Bulgariana: A Bulgarian Aggregator to Europeana

Mariana Damova

Ontotext, Tsarigradsko Chossee 47A, 1504 Sofia
`mariana.damova@ontotext.com`

Abstract. Europeana is the European virtual museum which was established in 2008. Its ambition is to create a common space allowing to access the cultural heritage of Europe from a single portal, by creating a network in all European countries. To make the initiative technically possible, Europeana has built a technological infrastructure to aggregate metadata from the different Europeana countries and memory institutions, while keeping the original digitized content on their sites. This paper presents Bulgariana, the Bulgarian chapter of Europeana. It is part of the technical infrastructure of Europeana, an established technical aggregator of Bulgarian cultural heritage content. It uses the two Europeana representation models ESE and EDM. Bulgariana is also a community building initiative putting in place a Bulgarian wide network of professionals and institutions working together to preserve and present Bulgarian cultural heritage around the world.

Keywords: Europeana, Bulgariana, Technical Aggregator, ESE, EDM, Cultural Heritage

1   Introduction

Europeana [4] digital museum has currently a collection of more than 20 million objects gathered from different memory institutions throughout Europe. The Bulgarian objects count close to 40K objects. They have been contributed to Europeana through the aggregators provided by Athena, Europeana Local, Carare and Linked Heritage European projects. Bulgariana is a technical aggregator to Europeana, developed and maintained by Ontotext and a community building initiative which aims at improving communication between Bulgarian organizations working in the cultural heritage domain, the improving of technical and organizational collaboration, the concentrating and focusing the efforts of the Bulgarian cultural heritage community, the providing of a default platform for IT research related to cultural heritage, and an opportunity for small museums and other institutions to host their collections.

2   Europeana

Europeana launched in 2008, as a project funded by the European Commission and endorsed by the European Parliament with the goal to make Europe's cultural and

scientific heritage accessible to the public. More precisely, the "Comité des Sages"r ecommended in 2011 to 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. For the five years of its existence it grew a network of over 180 heritage and knowledge organizations and IT experts across Europe. Europeana is based in the National Library of the Netherlands, the Koninklijke Bibliotheek.

Fig. 1. Europeana logo

The collected by Europeana objects fall into 4 categories, e.g. Image, Sound, Video, Text. Its technical architecture (cf. Fig. 2) ensures that the digitized cultural heritage objects remain on the servers of the data providers who ingest according to specific rules only a predefined set of metadata into a Solr database, which in turn is visible through the Europeana portal. Thus, when querying for an object the visitor obtains the standardized by Europeana metadata, and access to the original data provider's website.
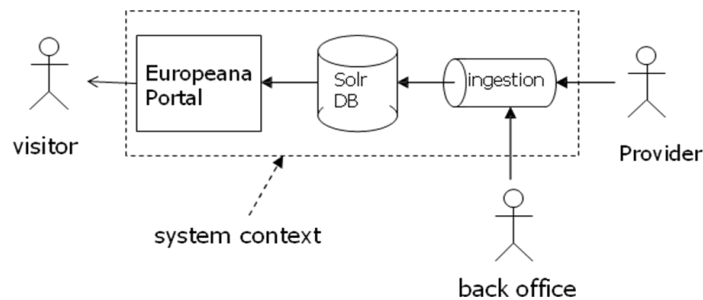

Fig. 2. Europeana architecture

Europeana aggregators are organized by country or by cultural heritage sector. The process of ingesting content is between 4 and 6 weeks.

3   Europeana Data Standards

Europeana maintains two metadata standards, ESE (Europeana semantic elements) [5] and EDM (Europeana data model) [6]. ESE is the first metadata representation model of Europeana. It is a flat model based on Dublin Core [3], and some Europeana specific metadata to an overall of 36 fields. Some of them are shown below.

```
dc:title            europeana:provider
dc:creator          europeana:dataProvider
dc:subject          europeana:rights
dc:description      europeana:type
dc:publisher        europeana:isShownBy
                    europeana:isShownAt
```

ESE model is limited and does not provide the ability to express complex semantic relations. EDM on the other hand is a semantic model, which presents a standard to derive facts in RDF [11]. It allows for building complex provenance, aggregation and authorship relations, and brings the ability for linking the Europeana collections with data from other RDF resources in the Linked Open Data cloud [7].

4   Bulgariana

Bulgariana (`http://bulgariana.eu`), (cf. Fig. 3) is a Bulgarian technical aggregator to Europeana set up by Ontotext. It has four components:
   a)   A digital repository that holds digitized cultural heritage objects, and gives the opportunity to small memory institutions to keep their collections;
   b)   A OAI-PMH [10] server for data harvesting by Europeana, where the metadata are uploaded;
   c)   Metadata conversion methodology, which includes using the Europeana tools Delving [1] and MINT [8];
   d)   Communication channel with Europeana.

Fig. 3. Bulgariana logo

The digital repository (`http://bulgarianheritage.bulariana.eu`, cf. Fig. 4) is based on DSpace [2], and provides with the following features, e.g. elabora-

tion 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, and publication of the digitalized collections on the web – UI layer enabling rich visualization (cf. Fig. 5), various search options, browse by thematic categories, etc.

Fig. 4. Bulgariana digital repository front page

Fig. 5. Presentation of a cultural heritage object in Bulgariana digital repository

The OAI-PMH server ensures the technical connection between Bulgariana and Europeana, and allows for monthly data harvesting of new content or of old content updates.

The metadata converstion methodology consists in a set of rules that help experts to provide ESE or EDM representations of the data described by cultural heritage subject matter experts. Additionally, to make the conversion efficient Delving conversion tool to ESE is being used. It allows to convert big amounts of data provided their original form is given in XML [12], and the metadata conversion schema for these data is inserted into the tool.

The communication channel with Europeana is set up via the OAI-PMH server and several administrative steps, which include data ingestion tests, made with the content ingestion text tool.

5   Bulgariana Collections

Currently Buglariana has two collections:

- Golden Pages from the Bulgarian Renaissance[1]
  This collection contains unique manuscripts of Bulgarian folk songs collected in the 19th century by Miladinov Brothers, renowned Bulgarian Folklorists. These songs have been studied and published in 2008 by the Bulgarian Academy of Sciences [9]. The metadata of this collection include images of the manuscripts (cf. Fig. 6) and transcriptions of the texts, as shown below.
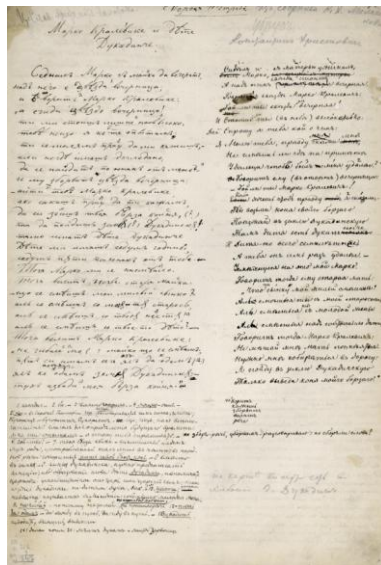

Fig. 6. Image of a folk song manuscript from Brothers Miladinov collection

---

[1]   http://bulgarianheritage.bulgariana.eu/jspui/handle/pub/1

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 мајќа:
"Ај ти, Марко, ај ти, синко милиј;
не си болен, синко, от господа,
тук си болен, синко, от гре'о'и,
да ти викна попоj (2), ду'овници,
лепо да се синко исповедиш,
да си кажиш твоите гре'о'и!"
….

- Pra-historic and Thracian Civilizations[2]

This collection contains unpublished Thracian archeological objects collected by Prof. Valeria Fol, Center of Thracology at the Institute for Balkan Studies at the Bulgarian Academy of Sciences (cf. Fig. 7). It is organized in several thematic sections, e.g. armery, towns, thombs, megaliths, necropoles, objects, sanctuaries, settlements, treasures. The metadata include the locations where the objects have been found and information about the memory institutions where they have been currently kept with their inventory numbers. Each object of the collection is supplied with an original description written by an expert scientist.

Fig. 7. Images of objects from the Pra-historic and Thracian Civilizations collection

These two collections went through the Europeana data ingestion procedure, and are now available for querying at `http://www.europeana.eu` under the collection names: europeana_collectionName: 20215*, or more precisely 2021501* and 2021502*.

6   Bulgariana Collections in EDM

In order to publish Bulgariana collections in Europeana, the metadata of the two collections were represented in ESE. They were consequently converted into EDM, linked to locations from the LOD cloud [7] and loaded into OWLIM triples store. The conversion of the 512 objects produced more than 18K triples. The converted model

---

[2]   `http://bulgarianheritage.bulgariana.eu/jspui/handle/pub/3`

presents the aggregations of the objects from the original Bulgariana site to Europeana. The screenshot in Fig. 8 show the results of the query about the source webpage (column 2) and the title (column 5) of the objects. Columns 1, 3 and 4 showcase the aggregations of the objects that are required by EDM, involving europeana and bulgariana based URIs and proxy's.

Fig. 8. Results of a SPARQL query over Bulgariana collections represented in EDM

Bulgariana collections in EDM are accessible via a SPARQL endpoint at `http://britishmuseum.ontotext.com` (onto/winner).

7 Conclusion and Future Work

This paper presented Bulgariana, a Bulgariana Aggregator to Europeana. The methodology of its creation was outlined along with its first collections, and data models. Along with the technical and technological side, community building activities of networking have been carried out, e.g. a Google group[3], two conferences held in Sofia and in Veliko Tyrnovo, and a working group at the Bulgarian Ministry of Culture has been established. Bulgariana will be extended with new collections of Bulgarian cultural heritage. The process of metadata modeling and data conversion will be automized, Bulgariana will be one of the Bulgarian national aggregators to Europeana, providing the IT infrastructure for preserving and presenting cultural heritage worldwide.

---

[3] `http://groups.google.com/group/cultural-heritage-digitalisation`

support of Europeana team for setting up the aggregator and the work of Lubomir Tsvetkov from Sirma Media, the contribution of Vladimir Alexiev and Svetoslav Petrov from Ontotext, the team of the Institute for Balcan Studies with a Center of Thracology, BAS, the team of the Institute for Literature, BAS, for the realization of Bulgariana and its first two collections.

References

1. Delving – `http://www.delving.eu/`.
2. DSpace – `http://www.dspace.org/`.
3. Dublin Core – `http://dublincore.org/`.
4. Europeana – `http://www.europeana.eu`.
5. Europeana Semantic Elements – `http://www.europeana.eu/schemas/ese/`.
6. Europeana Data Model – `http://pro.europeana.eu/edm-documentation`.
7. Linked Open Data – `http://linkeddata.org/`.
8. MINT – `http://mint.image.ece.ntua.gr/redmine/projects/mint/wiki/Introduction`.
9. Minkova, L., Antonova-Vassileva Luchia: Manuscript of Konstantin Miladinov in Ismail Spezenski's Archive, Bulgarian Academy of Sciences, Institute for Literature, Sofia, (2008)
10. OAI-PMH – `http://www.openarchives.org/pmh/`.
11. RDF – `http://www.w3.org/TR/rdf-concepts/`.
12. XML – `http://www.w3.org/XML/`.