



Bulgariana Collections in Europeana

Mariana Damova, PhD

*Veliko Tyrnovo
March 2012*

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
- **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
- **British Museum (UK)** ResearchSpace project funded by the Andrew W. Mellon Foundation support collaborative web-based research, information sharing for the cultural heritage scholarly community, a consortium lead by Ontotext
- **LODAC (Linked Open Data in Academia)** Japan's National Institute of Informatics and aggregates various information across multiple Japanese resources as LOD
- **The Polish Digital National Museum** aggregates artifacts from cultural institutions in the Digital Libraries Federation PIONIER Network: over 70 contributing institutions including universities, libraries, museums, archives, research.
- **The Gothenburg City Museum** provided close to 9K museum objects from two collections to build a use case within the MOLTO FP7 project for a knowledge representation infrastructure that allows querying RDF and presenting RDF results in natural language.
- **Bibliothek, The Hague**, aggregation of data from 150 library databases

Outline

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

Europeana

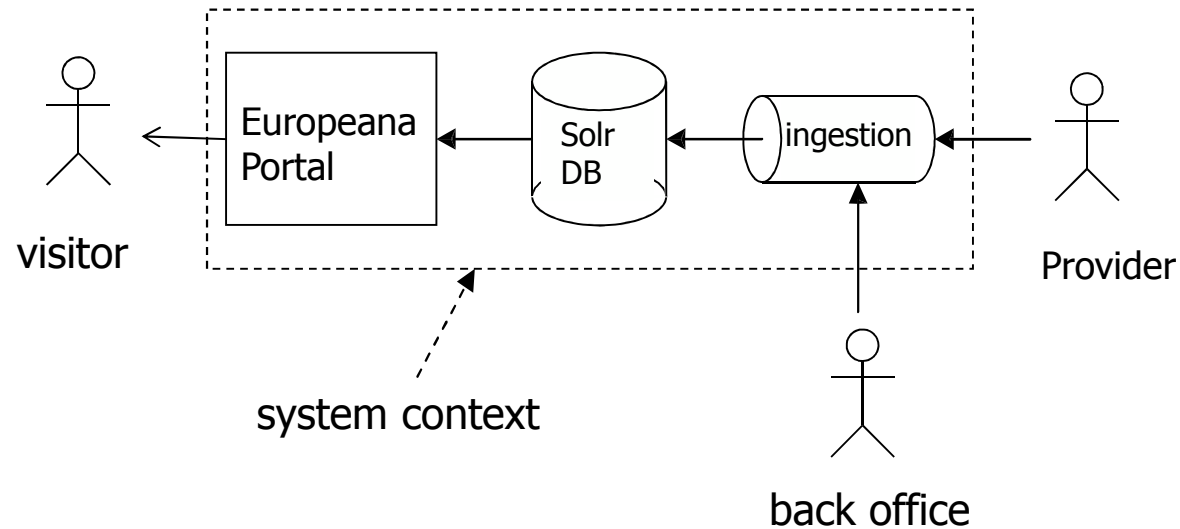


<http://www.europeana.eu>

- 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

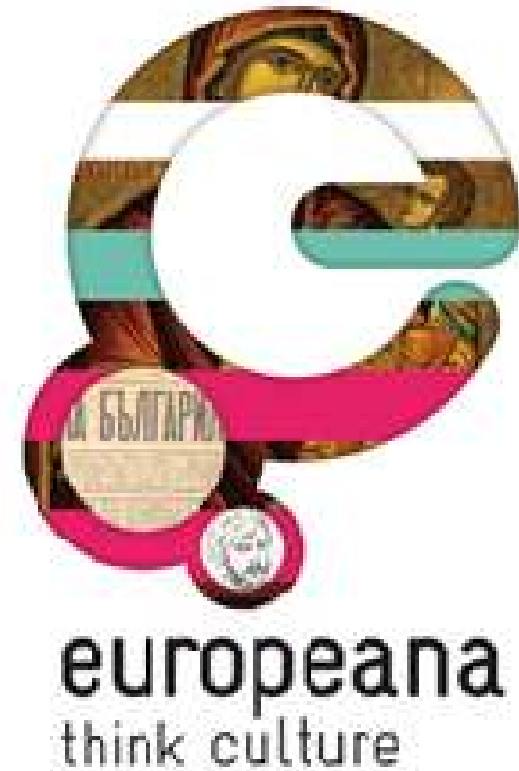


- Europeana data standards
- Europeana aggregators (by country or cultural heritage sector)
- Process of ingesting content (4-6 weeks)

bulgariana.eu

bulgariana.eu

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



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!"

....

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



Europeana Data Standards

Europeana Data Standards

- Unified metadata

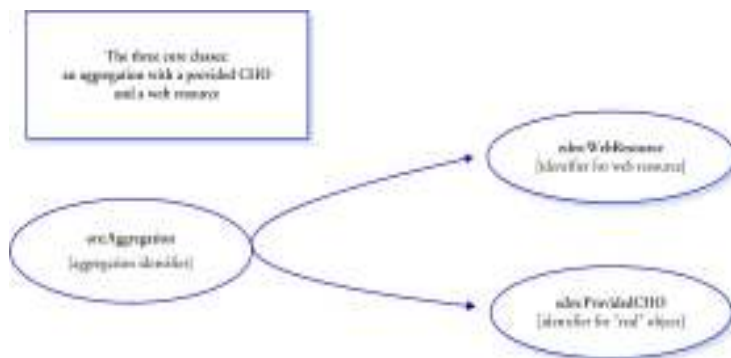
- ESE – Europeana Semantic Elements

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

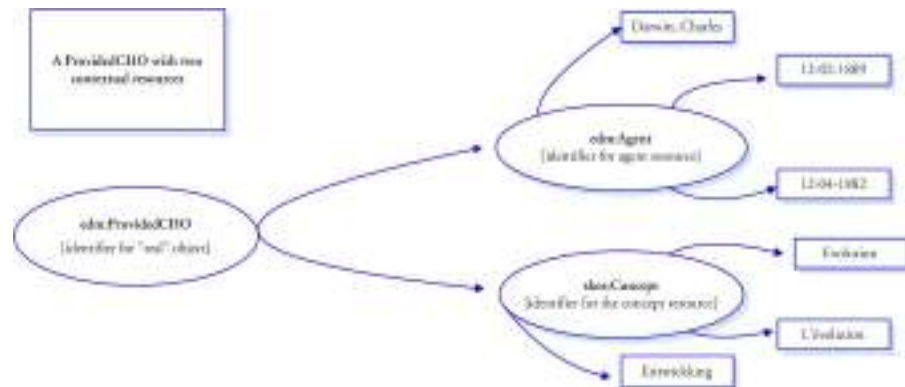
dc:title
 dc:creator
 dc:subject
 dc:description
 dc:publisher
 ...

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

- EDM - Europeana Data Model



Basic data model



Two contextual classes

Metadata mapping, conversion and ingestion

Metadata conversion

DELVING

tool used to convert entries into ESE

The screenshot displays the Delving S3P Creator application. The top window, titled "Local Data Sets", lists several data sets: "bg02test6", "csp047_01", "en02", and "en02_fix". Each entry includes a status message, such as "Data Set is available the local file store." or "There is a mapping created for ese." To the right of this window is a "Local Actions" panel with buttons for "Download Data Set", "Analyze Fields & Edit Facts", "Edit 'ese' Mapping", "Gather Record Statistics", and "Delete Local Data Set".

The bottom window, titled "Mapping 'ese' of 'en02_fix'", shows a "Template" configuration. It has three tabs: "Mapping", "Refinement", and "Normalization". The "Mapping" tab is active, showing a "Parent Record" with fields like "Record Number 1", "input.fdl_audio_record", "input.audio_example", "input.No", "input.PRIMARY_CLASSIFICATION", "input.SECONDARY_CLASSIFICATION", "input.THEME", "input.SOCIAL_FUNCTION", and "input.FORM". The "Groovy Code" section contains the following code:

```
input RECORDED_BY_THE_FOLKLOLIST * {
  dc.contributor $
}
input DATE_OF_RECORD * {
  dc.date $
}
input RECORDING_MEDIA * {
  dc.format $
}
input ARCHIVE_ITEM_NUMBER * {
  dc.identifier $
}
dc.language 'en'
input THEME * {
  dc.subject $
}
input TITLE_LYRICS % /* {
  dc.title $
}
input PRIMARY_CLASSIFICATION * {
  dc.type $
}
```

The "Output Record" section shows the resulting XML output, including fields like "dc.contributor", "dc.date", "dc.format", "dc.identifier", "dc.language", "dc.subject", "dc.title", "dc.type", "europeana:isShownAt", "europeana:language", "europeana:provider", "europeana:rights", "europeana:type", "europeana:url", "europeana:userTag", and "europeana:userTag-Tip".

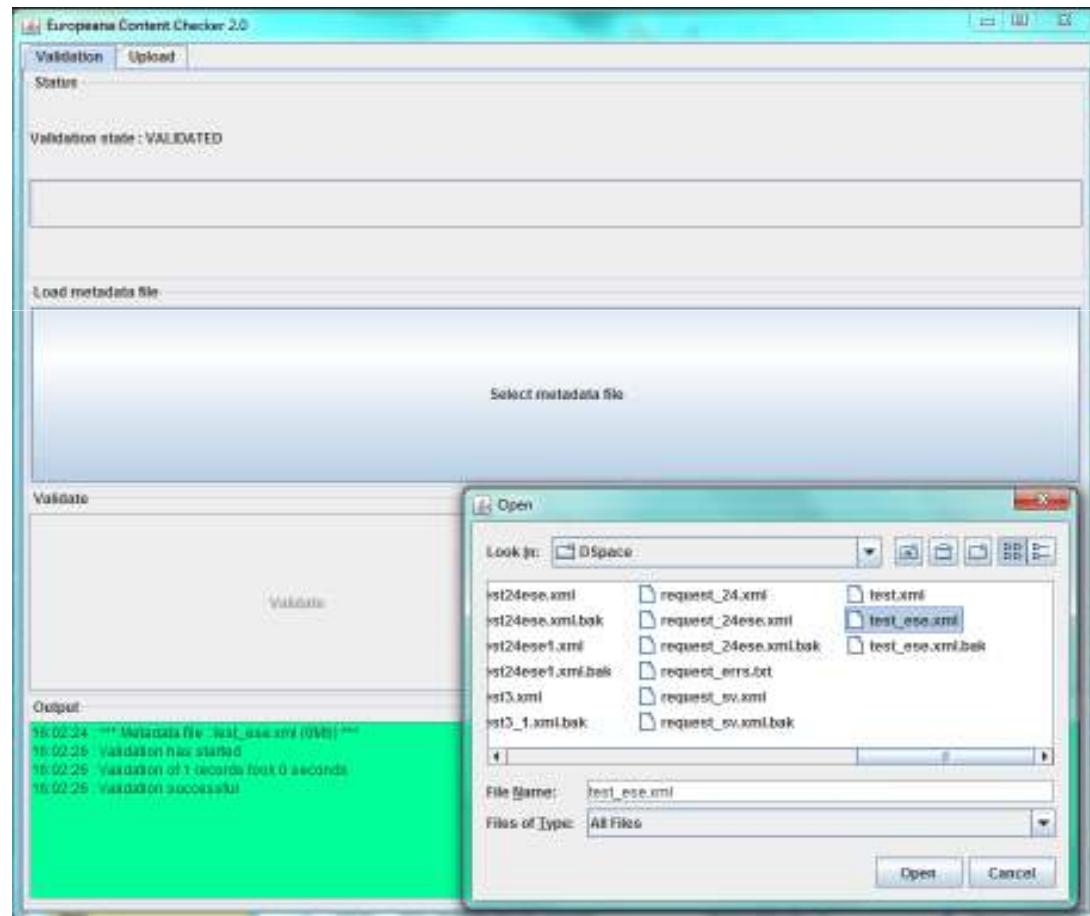
At the bottom of the application, a status bar indicates "Completed at 16:11:43 on 2011-11-02 with 1 normalized, and 0 discarded". There are also checkboxes for "Discard Invalid Records" and "Store Normalized XML", and a "Normalize" button.

Europeana ingestion – OAI-PMH

```
1 0. declara metadata format ese = http://www.europeana.eu/schemas/ese/
2 1. metadataPrefix="europeana" : should be "ese"
3 2. <request>http://192.168.130.213:8080/oai/request : local IP is wrong! reduce to path, eg /oai/request
4 3. If possible, indent the response properly
5 4. <identifier>oai:localhost:pub/13 : must not be localhost!
6 5. <metadata> : parasitic, remove it
7 6. <europeana> : should be <europeana:record>
8 7. <europeana:object> : remove!
9 8. <europeana:country> : I think remove. Check Metadata mapping and normalisation Guidelines v2.1.pdf
10 9. <europeana:language> : I think should be 3-letter
11 10. <europeana:rights> : must be a pure URL!
12 11. Check charset. It seems to be Unicode (UTF-8) but some letters are strange
13     ясь ке одамъ земл' Дуканиск';
14     Първи летли дури (15) пролѣти,
15
16
17 <<xml version="1.0" encoding="UTF-8" >>
18 <OAI-PMH xmlns="http://www.openarchives.org/OAI/2.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="
19 <responseDate>2012-01-17T07:34:08Z</responseDate>
20 <request verb="ListRecords" metadataPrefix="europeana">http://192.168.130.213:8080/oai/request</request>
21 <ListRecords>
22 <record>
23 <header>
24 <identifier>oai:localhost:pub/13</identifier>
25 <timestamp>2012-01-12T10:14:40Z</timestamp>
26 <setSpec>hdl_pub_2</setSpec>
27 </header>
28 <metadata>
29 <europeana xmlns:europeana="http://www.europeana.eu/schemas/ese/" xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:xs:
30 <dc:spatial>Вардарска Македонија</dc:spatial>
31 <dc:language>bg_BG</dc:language>
32 <dc:subject>Юнашка</dc:subject>
33 <dc:type>народна песен, рѣкомис</dc:type>
34 <dc:title>МАРКО И ДЕТЕ ДУКАДИНЧЕ</dc:title>
35 <dc:description>Седнал Марко с майка да вечерат,
36 мад него е дзвезда (1) Вечерница
37 и ја (2) 'ворит (3) Марко Кралевике:
38 'Ај егиди дзвездо Вечерницо!
39 Ти ми стоиш мошне (4) нависоко,
40 тебе нешто я ке (5) те олѣтам,
41 Ти си молам пра'у да ми кажиш:
42 али негде имаш догледамо
43 да се најдрит по-члак от мене?"
44 И му зборвит дзвезда Вечерница:
45 'Ај ти (6) тебе, Марко Кралевике!
46 Ајо сакаш пра'у да ти кажам,
47 да си зе'миш твоя бърза койна (7)
48 та са појдиш земя (8) Дуканискса (9),
```


Metadata ingestion

- administrative steps
- OAI-PMH access provided to Europeana
- ingestion tests



Digital Repository for Cultural Heritage

Digital Repository for Cultural Heritage

- 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...

developed by Sirma Media

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>

Community Building

- Google group
 - <http://groups.google.com/group/cultural-heritage-digitalisation> (35 members)
- Collaboration with IMI and UNIBIT
- Meeting in Sofia 30.01.2012 (75 participants)
- Intense networking as a result
- Broadcast at Bulgarian National Radio
- Working group for the Ministry of Culture
- Upcoming meeting in Veliko Tarnovo 19.03.2012
- About 5 project ideas for the upcoming FP7 and PSP calls

Conclusion

Aggregator for Bulgarian Cultural Heritage to
Europeana

Thank you for your attention!

mariana.damova@ontotext.com