



# Semantic Technologies for Cultural Heritage

## **Europeana Publishing of Bulgarian Cultural Heritage**

P1.1 Dev Semantic web pilot system

Mariana Damova, PhD  
[mariana.damova@ontotext.com](mailto:mariana.damova@ontotext.com)





# Ontotext Corp



- Who is Ontotext?
  - The leading semtech company in Eastern Europe and one of the leaders world-wide
  - 55 people: Bulgaria (Sofia, Varna), Austria, USA
  - Worked in this area since 2000
  - Venture funding and commercial clients since 2008
  - Bulgaria's most successful participant in EU FP5,6,7 research
  - [www.ontotext.com](http://www.ontotext.com)
- **360-degree semantic technology:**
  - Semantic Repository (OWLIM), ETL
  - Text Mining, Semantic Annotation and Search (KIM, U.Sheffield GATE, Teamware, MIMIR)
  - Web Mining and Crawling
  - Ontology Engineering and Exploitation
  - Master Data and Linked Data Management





# Ontotext Clients (selected)



- British Broadcasting Corporation (BBC)
  - Runs its World Cup 2010 sites on top of OWLIM
  - Next is BBC Sports (2011) and the **2012 Olympics**
- The National Archives (UK)
  - The UK Government's official archive contracted Ontotext to implement semantic search for the Government Web Archive
- AstraZeneca
  - Analysis and retrieval of clinical trial reports
  - Integration of biomedical databases for drug target identification
- KT
  - **Applications of OWLIM in Interactive TV and IPTV**
- All these clients need to:
  - integrate massive amounts of heterogeneous data
  - provide efficient search and navigation



# Outline



- Europeana
- bulgariana.eu
- Collections
- Europeana Data Standards
- Metadata mapping, conversion and ingestion
- Digital repository
- Conclusion



# Europeana



<http://www.europeana.eu>

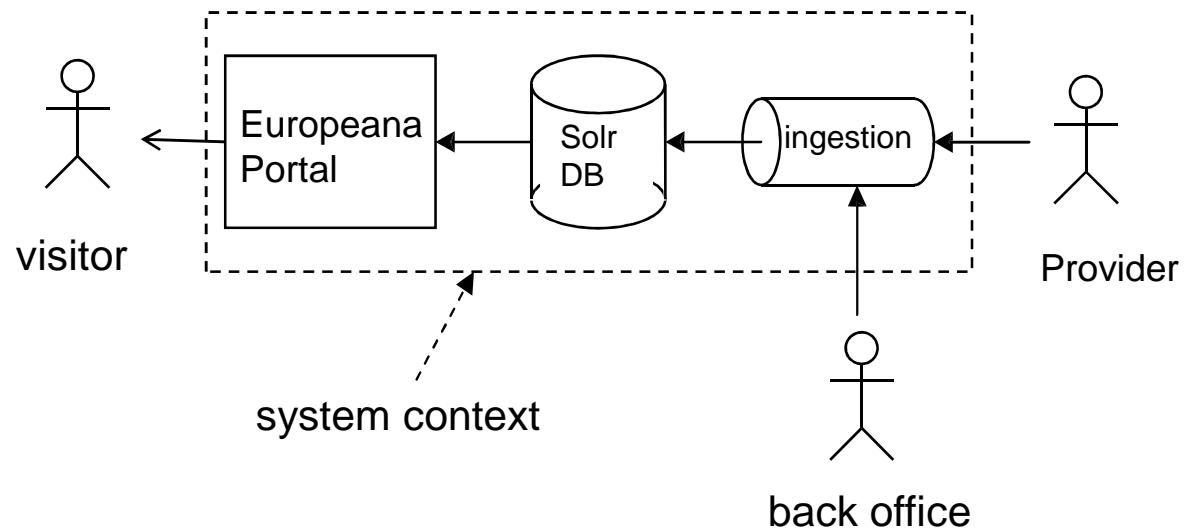
- Launched in 2008
- Project funded by the European Commission
- Based in the National Library of the Netherlands, the Koninklijke Bibliotheek
- Goal to make Europe's cultural and scientific heritage accessible to the public
- Over 180 heritage and knowledge organizations and IT experts across Europe
- 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.*



# Europeana



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



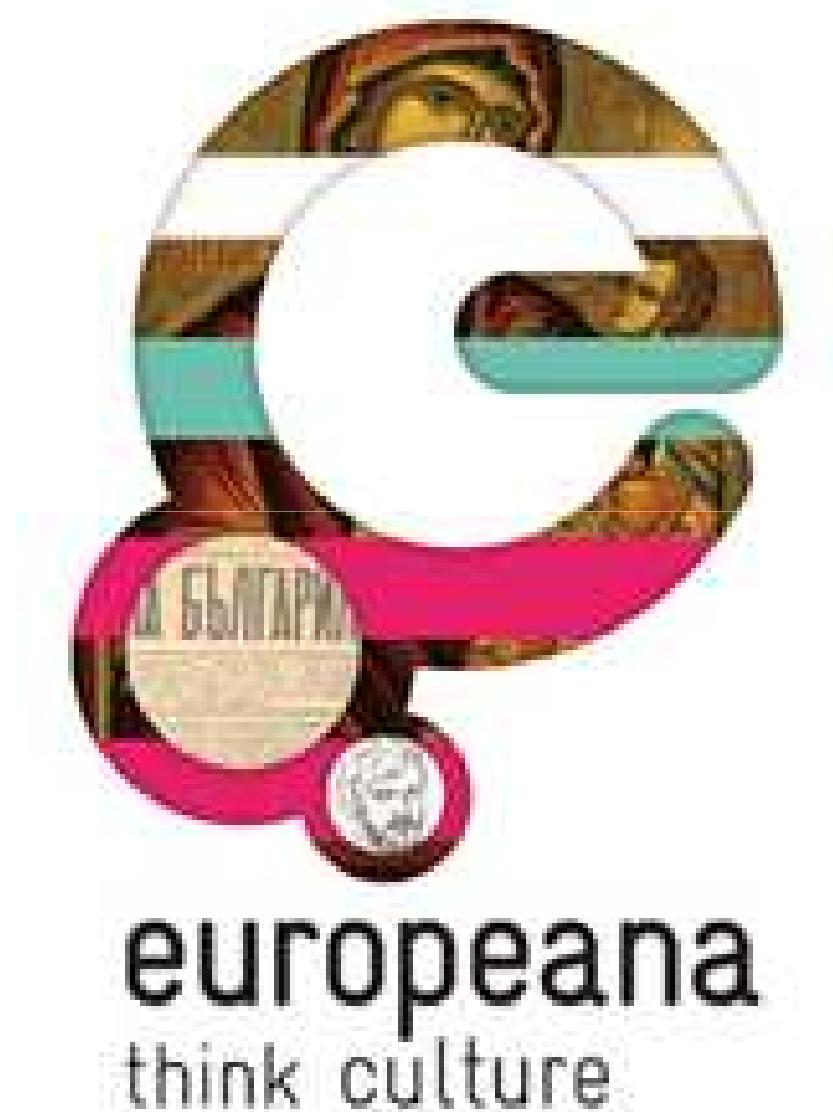
- Europeana data standards
- Europeana aggregators (by country or cultural heritage sector)
- Process of ingesting content (4-6 weeks)



# **bulgariana.eu**



- Main Purpose: BG aggregator for Europeana
- Secondary Purpose: networking and special interest group for BG Cultural Heritage





# Community Building



Google Group: Cultural Heritage Digitalisation

Meetings with

- professionals from the cultural heritage domain
  - Archeologists, Folklorists, Epigraphists, Thracologists, Researchers on Medieval Bulgaria, Museum workers
- professionals from the digitalization domain
  - IMI
- professionals who work on creating metadata for representing cultural heritage

Conferences

- Digital Preservation and Presentation of Cultural and Scientific Heritage (september 2011)

Identification of digital collections of Bulgarian Cultural Heritage

- F-museum
- Regional Museum of Kjustendil
- WebFolk

Connection with Europeana Officials and Technology Creators



# Collections



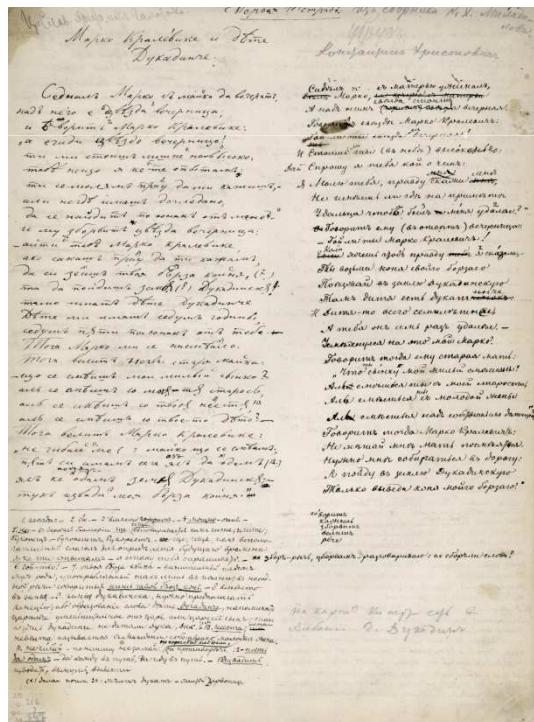
# Collections



## Golden Pages from the Bulgarian Renaissance

### Златни страници от Българското Възраждане

unique manuscripts of Bulgarian folk songs collected in 19<sup>th</sup> 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!”

....



# 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





# Collections



WebFolk

Веда Словена

an online collection of Bulgarian folk songs gathered in 1997 by a team from the Bulgarian Academy of Sciences

Website funded by

“Open Society” Foundation , available in English and Bulgarian



full audio record		audio example
NR	1	
PRIMARY CLASSIFICATION	Instrumental	
SECONDARY CLASSIFICATION	Entertainment (pleasure, fun)	
THEME	Magic	
SOCIAL FUNCTION	No particular occasion	
FORM	Not analysed	
TEXTURE	Monophonic	
METER	Non-measured	
WAY OF PERFORMANCE	Solo kaval (wood. instr.)	
ETHNIC DATA	Bulgarian	
TYPE	Authentic	
RECORDED BY THE FOLKLORIST	Petur Lyondev	
DATE OF RECORD	1966	
PLACE OF RECORD	Bojounitsa	
ARCHIVE ITEM - NUMBER	241.4048	
RECORDING MEDIA	Recording tape	
INFORMANT/S	Yordan	
INFORMANT - PLACE OF BIRTH	Bojounitsa	
COUNTY	Pleven	
FOLK MUSIC AREA	Middle North Bulgaria	

Originally, a collection made by the Bosnian-Croatian Stefan Verkovitch, 1874  
Entitled “Bulgarian Folk Songs from the Prehistoric and Prechristian Times”

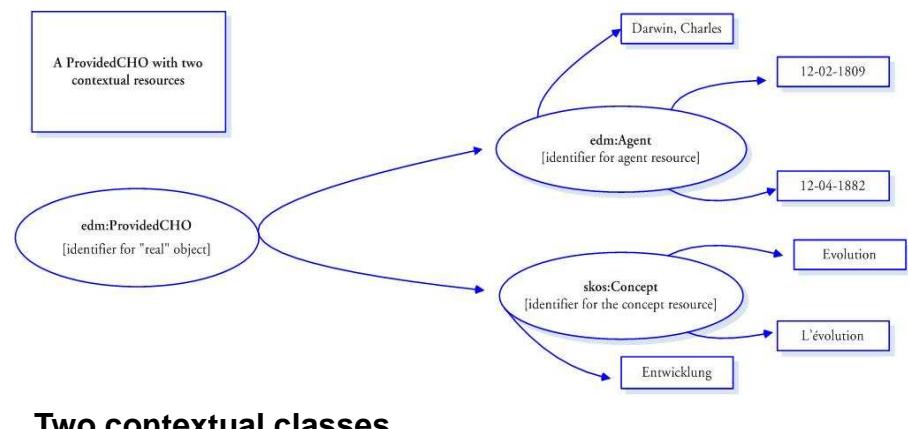
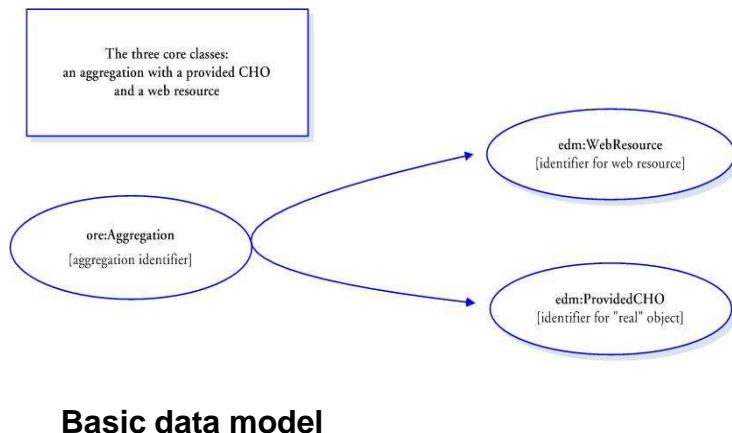


# Europeana Data Standards

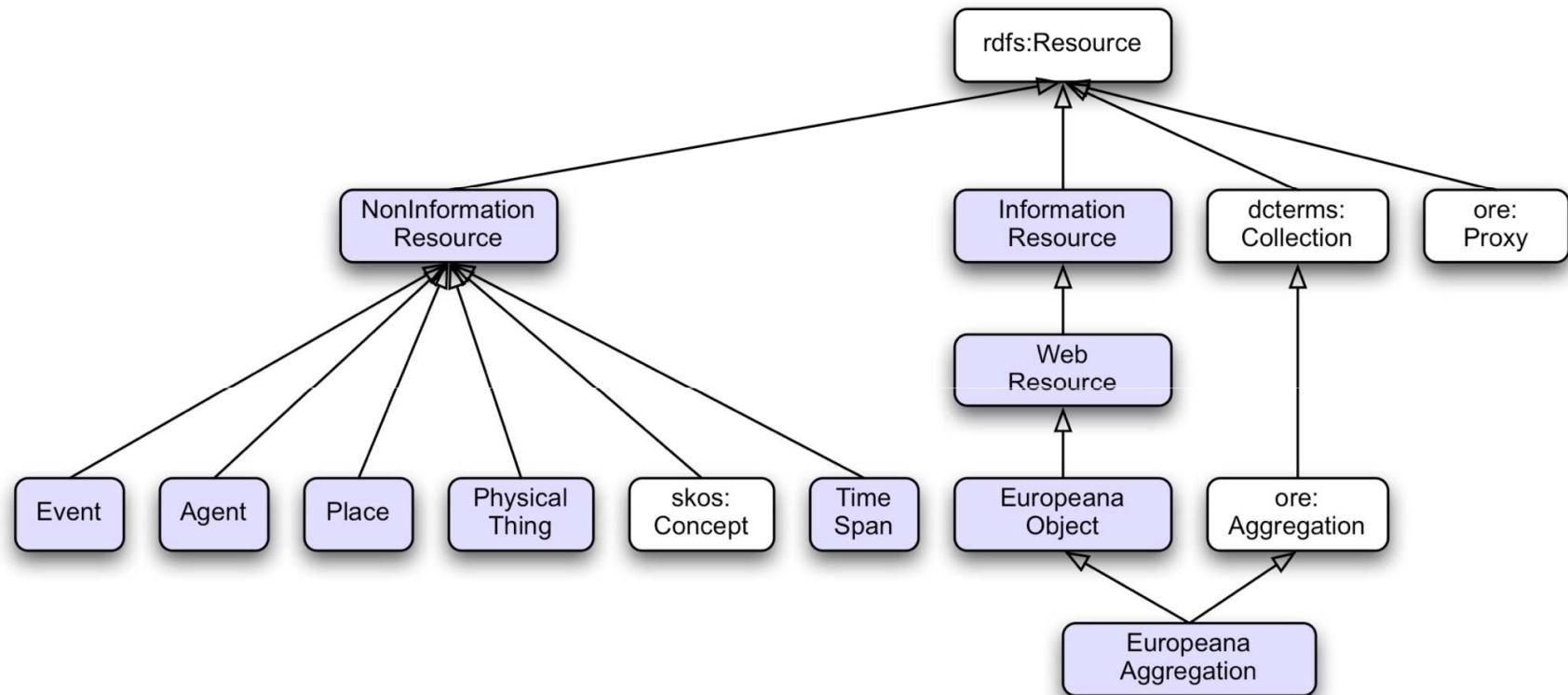
# Europeana Data Standards

- Unified metadata
  - ESE – Europeana Semantic Elements
    - DublinCore & Europeana fields
    - 36 fields: flat, limited ability semantic links
 

dc:title	europeana:provider
dc:creator	europeana:dataProvider
dc:subject	europeana:rights
dc:description	europeana:type
dc:publisher	europeana:isShownBy and/or europeana:isShownAt
...	...
  - EDM - Europeana Data Model

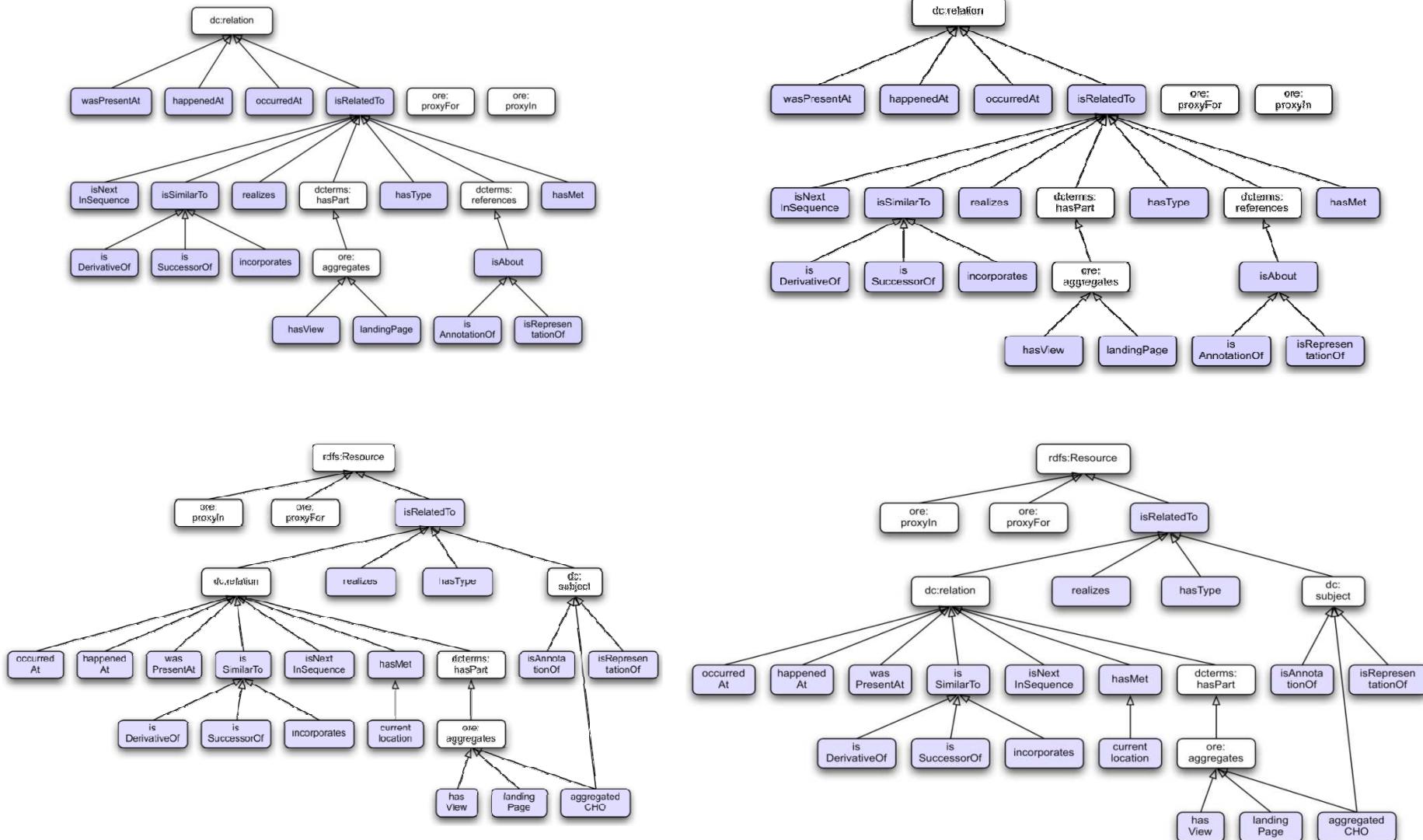


# EDM Classes

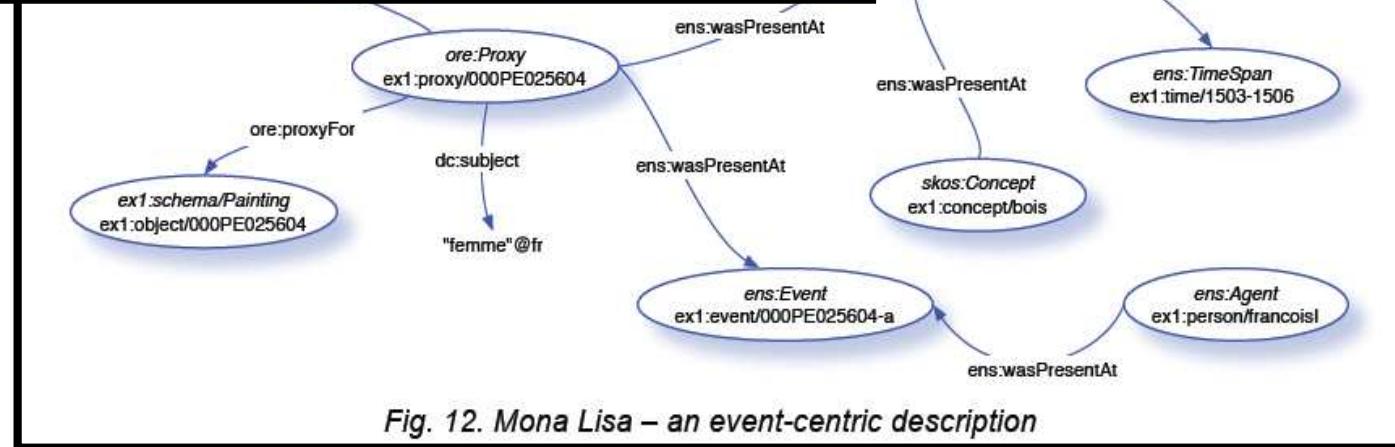
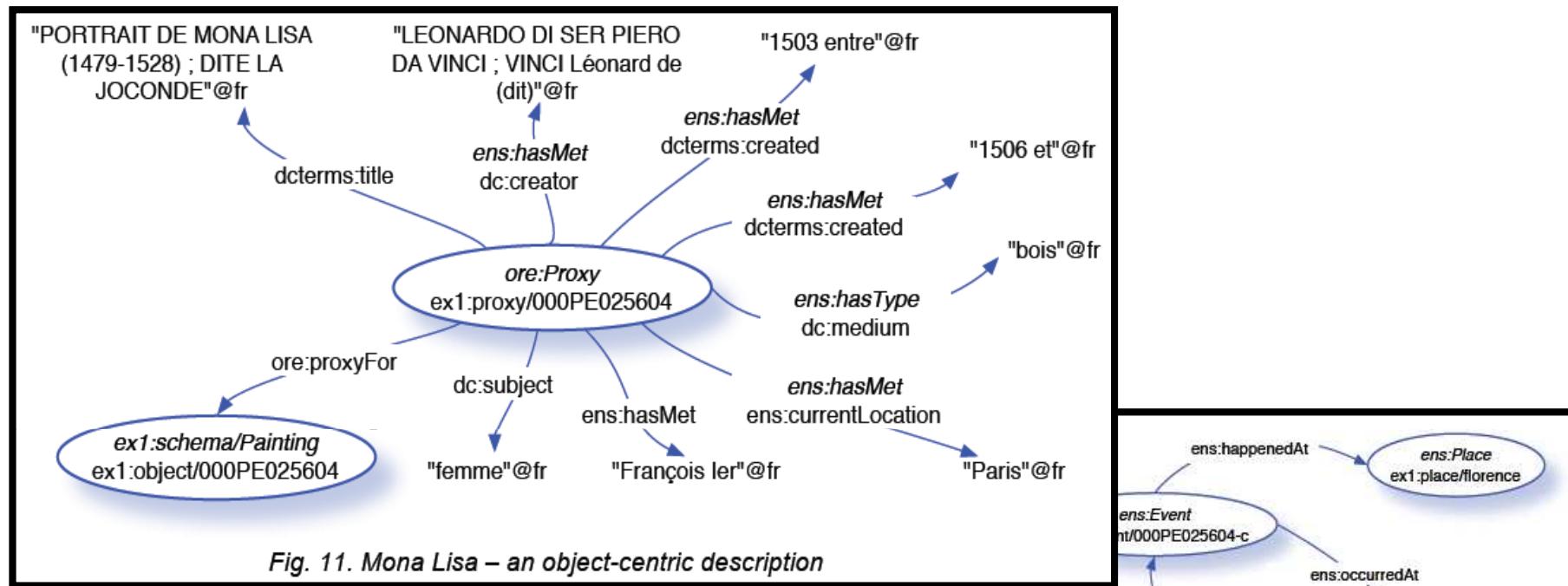


- Reuses 4 other ontologies (shown in white)
- This is the true spirit of semweb

# EDM Properties



# EDM Example: Mona Lisa





# Metadata mapping, conversion and ingestion



# Metadata mapping

- WebFolk

	WebFolk		Europeana	
	Veda Slovena - EN	Veda Slovena - BG	Source	Element
1	TITLE / LYRICS	заглавие и текст	DC	title
2	THEME	тематика	DC	subject
3	RECORDED BY THE FOLKLORIST	записвач	DC	contributor
4	DATE OF RECORD	година на записа	DC	date
5	PRIMARY CLASSIFICATION	музикална класификация	DC	type*
6	RECORDING MEDIA	начин на записване	DC	format
7	ARCHIVE ITEM - NUMBER	архивна сигнатура	DC	identifier
8	FORM	форма	Bulgariana	FORM
9	METER	размер	Bulgariana	METER
10	TEXTURE	фактура	Bulgariana	TEXTURE
11	WAY OF PERFORMANCE	начин на изпълнение	Bulgariana	WAY OF PERFORMANCE
12	SECONDARY CLASSIFICATION	функционална класификация	Bulgariana	SECONDARY CLASSIFICATION
13	SOCIAL FUNCTION	битова функция	Bulgariana	SOCIAL FUNCTION
14	INFORMANT/S	изпълнява	Bulgariana	INFORMANT/S
15	ETHNIC DATA	етнос	Bulgariana	ETHNIC DATA
16	INFORMANT - PLACE OF BIRTH	родно място на изпълнителя	Bulgariana	INFORMANT - PLACE OF BIRTH
17	COUNTY	околия	Bulgariana	COUNTY
18	PLACE OF RECORD	място на записа	Bulgariana	PLACE OF RECORD
19	TYPE	артефакти/типология	Bulgariana	TYPE
20	ARCHIVE ELABORATION	архивна обработка	Bulgariana	ARCHIVE ELABORATION
21			Europeana	country
22			Europeana	language*
23			Europeana	uri
24			Europeana	userTag
25			Europeana	year



# Metadata conversion

- WebFolk
  - crawling the website
  - processing HTML to XML
  - mapping schema

**DELVING**  
tool used  
to convert  
1400 entries  
into  
ESE

The screenshot shows the DELVING SIP Creator application. The main window displays a list of local data sets:

Local Data Sets
bg02test6
csyb47_01
en02
en02_fix

The bottom window shows a detailed view of the 'en02\_fix' mapping:

- Parsed Record:** Shows input fields like 'input.full\_audio\_record' pointing to 'Multimedia/Audio/M1000001.mp3'.
- Groovy Code:** Displays the mapping logic:

```
input.RECORDED_BY_THE_FOLKLORIST *{  
    dc.contributor it  
}  
input.DATE_OF_RECORD *{  
    dc.date it  
}  
input.RECORDING_MEDIA *{  
    dc.format it  
}  
input.ARCHIVE_ITEM_NUMBER *{  
    dc.identifier it  
}  
dc.language 'en'  
input.THEME *{  
    dc.subject it  
}  
input.TITLE_LYRICS % / *{  
    dc.title it  
}  
input.PRIMARY_CLASSIFICATION *{  
    dc.type it  
}
```
- Output Record:** Shows the generated XML output:

```
<record>  
<dc:contributor>Petur Lyndev</dc:contributor>  
<dc:date>1966</dc:date>  
<dc:format>Recording tape</dc:format>  
<dc:identifier>241.4048</dc:identifier>  
<dc:language>en</dc:language>  
<dc:subject>Magic</dc:subject>  
<dc:title>Tip = Dopunitejni danni za tipologijata</dc:title>  
<dc:type>Instrumental</dc:type>  
<europena:isShownAt>http://musicart.limbm.bas.bg/en/karton.asp?zapisID=1</europena:isShownAt>  
<europena:language>en</europena:language>  
<europena:provider>BULGARIAN ACADEMIC MUSIC PORTAL</europena:provider>  
<europena:rights>http://creativecommons.org/publicdomain/mark/1.0/</europena:rights>  
<europena:type>SOUND</europena:type>  
<europena:uri>en02_fix/F629AE44B7B3DCFED444D363E626EDF411EC69A8</europena:uri>  
<europena:userTag>Tip = Dopunitejni danni za tipologijata</europena:userTag>  
</record>
```



# Metadata ingestion

- administrative steps
- ftp access provided to Europeana
- ingestion tests



# The other Collections

## Process

- Metadata mapping
- Conversion of all data in ESE format
  - manually
  - with DELVING (original Europeana tool) or MINT (used by FP7 Athena)
- Preparation for Ingestion

## Pre-historic and Thracian Civilizations

- digitalization and metadata assignment
- storing in digital repository for cultural heritage
- preparing for Europeana publishing
- publishing in Europeana early 2012

## Golden Pages from the Bulgarian Renaissance

- specifying metadata
- storing in digital repository for cultural heritage
- preparing for Europeana publishing
- publishing in Europeana early 2012



# Digital Repository for Cultural Heritage



# Digital Repository for Cultural Heritage



## Sirma Media –

The mission of Sirma Media is to become a leading developer and supplier of interactive educational and cultural heritage environments enabling active acquisition of knowledge through direct experience.

- Selection of technical solution/platform to be used as digital repository for the needs of the project
- Setup and configuration of the digital repository – account management – organizations, users, roles
- Definition of the different types of entities – manuscripts, articles, artifacts, archeological sites, bibliographical records
- Elaboration of the metadata properties in accordance with the content providers requirements
- Migration of databases and digitalized artifacts from available online resources and cultural bodies collections
- Training of users to work with the admin panel of the digital repository – metadata input and editing, media files upload
- Publication of the digitalized collections on the web – UI layer enabling rich visualization, various search options, browse by thematic categories, etc...



# Conclusion

Aggregator for Bulgarian Cultural Heritage to Europeana



Thank you for your attention!