



Bulgariana and Europeana involving Semantic Technologies

Mariana Damova, PhD

*Round Table
Digitalization and Preservation of Cultural Heritage via Europeana
Sofia
November 2012*

Ontotext

- Top-5 provider of core Semantic Technology
- Established in year 2000; offices in Bulgaria, UK, USA
- Active both in research and commercial projects (FP7 funding for 10 years)
- **360° semantic technology – unique portfolio:**
 - **Semantic Databases:** high-performance RDF DBMS, scalable reasoning
 - **Semantic Search:** text-mining (IE), metadata generation, Information Retrieval (IR)
 - **Web Mining:** focused crawling, screen scraping, data fusion
 - **Linked Data Management and Data Integration**

Good recognition in the SemTech community

- Ontotext pages are ranked #1 for “semantic annotation” and “semantic repository” at GYM, #3 for “linked data management” at Google

Several joint ventures and subsidiaries

- Innovantage: leading online recruitment intelligence provider in UK

Ontotext Clients (selected)



British Broadcasting Corporation (BBC)

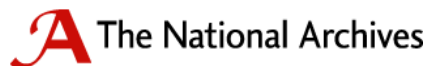
- Run its World Cup 2010 sites on top of OWLIM
- Since Mar'12 BBC Sports
- 2012 Olympics sections are driven by OWLIM and a Concept Extraction service developed by Ontotext



Press Association (UK)

- Analysis of Sports news
- Concept extraction
- Linked data generation

Top-3 USA media (not allowed to name)



The National Archives (UK) contracted Ontotext to implement semantic KB and semantic search for the Government Web Archive



British Museum (UK) Ontotext leads the development of Phase 3 of ResearchSpace project on collaborative research in cultural heritage; British Museum's public SPARQL end-point is powered by OWLIM



de Bibliotheek (Holland) aggregation of data from 150 library databases



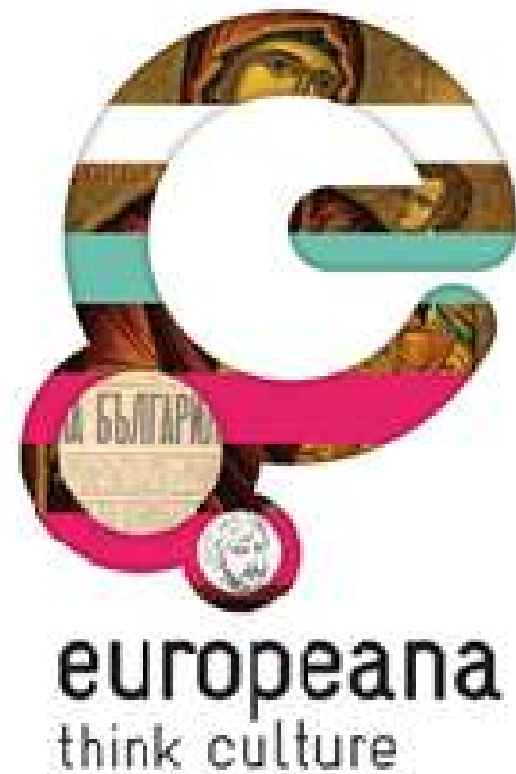
Outline

- bulgariana.eu
 - Community Building
 - Collections
 - Digital Repository – Bulgarian Heritage
- Europeana
 - Europeana Data Standards
- Semantic Technologies
 - Overview
 - OWLIM
- Europeana Semantic Data and Ontotext Technology
- Upcoming ...

bulgariana.eu

bulgariana.eu

- BG aggregator for Europeana
- Community building around BG Cultural Heritage



Community Building

- Google group
 - <http://groups.google.com/group/cultural-heritage-digitalisation> (about 50 members)
- Collaboration with IMI and UNIBIT
- Meeting in Sofia 30.01.2012 (75 participants)
 - endorsed by several Ministries and one member of the European Parliament
- Broadcast at Bulgarian National Radio
- Working group for the Ministry of Culture
- Communication with the Ministry of Culture
- Communication with Municipalities

Community Building

- Meeting in Veliko Tyrnovo 19.03.2012 (about 50 participants)
 - Ontotext, IMI- BAS, Regional Library of Veliko Tyrnovo organizers
- Participation in Europeana Plenary June 2012
- Meeting in Varna 17.10.2012 (about 50 participants)
 - Ontotext, Regional Library of Varna organizers
- Round Table with the Ministry of Culture 09.11.2012
- Upcoming meeting in March 2013
- Expert group about Digitalization with participation of NGOs
- Facebook page
- EU project SmartCulture
- Pilots ...

Collections

Collections

Golden Pages from the Bulgarian Renaissance

Златни страници от Българското Възраждане

unique manuscripts of Bulgarian folk songs collected in 19th century
by Miladinov Brothers, renowned Bulgarian Folklorists
published in 2008 by D-r Luchia Antonova,
Institute of Bulgarian Language, Bulgarian Academy of Sciences



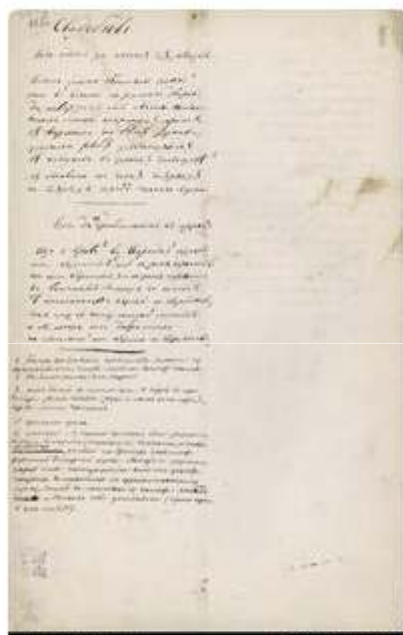
МАРКО КРАЛЕВИКИ БОЛЕН СЕ КАИТ И СЕ
ИСПОВЕДВИТ

Поболил се Марко Кралевике,
що си лежал токму три години,
от нищо се иляч (1) не на'ож'ал.
И му рече негва стара майка:
"Ай ти, Марко, ай ти, синко милий;
не си болен, синко, от господа,
тук си болен, синко, от гре'о'и,
да ти викна попой (2), ду'овници,
лепо да се синко исповедиш,
да си кажиш твоите гре'о'и!"

....

Golden Pages from the Bulgarian Renaissance in Europeana

[Return to search results](#)



View item at [Институт за Литература при БАН](#)

Rights: Институт за Литература при БАН

Кога да се приближават в църков

Contributor: [Лиляна Минкова](#) и [Лучия Антонова-Василева](#)

Time period: 19 век, Български ренесанс

Geographic coverage: [Вардарска Македония](#)

Тип: [народна песен, ръкопис](#) | ▶

Subject: [Сватбена](#) | ▶

Relation: Лиляна Минкова, Лучия Антонова-Василева. Ръкопис на Константин Миладинов в архива на Измаил Срезневски, Академично издателство "Проф. Марин Дринов", София, 2008.

Description: 1) Богатое воображение измышляе сплети из васильков ясли, чтобы наиболее чествовать сватов; 2) Большая патка для жаренья; 3) рев детей в смысле гула; 4) Порой в Горной Болгарии зовется поток: "кърви се леат като порой" - кровь льется потоком; 5) гремит гром; 6) умильно - я помню сомнения одного учившегося в России болгарина, выраженное в печати, о слове

[See more](#) ▶

Data provider: [Институт за Литература при БАН](#) | ▶

Provider: [Ontotext](#) | ▶ [Bulgaria](#) | ▶

Explore further!

Collections

Pra-historic and Thracian Civilizations

Праисторическа и Тракийска цивилизация

Unpublished Thracian archeological objects collected by Prof. Valeria Fol, Center of Thracology at the Institute for Balkan Studies at the Bulgarian Academy of Sciences



Pra-historic and Thracian Civilizations in Europeana



Explore Europe's cultural collections

europaana_dataProvider:"Институт за балканистика с Цен

Search

[Refine your search](#)

[Help](#)

Matches for: europaana_dataProvider:"Институт за балканистика с Център по тракология"

Filter your search:

By media type

IMAGE (500)

By language

By country

By copyright

By provider

Include content contributed by users



Results 1 - 12 of 500 Page: [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) →



Like



+1



Начелник



Съкровище от
Вълчитрън



Четири апликации за
конска амуниция



Идол



Bulgariana – a registered aggregator to Europeana

www.europeana.eu/portal/europeana-providers.html

Search... Search Safe Do Not Track Weather Facebook Speedtest

Slovenian National E-content Aggregator (1,900) ▶
Erfgoed Brabant (1,845) ▶
Nationale Aggregator (1,641) ▶
[Spletna galerija, Semantika d.o.o. \(1,517\)](#)
Ville de Bourg-en-Bresse (1,414) ▶
Universitätsbibliothek Regensburg (1,321) ▶
University College Dublin, UCD Library (1,246) ▶
[Niedersächsische Staats- und Universitätsbibliothek Göttingen \(1,125\)](#)
[Universitat Autònoma de Barcelona \(1,097\)](#)
[Saxon State and University Library, Dresden \(1,081\)](#)
[ICAB \(1,012\)](#)
[Arxiu Històric de Girona \(999\)](#)
Mediateca de la Universidad de Oviedo - Encycloasturias (986) ▶
EuropeanaConnect - EOD (975) ▶
Ontotext (516) ▼

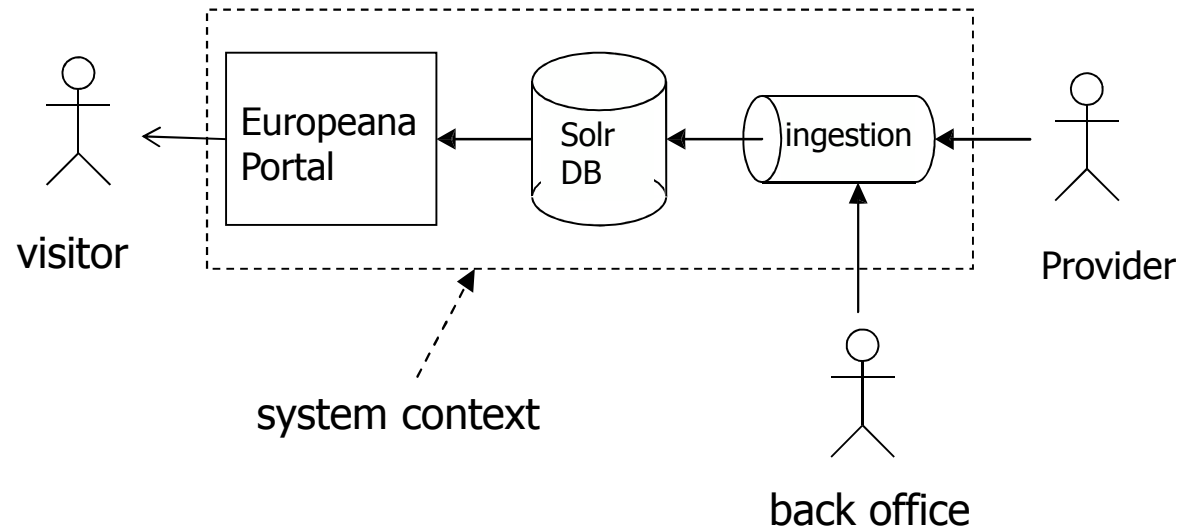
- [Ontotext \(516\)](#)
- [Институт за балканистика с Център по тракология \(500\)](#)
- [Институт за Литература при БАН \(14\)](#)
- [Институт за балканистика с Център по тракология - БАН \(2\)](#)

[Institut d'Estudis Ilerdencs. Servei d'Arxiu i Llegats \(457\)](#)
[Bibliothèques Virtuelles Humanistes \(384\)](#)
[Universitat Politècnica de Catalunya \(223\)](#)
[Universitat de Girona. Biblioteca. Fons Bertrana \(196\)](#)
["Department of Information Management, Hacettepe University" \(170\)](#)
[Biblioteca de Comerç i Turisme de la Generalitat de Catalunya \(131\)](#)
[Swiss National Library \(100\)](#)
[Helsingin kaupunginkirjasto \(97\)](#)

since February-March 2012

Europeana

- Collection types: Image, Sound, Video, Text
- Present Europeana Architecture



- Europeana data standards
- Europeana aggregators (by country or cultural heritage sector)
- Process of ingesting content (4-6 weeks)

Digital Repository – Bulgarian Heritage


Bulgarian Heritage



Вхх

<p>Търсене в Архива</p> <input type="text"/> <input type="button" value="Преминете"/> <p>→ Начало</p> <p>Разглеждане</p> <p>→ Раздели & Колекции</p> <p>→ Обекти</p>	<p>Bulgarian Heritage ></p> <p>Bulgariana ви приветства с новото си дигитално хранилище Bulgarian Heritage</p> <p>Очакват ви вълнуващи срещи с българското културно наследство.</p> <p>Повече информация за Bulgariana може да получите на http://bulgariana.eu/</p> <p>Търсене</p> <p>Въведете текст на кирилица в полето, за да търсите в нашите колекции.</p>	<p>Bulgariana</p> <p>Новини</p> <p>Българското участие в Европеана. Сътрудничество и развитие 17.10.2012 в град Варна се проведе поредната среща-семинар за българското културно наследство и</p>
---	---	--

Bulgarian Heritage



europaana
think culture

Търсене в Архива

→ Начало

Разглеждане
→ Раздели & Колекции
→ Обекти

Bulgarian Heritage >

Раздели и Колекции

Това са нашите колекции. Кликнете върху името на някоя от тях, за да видите началната страница на раздел или колекция.

Златни Страници от Българското Възраждане
Братя Миладинови

Праисторическа и Тракийска цивилизации
Въоръжение
Градове
Гробници
Мегалити
Некрополи
Предмети
Светилища
Селища
Съкровища

Links

- <http://bulgariana.eu>
- <http://bulgarianheritage.bulgariana.eu>
- <http://www.europeana.eu>
 - europeana_collectionName: 20215*
 - for the individual sets use europeana_collectionName: 2021501* (or 2021502*)
- <http://britishmuseum.ontotext.com>

Europeana

Europeana



<http://www.europeana.eu>

- Launched in 2008
- Project funded by the European Commission
- Based in the National Library of the Netherlands, the Koninklijke Bibliotheek
- Goal to make Europe's cultural and scientific heritage accessible to the public
- Over 180 heritage and knowledge organizations and IT experts across Europe
- Europeana Collection: 5M objects in 2009, 10M in 2010, 20M at present
- Endorsed by the European parliament in 2010
- **2011** "Comité des Sages" makes recommendations about Europeana
to put online the collections held by Europe's libraries, archives, museums and audiovisual archives – vast numbers of books and periodicals (there are some 2.5bn items in Europe's libraries alone), and millions of hours of film and video covering the whole of Europe's diverse history and culture.
- September 2012 published all Europeana data for re-use via API
- Europeana 2.0

Europeana Data Standards

Europeana Data Standards

- Unified metadata

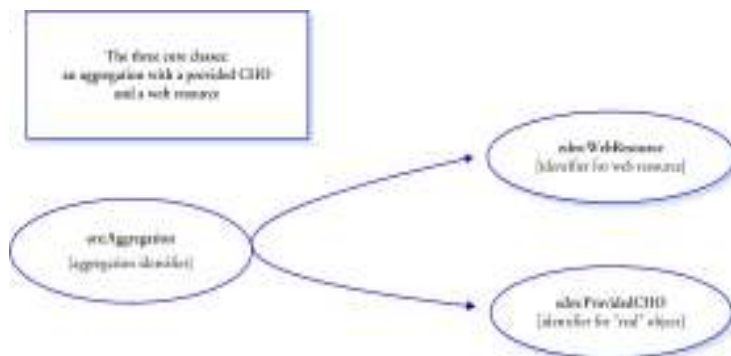
- ESE – Europeana Semantic Elements

- DublinCore & Europeana fields
 - 36 fields: flat, limited ability semantic links

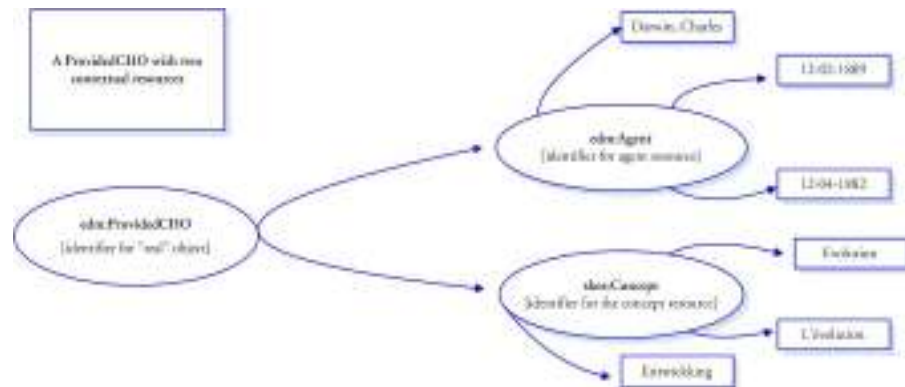
dc:title
 dc:creator
 dc:subject
 dc:description
 dc:publisher
 ...

europeana:provider
 europeana:dataProvider
 europeana:rights
 europeana:type
 europeana:isShownBy and/or europeana:isShownAt
 ...

- EDM - Europeana Data Model



Basic data model



Two contextual classes

Semantic Technologies

Tim Berners-Lee



The founder of the World Wide Web.

In 1998 he proposes his vision for the Semantic Web in a paper called "Semantic Web Road Map"

<http://www.w3.org/DesignIssues/Semantic>

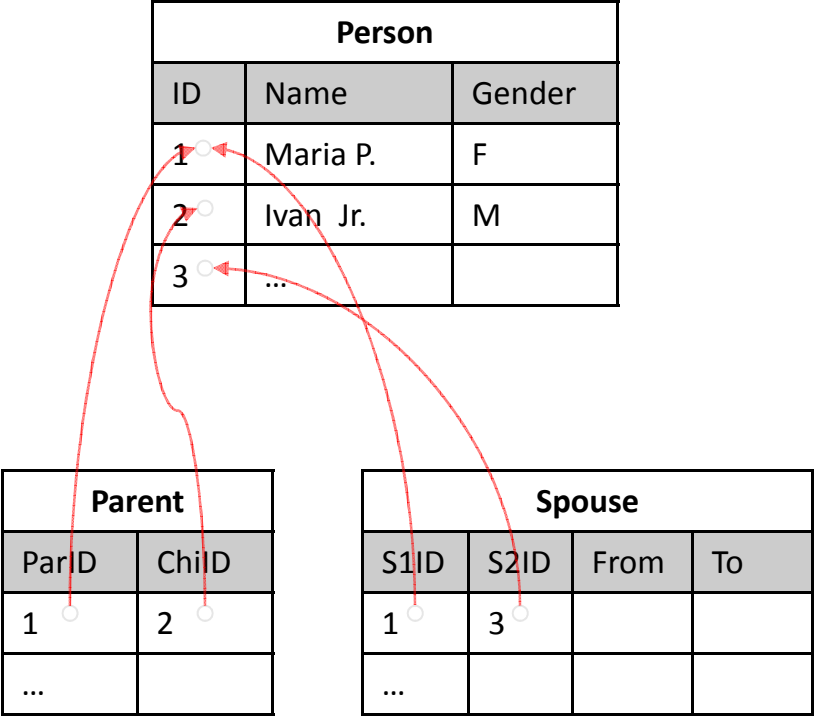
This vision receives further development in Tim Berners-Lee et al., "The Semantic Web", Scientific American, May 2001

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation." [Berners-Lee et al. 2001]

Semantic Technologies: Major Characteristics

- *Easy integration of multiple data-sources*
 - once the schemata of these sources is semantically aligned, the inference capabilities of the engine supports the interlinking and combination of the facts from the different sources;
- *Easy querying against rich or diverse data schemata*
 - inference is applied to match semantics of the query to the semantics of the data, regardless of the vocabulary and the data modeling patterns used for encoding of the data;
- *Based on open standards (W3C)*
 - RDF, RDFs
 - OWL
 - SPARQL

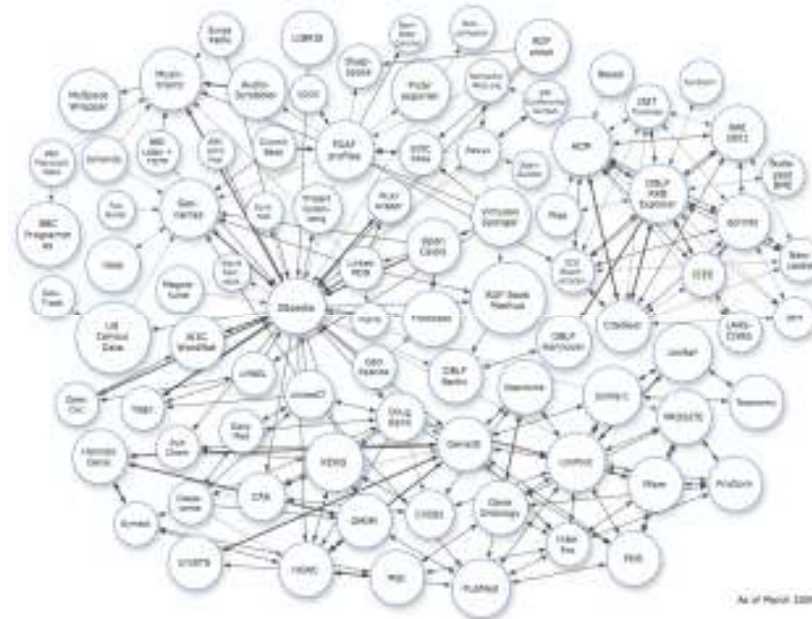
Physical data representation: RDF vs. RDBMS



Statement		
Subject	Predicate	Object
myo:Person	rdf:type	rdfs:Class
myo:gender	rdfs:type	rdfs:Property
myo:parent	rdfs:range	myo:Person
myo:spouse	rdfs:range	myo:Person
myd:Maria	rdf:type	myo:Person
myd:Maria	rdf:label	"Maria P."
myd:Maria	myo:gender	"F" ○
myd:Maria	rdf:label	"Ivan Jr."
myd:Ivan	myo:gender	"M" ○
myd:Maria	myo:parent	Myd:Ivan
myd:Maria	myo:spouse	myd:John
...	○	○

Linking Open Data

- Linking Open Data (LOD) W3C SWEO Community project
<http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>



- Initiative for publishing “linked data” – a set of principles, which allows browsing of RDF data, spread across different servers, in the way HTML is browsed

OWLIM

OWLIM is a family of semantic repositories, or RDF database management systems, with the following characteristics:

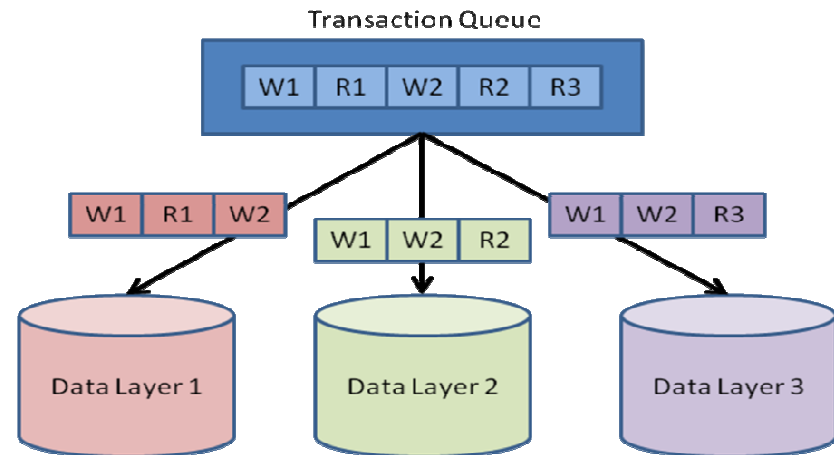
- native RDF engines, implemented in Java
- delivering full performance through both Sesame and Jena
- robust support for the semantics of RDFS, OWL 2 RL and OWL 2 QL
- best scalability, loading and query evaluation performance

OWLIM is used in a large number of research projects and software tools. Independent opinions justifying our bold claims are referred to here.

The presentation Lowering the Cost of Data and Content Integration and enabling Searching and Querying of Billions of Facts on the Web presents the key features of OWLIM alongside an introduction to the benefits of using RDF databases for data integration and a discussion on linked data management.

OWLIM Replication Cluster

- Distribution through data replication is used to ensure:
 - Better handling of concurrent user requests
 - Failover support
- How does it work?
 - Every user request is pushed in a transaction queue
 - Each data write request is multiplexed to all repository instances
 - Each read request is dispatched to one of the instance only
 - To ensure load-balancing, each read requests is send to the instance with smallest execution queue at this point in time



Europeana Semantic Data and Ontotext Technology

Europeana Data in EDM

- 268GB of data
- cultural objects data and linkages to other datasets

Dataset size:

NumberOfStatements=3,899,531,218

NumberOfExplicitStatements=993,332,911

NumberOfEntities=264,523,842

EDM model

SKOS

Prototype available at

<http://europeana.ontotext.com>

to become

-> <http://data.europeana.eu>



data.europeana.eu

Welcome on the SPARQL end-point of [data.europeana.eu](#)!

data.europeana.eu currently contains open metadata on 20 million texts, images, videos and sounds gathered by Europeana. Data is following the terms of the [Creative Commons CC0 public domain dedication](#). Data is described the [Resource Description Framework \(RDF\)](#) format, and structured using the [Europeana Data Model \(EDM\)](#). We give more detail on the EDM data we publish on the [technical details page](#).

Please take the time to check out the [list of collections currently included in the pilot](#).

The terms of use and external data sources appearing at data.europeana.eu are provided on the [Europeana Data sources](#) page.

Sample queries are available on the [sparql](#) page.

Powered by [OWLIM](#). © 2009-2012 Ontotext AD. All rights reserved.



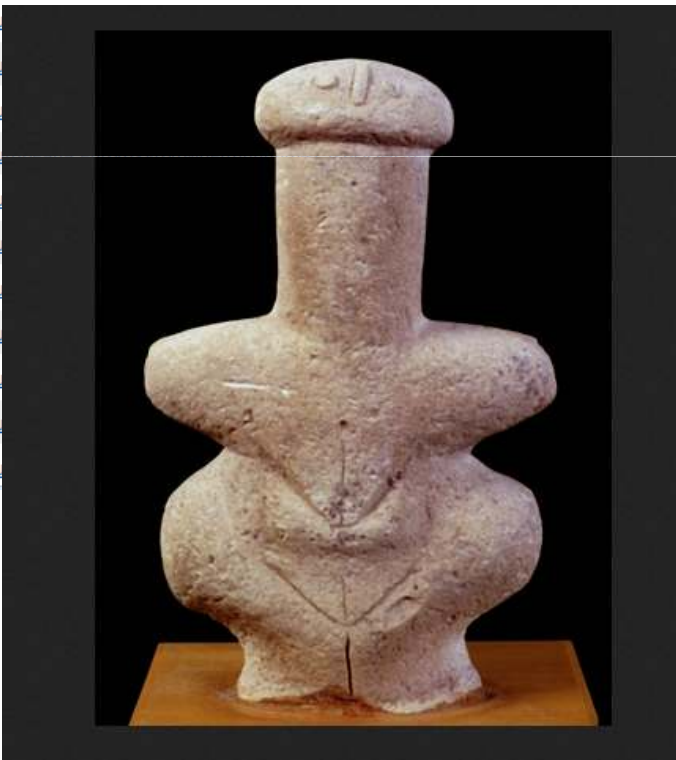
Europeana objects with their images

SPARQL Query

Results for # Europeana objects... (100 of 0)

View as [Exhibit](#) Download SPARQL Results in: [JSON](#) | [XML](#)

object	image
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.planetware.com/i/photo/the-hague-nl282.jpg
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.mcw.gov.cy/mcw/da/da.nsf/All/8308A7AB9CA6CB5EC...
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.theeuropeanlibrary.org/portal/images/treasures...
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.theeuropeanlibrary.org/portal/images/treasures...
http://data.europeana.eu/proxy/provider/00000/047326BDAC0...	http://www.theeuropeanlibrary.org/portal/images/treasures...



Upcoming ...

Europeana Creative - PSP project

lead by the Austrian National Library

26 partners

Objective: experimenting with re-use of cultural content for creativity

Project: Europeana re-use framework and 6 pilots in different domains such as education, tourism, etc.

Ontotext: participate in the infrastructure for re-use with the semantic repository OWLIM, and data integration

Upcoming ...

- Further collaboration with Europeana
- Based on this Round Table
 - Initiatives for building financial instruments for Bulgaria
- Community building
- Pilots with Bulgarian Partners
- EU project proposals for EU PSP Content for Europeana

Thank you for your attention!

mariana.damova@ontotext.com