

Automated Metadata Harvesting Among Greek Repositories in the Framework of EuropeanaLocal: Dealing with Interoperability

Alexandros Koulouris^{1,4} Emmanouel Garoufallou^{2,4} and Evangelos Banos^{3,4}

¹Department of Librarianship and Information Systems, Technological Educational Institute (T.E.I.) of Athens, Agiou Spyridonos Str., 12210 Egaleo, Athens, Greece, koulouris.a@gmail.com, akoul@teiath.gr

²Department of Library Science and Information Systems, Technological Educational Institution (T.E.I.) of Thessaloniki, P.O. BOX 141, 57400 Thessaloniki, Greece, garoufallou@yahoo.co.uk, mgarou@libd.teithe.gr

³Evangelos Banos, Openarchives.gr, vbanos@gmail.com

⁴Veria Central Public Library, 8, Ellis Str., P.O. BOX 236, 59100 Veria, Greece

Abstract: EuropeanaLocal is a best practice network project, which will help Europeana to enhance its content and service by applying automated metadata harvesting among distributed repositories. Greek content providers and its metadata aggregator, the Veria Central Public Library (VCPL), are in a testing metadata harvesting period, in the framework of EuropeanaLocal. This paper analyzes the practices that the Greek Institutional Repositories follow in order to transform their metadata schemas to European Semantic Elements (ESE) profile and to export this profile through OAI-PMH to the VCPL aggregator. In addition, it describes the way in which the VCPL aggregates the ESE exported metadata output. Finally, it emphasizes on the transformation and aggregation tools that have been launched on a local level, before Europeana disseminates its official ones.

Keywords: Europeana, EuropeanaLocal, Public Libraries, Cultural Content, Digital Projects, Interoperability, Repositories, Harvesting, Metadata, European Semantic Elements

1. Introduction

Europeana is an evolving service, which will constitute an umbrella of European metadata from distributed cultural organizations. The European Digital Library “Europeana” is enhanced in this effort by various European projects. One of the most representatives is EuropeanaLocal, which will feed Europeana with metadata of local cultural content. Europeana will collect these metadata in an automated metadata harvesting procedure. Greece is participating in EuropeanaLocal with six Greek content providers, a meta-search engine of Greek repositories (openarchives.gr) and a metadata aggregator, the Veria Central Public Library (VCPL). These organizations are in a testing metadata harvesting period, in the framework of EuropeanaLocal. The practices that the Greek repositories follow in order to transform their metadata schemas to European Semantic Elements (ESE) profile and to export this profile through OAI-PMH to the VCPL aggregator are analyzed. Emphasis is given on the transformation and aggregation tools that have been launched on a local level, by VCPL. Finally, results from the test harvesting that have been conducted until the time of papers’ writing are presented.

2. Europeana and EuropeanaLocal: brief analysis

Europeana currently gives access to 5.9 million items representing all Member

States including film material, photos, paintings, sounds, maps, manuscripts, books, newspapers and archival papers, rising to a target of 10 million by 2010 (Davies, 2008; Annual Report, 2010). The Europeana service (Koninklijke Bibliotheek, 2009) is designed to increase access to digital content across Europe's cultural organizations (i.e. libraries, museums, archives and audio/visual archives). This process will bring together and link up heterogeneously sourced content, which is complementary in terms of themes, location and time. By February 2010, Europeana's active partner network consists of 180 organizations.

In order to achieve these goals European Union launched in June 2008 the EuropeanaLocal project in the framework of eContentPlus program. Up to 2011, the EuropeanaLocal partners aim to make available to Europeana more than 20 million items, held across 27 countries. At the same time, they are committed to exploring and developing efficient and sustainable processes and governance procedures so that the growing numbers of regional and local institutions can easily make their content available to Europeana into the future by adopting and promoting the use of its infrastructure, tools and standards (McHenry, 2008).

Key challenges, like metadata, system and semantic interoperability, the harvesting protocol compatibility, the data format diversity, arise. EuropeanaLocal will ensure that the approaches, standards and tools developed by Europeana are widely adopted, thereby supporting the interoperability of content. In brief, this will involve the establishment of a harvestable network of OAI-PMH compliant metadata repositories, which will aggregate local content and which will complement the developing Europeana network. In addition, tools such as Europeana's metadata 'installer' and automated metadata conversion tools, will be made available for EuropeanaLocal partners to use (McHenry, 2008).

The cultural heritage management prototype that Europeana and EuropeanaLocal will implement is the project's innovation. This distributed and automated metadata harvesting model based on OAI-PMH, is represented in figure 1, which describes the Europeana services structure.

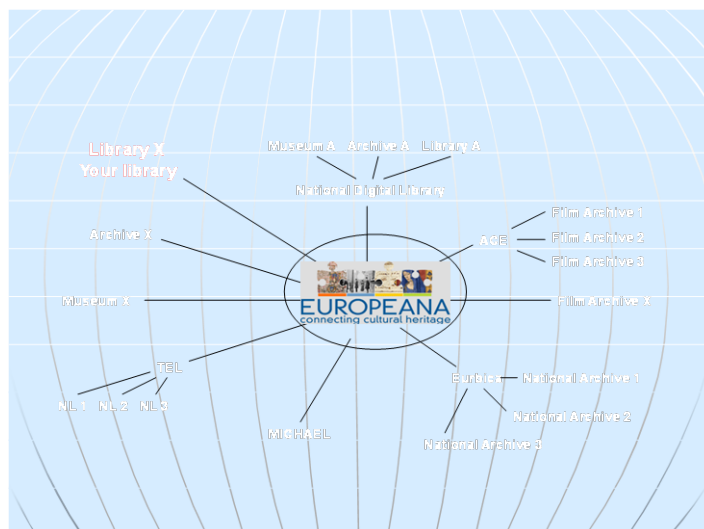


Figure 1. Possible Europeana structure.

3. EuropeanaLocal in Greece: current status and results

Until March 2010, seven Greek cultural organizations, participate in EuropeanaLocal and they will made their metadata available through Europeana service. These organizations are:

- Veria Central Public Library (VCPL) (aggregator)
- American Farm School (AFS)
- Music Library “Lilian Voudouri”
- National Documentation Centre of Greece (NDC)
- National Technical University of Athens (NTUA)
- openarchives.gr (Greek digital repositories meta-search engine)
- Technical Chamber of Greece – Regional Department of Corfu

Among these repositories, VCPL is the national aggregator in the framework of EuropeanaLocal until the finalization of the project, in June 2011. Openarchives.gr is an established meta-search engine of Greek digital repositories which provides consulting and technical support whereas the other organizations are content providers.

In order to analyze the “landscape” in Greek cultural organizations, a metadata, content and technical survey was conducted. This survey was part of the main one that was conducted through EuropeanaLocal partners, from November 2008 till January 2009. The results of the survey are accessible, only for partners, via EuropeanaLocal’s website.

The main conclusion derived from the survey, is that some Greek organization that was at the initial plan for participating at the first phase of the project, will not be able to participate at this stage. Some of the reasons are:

- They are hesitant to provide their digital content widely available though the World Wide Web
- The project’s innovation and the lack of previous experience
- The lack of interoperability standards (e.g. OAI-PMH) that has been adopted by Europeana
- The low level of digitization
- Non-established repositories

In addition, existing fears on intellectual property rights prevent local repositories in participating in EuropeanaLocal. However, this problem has been solved, because the content does not migrate in Europeana, and its provider (repository) is responsible for the content and controls the access.

However, from the organizations that already participate in the project useful remarks were extracted. Most of them use open-source software to provide their digital content. The majority uses DSpace, which is interoperable in metadata schemes; Dublin Core is usually the scheme that the DSpace implements. They strictly follow Dublin Core v1.1 qualified or their application profiles are compatible with this version. In addition, OAI-PMH has been already implemented, and in cases where is not in use, it is already been embedded into the system. That means that it can be implemented any time. Finally, the content varies in formats, from PDF files to AVI. However, all types of forms are compatible with Europeana data format standard (Koulouris and Garoufallou, 2009).

The EuropeanaLocal (2008) helps the Greek cultural organizations to gain new knowledge, to improve their content, metadata, interoperability, and to provide

new services. These are some of the reasons that the EuropeanaLocal Greek team encourages the institutions to participate in the future phases of the project (Trohopoulos, Koulouris, Garoufallou and Siatiri, 2009).

4. Harvesting procedures among Greek repositories

Greek content providers and its metadata aggregator, the Veria Central Public Library (VCPL), are in a testing metadata harvesting period, in the framework of EuropeanaLocal. This section analyzes the practices that the Greek Institutional Repositories follow in order to transform their metadata schemas to European Semantic Elements profile (2009) – that has been developed by Europeana – and to export this profile through OAI-PMH to the VCPL aggregator. In addition, it describes the way in which the VCPL aggregates the ESE exported metadata output. Finally, it emphasizes on the transformation and aggregation tools that have been launched on a local level, before Europeana disseminates its official ones.

It is worth noted that in this process only the metadata are harvested and copyright agreements, firstly between the local content providers and the national aggregator for each country, and secondly between each national aggregators and Europeana should be signed. For example, in Greece, and for the duration of the project EuropeanaLocal (June 2008 – June 2011), each content provider will sign bilateral agreement with VCPL. Then, VCPL will sign an agreement with Europeana. As a result, each Greek content provider has the obligation to export the ESE metadata using OAI-PMH to VCPL, and on the other hand VCPL has the obligation to collect and harvest automatically the ESE metadata from the content providers and to disseminate them to Europeana. Specifically, Europeana will apply automated harvesting to VCPL and to all national aggregators. More on the harvesting procedure and the progress on ESE transformation and aggregation that have been implemented by the Greek EuropeanaLocal partners and the support team are analyzed below.

4.1 Europeana Semantic Elements Compliance

One of the most important aspects in the process of creating a Europeana Compliant digital repository is the support for ESE, which is virtually a new Dublin Core Profile, developed by Europeana in order to fulfill its operational requirements. Existing digital repository software in general does not support ESE by default as it is the case with Dublin Core. Nevertheless, the nature of the formats makes it feasible to alter existing software and data in order to add support for ESE. Specific information about the process can be found at the DSpace plugin for Europeana Semantic Elements webpage (Banos, 2010).

The first step in the process is to use the Europeana XML Namespace <http://europeana.eu/schemas/ese/> and augment existing systems' configuration in order to support the additional ESE elements. Figures 2 and 3 illustrate a sample metadata set for a specific record from the VCPL digital repository using Dublin Core and ESE formats.

```

<record>
<header>
<identifier>oai:medusa.libver.gr:123/903</identifier>
<timestamp>2010-02-07T15:35:54Z</timestamp>
<setSpec>hdl_123_902</setSpec>
</header>
<metadata>
<oai_dc:dc xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
<dc:contributor>Καλλιβωκάς, Αντώνης Δ.</dc:contributor>
<dc:contributor>Ποταμιάνος, Δημήτριος</dc:contributor>
<dc:date>2009-10-14T20:26:31Z</dc:date>
<dc:date>2009-10-14T20:26:31Z</dc:date>
<dc:date>1899</dc:date>
<dc:date>1899</dc:date>
<dc:identifier>http://medusa.libver.gr/handle/123/903</dc:identifier>
<dc:format>784</dc:format>
<dc:publisher>
Εν Αθήναις, Εκ του βιβλιοεκδοτικού καταστήματος Αναστασίου Δ. Φέξη, 1899
</dc:publisher>
<dc:subject>Ιατροδικαστική</dc:subject>
<dc:title>
Ιατροδικαστική : μεθ' ερμηνείας των σχετικών νόμων, διατάξεων κτλ. : προς χρήση των ιατρών και νομικών/ Α.Καλλιβωκάς, Δ
</dc:title>
<dc:type>Βιβλίο</dc:type>
</oai_dc:dc>
</metadata>
</record>

```

Figure 2. Sample Dublin Core Record

```

<record>
<header>
<identifier>oai:medusa.libver.gr:123/903</identifier>
<timestamp>2010-02-07T15:35:54Z</timestamp>
<setSpec>hdl_123_902</setSpec>
</header>
<metadata>
<europeana:record>
<dc:contributor>Καλλιβωκάς, Αντώνης Δ.</dc:contributor>
<dc:contributor>Ποταμιάνος, Δημήτριος</dc:contributor>
<dc:date>2009-10-14T20:26:31Z</dc:date>
<dc:date>2009-10-14T20:26:31Z</dc:date>
<dc:date>1899</dc:date>
<dc:date>1899</dc:date>
<dc:identifier>http://medusa.libver.gr/handle/123/903</dc:identifier>
<dc:format>784</dc:format>
<dc:publisher>
Εν Αθήναις, Εκ του βιβλιοεκδοτικού καταστήματος Αναστασίου Δ. Φέξη, 1899
</dc:publisher>
<dc:subject>Ιατροδικαστική</dc:subject>
<dc:title>
Ιατροδικαστική : μεθ' ερμηνείας των σχετικών νόμων, διατάξεων κτλ. : προς χρήση των ιατρών και νομικών/ Α.Καλλιβωκάς, Δ
</dc:title>
<dc:type>Βιβλίο</dc:type>
<europeana:object>
http://medusa.libver.gr/retrieve/7070/GRVER_00000000000000022.jpg.jpg
</europeana:object>
<europeana:provider>Veria Central Public Library</europeana:provider>
<europeana:type>TEXT</europeana:type>
<europeana:isShownAt>http://medusa.libver.gr/handle/123/903</europeana:isShownAt>
</europeana:record>
</metadata>
</record>

```

Figure 3. Sample Europeana Semantic Elements Record

4.2. Batch import of the ESE elements

After implementing ESE support, the repository has to be populated with the appropriate metadata values. This task can be either performed manually through the appropriate interface of each digital library or automatically using special software tools developed for this purpose. It must be noted that due to the wide usage of the DSpace software among almost all digital repositories, the focus has been the implementation of tools for this specific platform. National Documentation Center has developed a DSpace plugin written in Java, capable of the addition of the Europeana schema in any DSpace repository and the automated completion of the metadata fields. Thorough instructions have also been published in English and in Greek in order to allow anyone to utilize this software (Banos, 2010). Alternative, a simpler PHP Metadata Updater script has been implemented in order to allow the batch insertion of ESE metadata in existing DC records. (Banos, 2010)

4.3 DSpace OAI-PMH support for ESE

The final step in the process of achieving the ESE compliance of a DSpace

digital repository is the modification of the OAI-PMH interface in order to export not only DC but also ESE metadata. This is achieved through a new DSpace Crosswalks Plugin which implements these features. To illustrate the results of this plugin a simple example is listed below. Currently, the VCPL DSpace installation exports DC metadata through the following URL: http://medusa.libver.gr/oai/request?verb=ListRecords&metadataPrefix=oai_dc Using the DSpace ESE Plugin, it is now possible to retrieve ESE metadata through the following URL: <http://medusa.libver.gr/oai/request?verb=ListRecords&metadataPrefix=ese> Last but not least, source code and instructions for installation are available at the DSpace plugin for Europeana Semantic Elements webpage (Banos, 2010).

4.4 Support for legacy digital libraries

Except from DSpace and other modern digital libraries software, there are also numerous digital libraries built with older or closed source technologies which do not support OAI-PMH or any other form of automatic metadata exchange. In these cases, special techniques should be applied in order to extract metadata through plain HTTP requests.

DEiXTo (or ΔEiXTo) (Donas, 2010) is a powerful web data extraction tool that is based on the W3C Document Object Model (DOM). It allows users to create highly accurate "extraction rules" (wrappers) that describe what pieces of data to scrape from a web page. When used appropriately, DEiXTo can extract meaningful metadata from digital libraries and generate DC and ESE metadata records according to the standards. These records can be utilized by any standards compliant metadata harvester in order to insert the specified repository in Europeana.

DEiXTo is currently used to import Music Library "Lilian Voudouri" in Europeana and gives the possibility to extract metadata from virtually any web site willing to participate in Europeana, regardless of its current technology.

4.5 Aggregation

In order to optimize the process of aggregating the metadata of all European digital repositories, Europeana suggests that aggregators should be implemented on national or regional scale. VCPL has implemented an aggregator (Veria Central Public Library, 2010) which ingests ESE metadata from current Europeana participants in Greece and after performing some rudimentary checks, exports this metadata to Europeana. All new participants willing to participate in Europeana Local from Greece must be checked and ingested from the VCPL Aggregator which in turn propagates their metadata to Europeana (see figure 4).

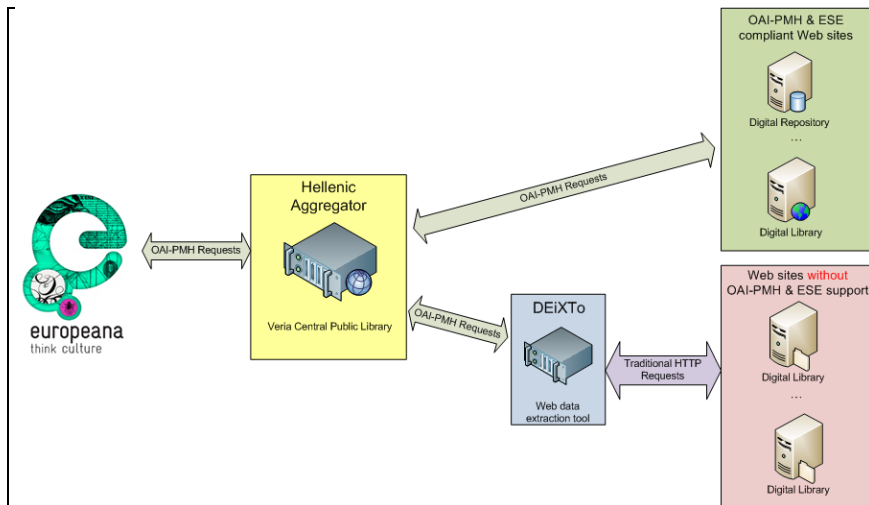


Figure 4. VCPL Aggregator

4.6 Current Status

As of March 2010, all content providers except Music Library “Lilian Voudouri” use the DSpace software for their digital libraries and have followed closely the aforementioned instructions, thus implementing full support for ESE. Sample metadata records from these libraries have been tested successfully with the Europeana Content Checker, a special tool provided by Europeana in order to check metadata compliance with the standards. Additionally, all these libraries have been harvested successfully by the VCPL Aggregator.

5. Conclusions

It is evident that in a couple of years Europeana, the European Digital Library, will be one of its kind worldwide since it will compile a variety of media from across Europe, all house under one umbrella. These collections and cultural heritage resources of immense importance will change the ‘digital library map’ in Europe. It is therefore our obligation as Greek harvesting point to assist, drive and enrich the quality of metadata, build tools that will assist small and medium libraries, archives and museums in order to be part of this effort that the European Community undergoes.

This vision for the future is a common vision for every European country and Greek content providers need to be part of this. We hope that in the following months the Greek section of EuropeanaLocal to be first in the agenda of Greek Library, Museum and Archive community.

Acknowledgement

We thank the Greek EuropeanaLocal team and the content providers for their support that make this paper possible.

References

Annual Report. Grant Agreement Number: 558001 Europeana v1.0. 1 February 2009 – January (2010).

- http://version1.europeana.eu/c/document_library/get_file?uuid=df49b382-aa16-497f-8405-948b532dc0d7&groupId=10602 [accessed: 31/03/2010].
- Banos, E. (2010). DSpace plugin for Europeana Semantic Elements (ESE), <http://vbanos.gr?p=189> [accessed 25/3/2010]
- Davies, R. (2008). Europeana: an Infrastructure for Adding Local Content. *Ariadne*, 57, <http://www.ariadne.ac.uk/issue57/davies/> [accessed: 28/03/2010].
- Donas. K. (2010). DEiXTo, <http://www.deixto.com> [accessed 25/3/2010]
- European Commission (2008). *eContentplus*, http://ec.europa.eu/information_society/activities/econtentplus/index_en.htm [accessed: 15/8/2009].
- Europeana Semantic Elements specifications. Version 3.2.1, 06/11/2009. Europeana v1.0 (2009). http://www.europeanfilmgateway.eu/downloads/Europeana_Semantic_Elements_Specifications_v3_2_1_2009-11-06-1.pdf [accessed: 31/03/2010].
- EuropeanaLocal (2008). *EuropeanaLocal: connecting cultural heritage*, <http://www.europeanlocal.eu> [accessed: 28/03/2010].
- Koninklijke Bibliotheek (2009). *Europeana*, <http://www.europeana.eu> [accessed: 28/03/2010].
- Koulouris, A. and Garoufallou, E. (2009). Managing cultural heritage content in the context of Europeana. *Proceedings of the International Conference of Marketing and Management Sciences (ICMMS 2009)*, Kos Island, Greece, September 11-14, 2009 (to appear).
- Lagoze, C., et al. (2008). *The Open Archives Initiative Protocol for Metadata Harvesting*, <http://www.openarchives.org/OAI/openarchivesprotocol.html> [accessed: 15/8/2009].
- McHenry, O. (2008). EuropeanaLocal – its role in improving access to Europe’s cultural heritage through the European Digital Library. *11th Annual International Conference «EVA 2008 Moscow»*, Moscow. http://conf.cpic.ru/upload/eva2008/reports/dokladEn_1509.pdf [accessed: 28/03/2010].
- Trohopoulos, I., Koulouris, A., Garoufallou, E. and Siatri, R. (2009). Harvesting metadata of Greek institutional repositories in the context of EuropeanaLocal. *Workshop on harvesting metadata: practices and challenges, 13th European Conference on Digital Libraries (ECDL 2009)*, Corfu, Greece, September 29 - October 2, 2009.
- Veria Central Public Library (2010). Europeana Local Aggregator <http://aggregator.libver.gr> [accessed 25/3/2010]