.

Preferred identifier
GAA1591 (BM Public Reference Number)

Identifiers
465772 (BM Codex Id)
Gem 1918* (BM Greece & Roman Number)
1924,0514.6 (BM Registration Number)
(Nuxeo UID: BM-GAA1591)

Current owner
The British Museum

Current keeper
The British Museum

RM Department Greek and Roman

New Annotation

Title: Referred value: Suggested value:

Bought in 1924 at unknown auction.

Add topic... Add

Save Cancel

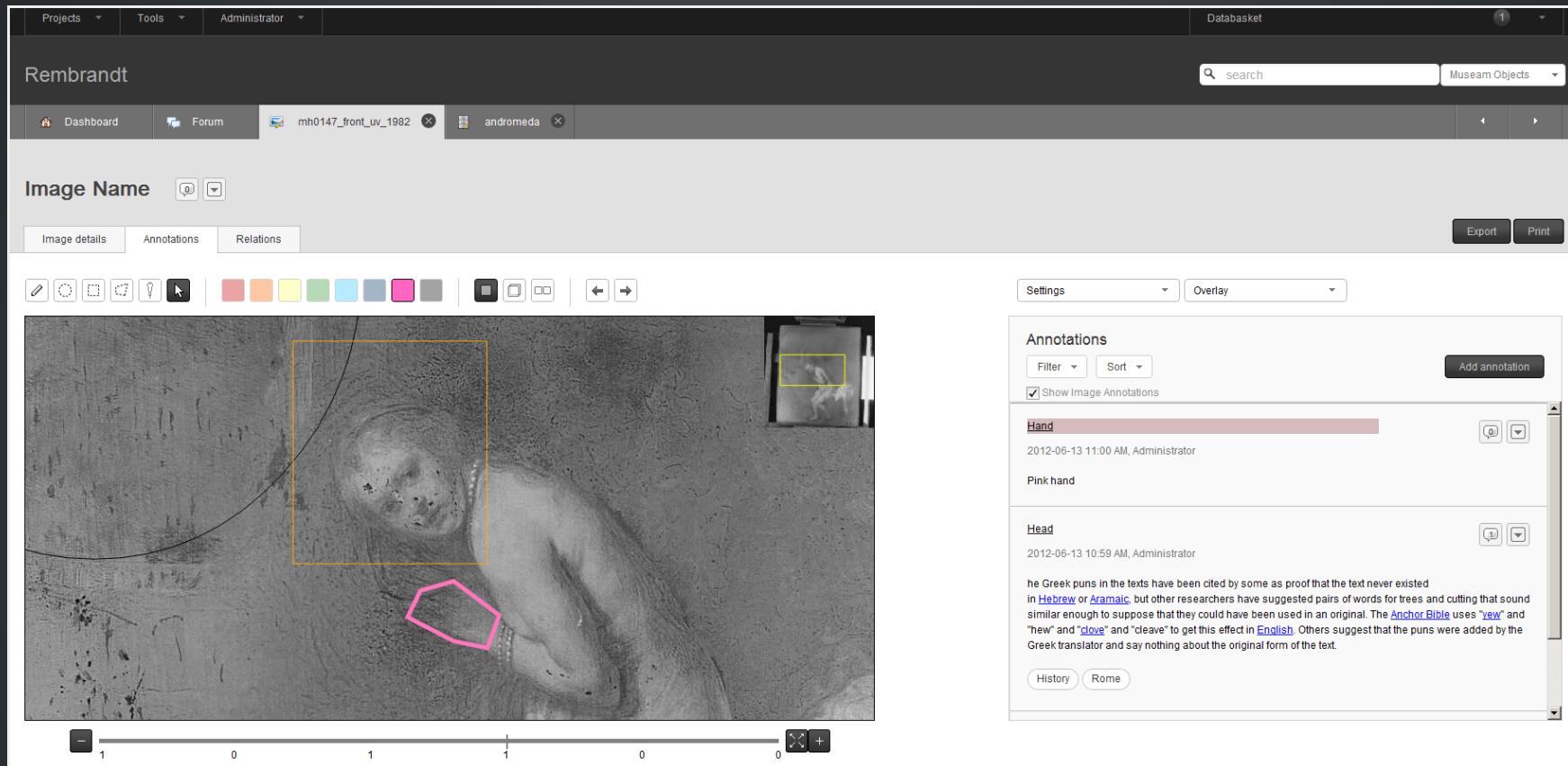
Related Content

Images

[View all images](#)

Documents

RESEARCHSPACE: SEMANTIC IMAGE ANNOTATION



BULGARIANA

- A Bulgarian aggregator to Europeana

EXAMPLE: THRACIAN GOLD (RHYTON)



The collage consists of four images arranged in a triangular pattern. The top-left image shows a Thracian gold rhyton with intricate relief carvings. The top-right image is a photograph of the ancient Thracian fortress of Perperikon, featuring large stone rock formations and a stone wall. The bottom-left image shows a close-up of a stone relief with geometric or animal-like patterns. The bottom-right image depicts a large, ornate Thracian tomb with a golden-yellow hue.

Europeana
think culture

Търсене в Архива Преминете

Bulgarian Heritage > Праисторическа и Тракийска цивилизации > Предмети >

Начало

Моля, използвайте този идентификатор за цитиране или линк към този публикация:
<http://bulgarianheritage.bulgariana.eu/jspui/handle/pub/561>

Разглеждане

Раздели & Колекции
Обекти



RHYTON AT EUROPEANA

- Now any European citizen can find it!

 europeana
think culture

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

[Return to search results](#)


[View](#)



View item at
[Институт за балканистика с Център по тракология](#)

 Share

Riton, страна Б 0

Description: Троянският принц Парис (Александър), седнал между Атина и Хера, е облечен в т. н. фригийско или „тракийско“ облекло – къс, препасан хитон, пътно 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. Назоване 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 take your data and make beautiful apps with it

BULGARIANA COLLECTION FEATURED IN OPEN CULTURE

Bulgarian Collection Featured in Europeana's First Free iPad App

June 5, 2013

[Europeana Released its First Free iPad App](#) called Europeana Open Culture. It is an engaging application featuring 350,000 images from selected Europeana collections, broken down in 5 categories: Maps and Plans, Treasures of Art, Treasures of the Past, Treasures of Nature and Images of the Past.

Jill Cousins, Europeana Executive Director says: "We're really pleased to launch the first ever Europeana app. By downloading Europeana Open Culture, more people can now explore, share and have fun with Europe's cultural heritage"

We are really proud that ancient Bulgarian treasures were selected for this first iPad application! The app showcases ancient gold items of significant historic value, published as part of the [Bulgarian.eu collection](#) "Bulgarian Heritage: Pra-historic and Thracian Civilizations". These items were contributed by the Institute of Balkan Studies and Thracology, Bulgaria (Prof Dr Valeria Fol) and published by Ontotext and <http://bulgariana.eu> as part of a project funded by the Bulgaria-Korea IT Cooperation Center (ITCC).



EUROPEANA CREATIVE

EUROPEANA CREATIVE

- EXPLORE THE EUROPEANA LABS ONLINE PLATFORM
- FIND OUT HOW TO PROVIDE & USE CULTURAL HERITAGE CONTENT
- TAKE ON OUR OPEN INNOVATION CHALLENGES
- BE INSPIRED AND BUILD YOUR OWN APPS AND GAMES

Culture.si photo library, Slovenia. Photo Miran Kambič

Europeana Creative explores new ways to engage with Europe's digital cultural heritage for business, education and entertainment. It seeks creative entrepreneurs to push Europe's rich cultural heritage into the new media era. Europeana Creative develops innovative applications that are focused on education, tourism, social

networks and design. In open innovation challenge events, developers, creative entrepreneurs and cultural heritage experts meet to identify, incubate and spin off viable projects. Europeana Creative has 26 partners from 14 EU member states and is coordinated by the Austrian National Library.



europeana
creative



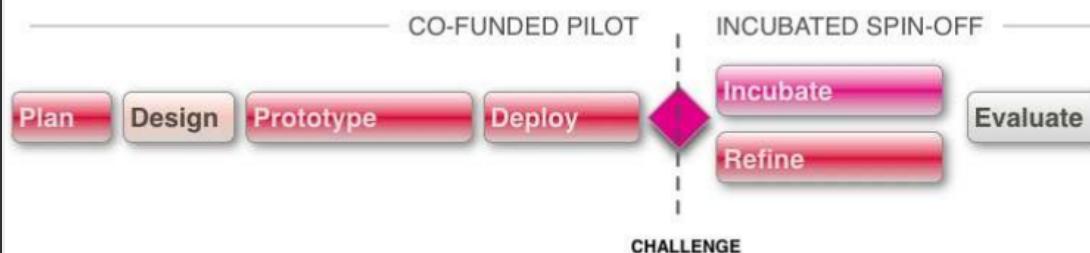
Europeana Creative is co-funded by the European Union, through the ICT Policy Support Programme as part of the Competitiveness and Innovation Framework Programme (CIP).

europeanacreative.eu

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...

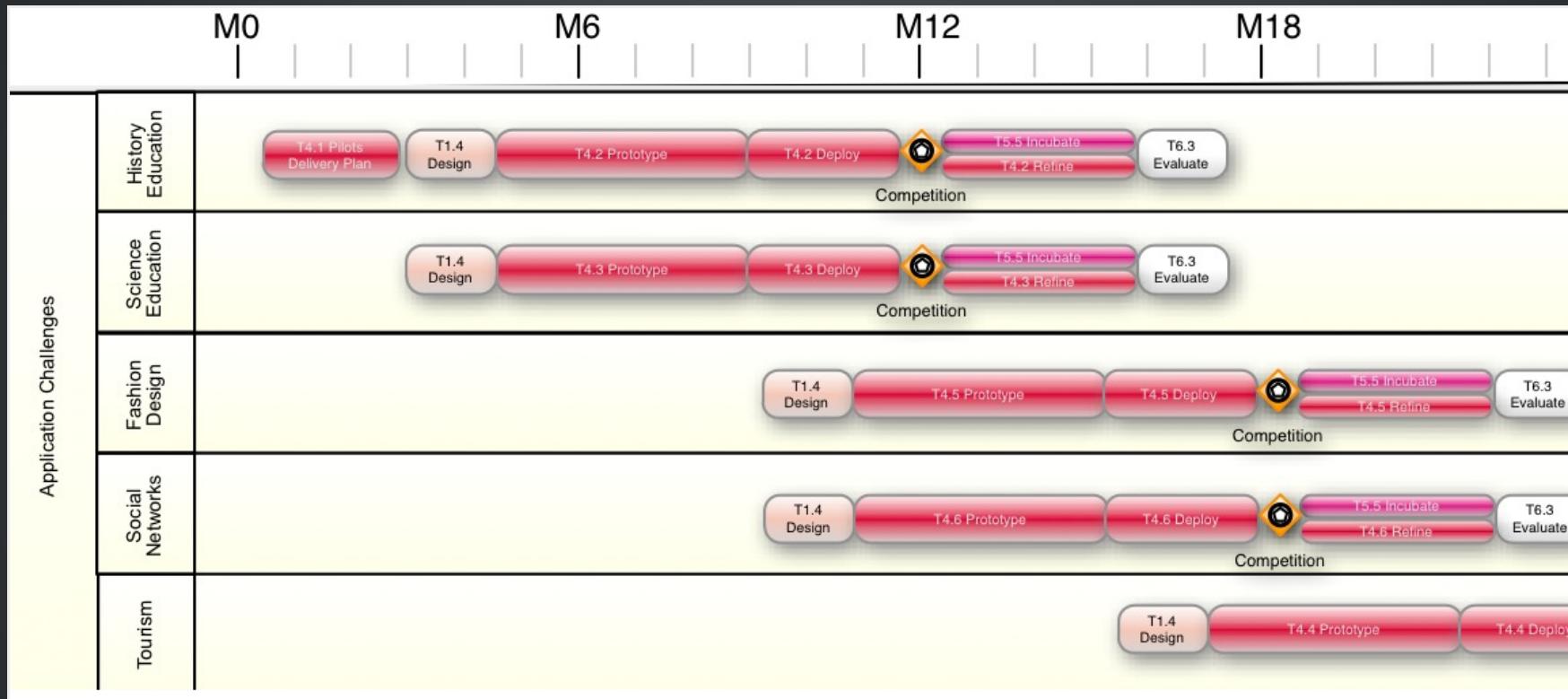
Pilot → Challenge → Spin-Off



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



CREATIVE OPEN CHALLENGES AND INCUBATION SUPPORT

ONTOTEXT IN EUROPEANA

Ontotext works on fundamental backend technologies important for 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, 1.5 years old
- <http://europeana-test.ontotext.com/sparql>
working with Europeana to update it
- Provides SPARQL querying

SPARQL QUERIES

Eg Polish Periodicals by library and decade

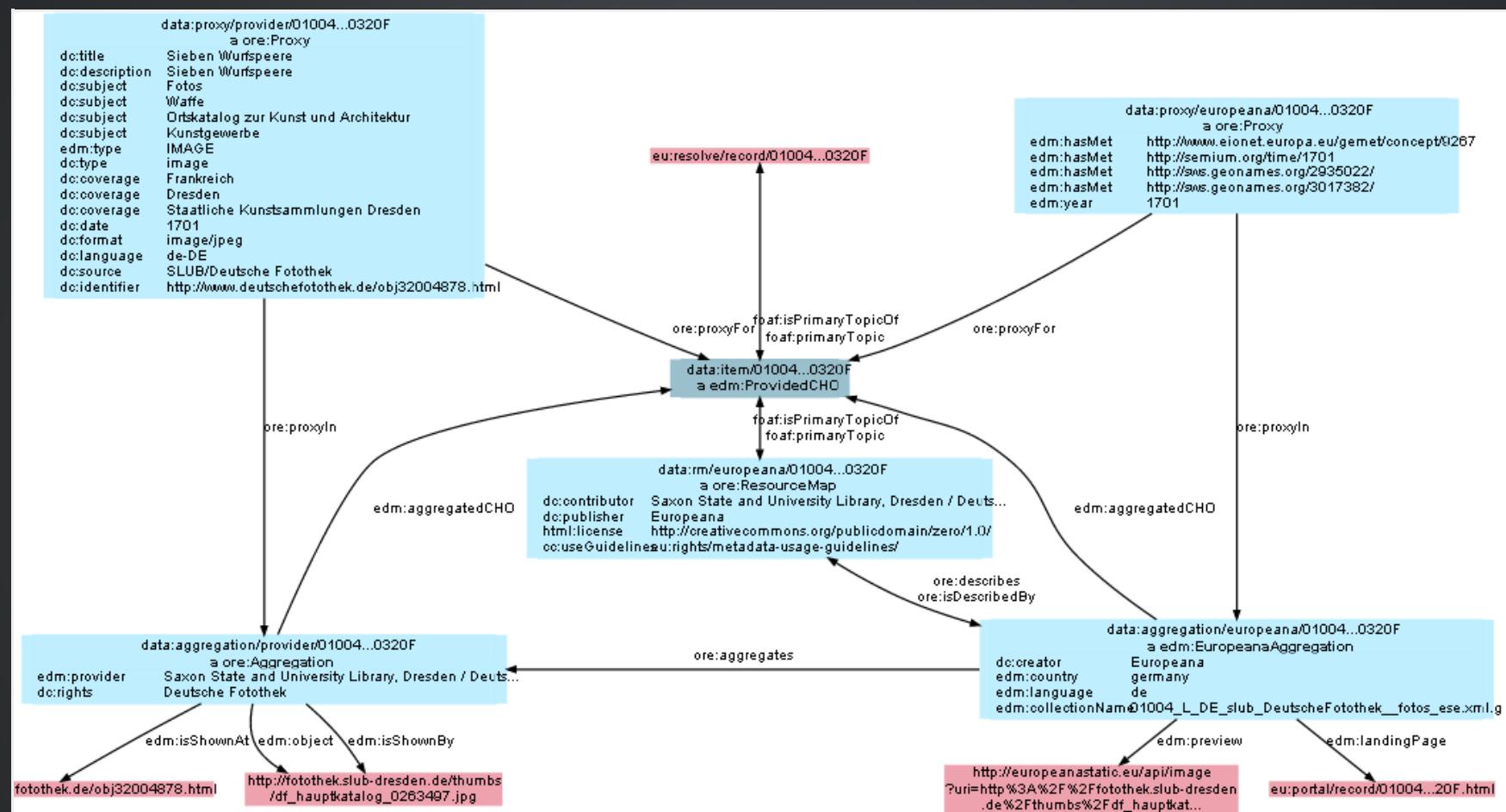
<http://europeana-test.ontotext.com/sparql>

```
select
  ?date
  (sum(?n1) as ?Uniwersytetu_Warszawskiego)
  (sum(?n2) as ?Politechniki_Lubelskiej)
  (sum(?n3) as ?Bałtycka)
{
  ?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 and decade (you can [jsfiddle](#) with it)

EDM OBJECT GRAPH



GETTY VOCABULARIES

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 is publishing them as Linked Open Data:

<http://vocab.getty.edu>

- AAT LOD published in Feb 2014, already sees numerous use cases
- First industrial use of the latest ISO standard on thesauri: ISO 25946
- Contributed to ISO 25946 ontology
- Special session at CIDOC Congress (Dresden, Sep 2014)

USE OF ISO 25946 IN GETTY LOD

