A Systemic Approach for Effective Semantic Access to Cultural Content

Ilianna Kollia∗, Vassilis Tzouvaras, Nasos Drosopoulos and George Stamou
School of Electrical and Computer Engineering, National Technical University of Athens, Zographou Campus 15780, Athens, Greece

Abstract. A large on-going activity for digitization, dissemination and preservation of cultural heritage is taking place in Europe, United States and the world, which involves all types of cultural institutions, i.e., galleries, libraries, museums, archives and all types of cultural content. The development of Europeana, as a single point of access to European Cultural Heritage, has probably been the most important result of the activities in the field till now. Semantic interoperability, linked open data, user involvement and user generated content are key issues in these developments. This paper presents a system that provides content providers and users the ability to map, in an effective way, their own metadata schemas to common domain standards and the Europeana (ESE, EDM) data models. The system is currently largely used by many European research projects and the Europeana. Based on these mappings, semantic query answering techniques are proposed as a means for effective access to digital cultural heritage, providing users with content enrichment, linking of data based on their involvement and facilitating content search and retrieval. An experimental study is presented, involving content from national content aggregators, as well as thematic content aggregators and the Europeana, which illustrates the proposed system capabilities.

Keywords: cultural heritage access, metadata schema mapping, European data model, Europeana, linked open data, semantic query answering, query rewriting, cultural resource discovery and enrichment

1. Introduction

Digital evolution of the Cultural Heritage Field has grown rapidly in the last few years. Following the early developments at European level and the Lund principles[15], massive digitisation and annotation activities have been taking place all over Europe and the United States. The strong involvement of companies, like Google, and the positive reaction of the European Union have led to a variety of, rather converging, actions towards multimodal and multimedia cultural content generation from all possible sources, such as galleries, libraries, archives, museums and audiovisual archives. The creation and evolution of Europeana, as a unique point of access to European Cultural Heritage, has been one of the major achievements in this procedure. More than 15 million objects, expressing the European cultural richness, are currently accessible through the Europeana portal, with the target pointing to double this number within the next five years.

As a consequence of the above, research in digital cultural heritage (DCH) is rapidly becoming data intensive, in common with the broader humanities, social science, life and physical sciences. Despite the creation of large bodies of digital material through mass digitisation programmes, only a small proportion of all cultural heritage material has been digitised to date, and there is significant commitment to further digitisation at national and institutional levels across Europe [24]. An estimate of the vast amount of data (around 77 million books, 358 million photographs, 24 million hours of audiovisual material, 75 million of works of art, 10,5 billion pages of archives) still to be digitized and the related cost (about 100 billion euro) is provided in the recent European Report of the Comite’ des Sages [19]. Further, substantial amounts of born-digital material are related with cultural heritage, such as data produced by scientific research and by digital

∗Corresponding author. E-mail: ilianna2@mail.ntua.gr

0000-0000/11/$00.00 © 2011 – IOS Press and the authors. All rights reserved
In the 2008 European Semantic Web Conference, a specific emphasis was put on achieving a certain level of common understanding in a machine-understandable manner. This means that the ability to automatically process the information in a machine-understandable manner is key. Therefore, systems that understand these semantics, such as reasoning tools, ontology querying engines, and more, are needed to provide unified access to the continuum of data. This leads to a unifying model, as is the case with Europeana, for accessing aggregated data from many heterogeneous and diverse data sources. The current paper presents a system including an ingestion mechanism, which provides users and content providers the ability to deliver more complete answers as new data sources appear on the Web. Linked data consist one of the main technologies identified so far for extending Europeana in the following three years [25].

The main approach to interoperability of cultural content metadata has been the usage of well-known standards in the specific museum, archive and library sectors (Dublin Core, Cidoc-CRM, LIDO, EAD, METS) [17,18] and their mapping to a common data model used - at the Europeana level: European Semantic Element (ESE, 2008), European Data Model (EDM, 2010) - to provide unified access to the centrally accessed, distributed all over Europe, cultural content [22]. In this framework, research in cultural heritage has to treat collections of data from many heterogeneous data sources as a continuum, overcoming linguistic, institutional, national and sectoral boundaries [20,26].

On the other hand, the Web has evolved in recent years, from a global information space of linked documents to one where both documents and data are linked. Underpinning this evolution is a set of best practices for publishing and connecting structured data on the Web known as Linked Data [23]. The adoption of the Linked Data best practices has lead to the extension of the Web with a global data space connecting data from diverse. The Web of Data enables new types of applications such as browsers which allow users to start browsing in one data source and then navigate along links into related data sources and, search engines that crawl the Web of Data by following links between data sources and provide expressive query capabilities over aggregated data, similar to how a local database is queried today. The Web of Data also opens up new possibilities for domain-specific applications. Unlike Web 2.0 mashups which work against a fixed set of data sources, Linked Data applications operate on top of an unbound, global data space. This enables them to deliver more complete answers as new data sources appear on the Web. Linked data consist one of the main technologies identified so far for extending Europeana in the following three years [25].

The system workflow presented in this paper includes a semantic enrichment and content linking part, which is proposed to be based on semantic query answering techniques. It is shown that query answering techniques can be used for assisting users to enrich metadata of their content, taking advantage of relevant sources, data and knowledge stores, or to link their data to relevant ones provided by other sources; thus contributing in the generation of a linked data knowledge substrate that can be further used to facilitate and make feasible the effective access to large amounts of cultural content.

The paper is organised as follows: Section 2 describes the workflow and architecture of the proposed system. Specific attention is given to user assisted mapping of data models to the new EDM Europeana data model and to inclusion of the metadata enrichment module. Section 3 presents the query answering method, describing the different possible approaches based on the targeted query and ontology properties. Targeted resource discoveries in the cultural heritage section, as well as query implementation issues are discussed in Section 4. An experimental study is pre-
sent in Section 5 which illustrates the usage of the proposed system, based on experiments with content provided by the Hellenic Ministry of Culture and Tourism aggregation portal\(^1\), with potential extensions to a large number of object metadata provided by project Athena to Europeana. Conclusions and further work are given in Section 6 of the paper.

2. The Semantic Mapping System

Many different metadata schemas or in a broader sense, sets of elements of information about resources, are being used in the areas of Digital Libraries and eLearning, across a variety of technical environments and scientific disciplines. Europeana is being developed to provide integrated access to digital objects from cultural heritage organisations, encompassing material from museums, libraries, archives and audio-visual archives as the single, direct and multilingual gateway to Europe’s cultural heritage. Several cross-domain, vertical or thematic Aggregators are being deployed at regional, national and international level in order to reinforce this initiative by collecting and converting metadata about existing and newly digitised resources.

The currently employed Europeana Semantic Elements V3.3 (ESE) is an updated version of the metadata set used in the Europeana prototype in November 2008 that has been amended to include additional elements for the Rhine release of the portal in July 2010. It is a Dublin Core-based application profile providing a generic set of terms that can be applied to heterogeneous materials thereby providing a baseline to allow contributors to take advantage of their existing rich descriptions. The later constitute a knowledge base that is constantly growing and evolving, both by newly introduced annotations and digitisation initiatives, as well as through the increased efforts and successful outcomes of the aggregators and the content providing organisations.

The new Europeana Data Model is introduced as a data structure aiming to enable the linking of data and to connect and enrich the descriptions in accordance with the Semantic Web developments. Its scope and main strength is the adoption of an open, cross-domain framework in order to accommodate the growing number of rich, community-oriented standards such as LIDO for museums, EAD for archives or METS for libraries. In the same time, EDM has to serve as the means of interoperability between Europeana and a large number of diverse systems and collections of structured and unstructured data, addressing the growing needs of Europe’s digital cultural heritage. Apart from its ability to support standards of high richness, EDM also enables source aggregation and data enrichment from a range of third party sources while clearly providing the provenance of all information. One of the crucial purposes of EDM is to answer the basic queries ‘Who?’, ‘What?’, ‘When?’ and ‘Where?’ for every digital object and to make connections between the networks that will animate Europeana’s content.

The described system handles the ingestion, storage and maintenance of cultural content metadata, offering content providing organisations and aggregators the tools for data alignment, transformation, resource linking, enrichment, reconciliation and re-mediatisation. It has been deployed for several standards or specialized models such as LIDO, Dublin Core, ESE, CARARE’s MIDAS-based schema, EUScreen’s EBU-Core based approach and, it is being used for the prototyping of EDM. It allows for the ingestion of semi-structured data and offers the ability to intuitively align and take advantage of a well defined, machine understandable schema. The underlying data representation is XML while the user’s actions are translated into XSL transformations. The common model functions as an anchor for the repository, to which various data providers can be attached and become, partly at least, interoperable. The metadata mapping tool formalizes the notion of ‘crosswalk’ by hiding technical details and permitting the semantic equivalences to emerge as the centrepiece. As a result, content providers, researchers or domain experts, who are typically not programmers, can easily enter the translation logic that can be automatically converted into executable code. It involves a graphical environment where a user attempts to achieve interoperability by mapping one pair of elements at a time. A user import is not required to include the schema used. Simplification of the actual work for the user is done by reducing the set of elements that have to be mapped to only those that are populated and, to avoid inconsistencies between schema declaration and actual usage. The structure that corresponds to a user’s specific import is visualized in the mapping interface as an interactive tree that represents a snapshot of the XML schema that the user is going to use as input for the mapping process. The user

\(^1\)http://collections.culture.gr
I. Kollia et al. / A Systemic Approach for Effective Semantic Access to Cultural Content

is able to navigate the tree and access element statistics for the specific import, e.g. see Figure 1.

The interface then presents an intuitive and visual appealing environment for the user to define mappings, without sacrificing any of the functionality needed to properly achieve the task of schema mapping (e.g., conditional mappings, string manipulation functions). It can be configured to provide the user with groups of high level elements that constitute separate semantic entities of the target schema. These top level elements are presented on the right side of the screen while clicking the corresponding button, the set of the sub-elements that are part of that group are presented in the middle part as can be seen in Figure 1. This part of the user interface has a tree structure of embedded boxes that represents the internal structure of the complex element. The user is able to interact with this structure by clicking to collapse and expand every embedded box that represents an element along with all relevant information (attributes, annotations) defined in the XML schema. To perform an actual mapping between the input and the target schema, a user has to simply drag a source element and drop it on the respective target in the middle.

The user interface of the mapping tool is schema aware regarding the target data model. That means that certain operations are enabled or restricted accordingly based on constraints for elements in the target XML schema. For example, if an element can be repeated then an appropriate button appears to indicate and implement its duplication. User’s mapping actions are expressed through XSLT stylesheets, i.e. a well-formed XML document conforming to the namespaces in XML recommendation. XSLT stylesheets are stored and can be applied to any user data, can be exported and published as a well-defined, machine understandable crosswalk and, shared with other users to act as template for their mapping needs.

One of the main points that have guided the system’s development is the apparent need for the preservation and alignment of as much of the original data richness as possible. That is also reinforced by the parallel, ongoing provider efforts to increase the expressivity of their systems and, subsequently, that of the submitted data structures. The aggregation is only the first effort on the part of providers and aggregators towards the
efficient mediation and reuse of their knowledge bases. The support for semantic data models such as EDM enables the repository for deployment and, most importantly, information reuse through knowledge modelling and data interoperability research activities. The aim is to support further resource linking between different collections, reconciliation across the repository and with external authorities and, enrichment of the information resources.

The work-flow for the creation of EDM RDF consists of the following phases shown in Figure 2, each being responsible for specific services that ensure the quality of the ingestion process. Harvesting/delivery refers to the collection of the metadata from content providers through common data delivery protocols, such as OAI-PMH, HTTP and FTP.

Schema Mapping aligns harvested metadata to a common reference model. It consists of a graphical user interface that assists content providers in mapping their metadata structures and instances to a rich, well defined schema, using an underlying machine-understandable mapping language. It supports the sharing and reuse of metadata crosswalks and the establishment of template transformations. Value Mapping focuses on the alignment and transformation of a provider’s list of terms to the authority file or external source introduced by the reference model. A more common task is data normalisation such as in the case of dates, geographical locations or coordinates, country and language information or name writing conventions. Revision/Annotation enables the addition of annotations, editing single or group of items in order to assign metadata not available in the original context, or to use additional terminologies. The output of this process is the EDM instances as also illustrated in the EDM RDF preview in Figure 3.

3. Semantic Cultural Content Access

Content metadata can be terminologically described in order to be more easily interconnected and used in conjunction with other, useful, possibly complementary content and information, independently published in the web. The result is to provide a semantically integrated cultural domain knowledge that can be used for effective access to cultural content. This is achieved by semantically characterizing content through mappings to ontological domain knowledge. Thus, one of the main goals of semantic integration is to develop formal methods for the description and representation of such mappings, supporting the construction of an interlinked information space for the cultural content.

Nevertheless, semantic characterisation of cultural content, stored in cultural content metadata, would be of little use, if the end users were not in position to pose their queries in terms of a rich integrated ontological knowledge, instead of in terms of a data storage schema, as currently is the case, which highly limits the scope of a query. Semantic query answering refers to this notion of finding answers to queries posed by users based not only on the explicitly stated data, i.e., the data that are stored in databases, but also on the implicit ones which are found by reasoning on the schema of the ontology.

The problem of answering conjunctive queries in terms of ontologies represented in description logics [2] (the underlying framework of the W3C’s Web Ontology Language -OWL) has been proven to be difficult, suffering from very high worst-case complexity (higher than other standard reasoning problems) that is not relaxed in practice [7]. This is the reason that methods and techniques targeting the development of practical systems mainly follow two distinct directions. The first suggests the reduction of the ontology language expressivity used for the representation of conjunctive queries vocabulary, while the second sacrifices the completeness of the query answering process, providing as much expressivity as it is needed.

The attempts to provide scalable semantic query answering over ontologies expressed in larger fragments of OWL introduced the notion of approximation. Approximate reasoning usually implies unsoundness and/or incompleteness, however in the case of semantic query answering most systems are sound. Typical examples of incomplete query answering systems are the well-known triple stores (Sesame, OWLIM, Virtuoso, AllegroGraph, Mulgara etc).

In this paper we propose the interweaving of the mapping system presented in Section 2 with query answering and illustrate the performance of the resulting approach in real life access to digital culture heritage of the Hellenic Cultural Content Aggregator and of Europeana. In this framework, query answering w.r.t ontology schemas over data coming from heterogeneous sources can be used to suggest similarities between data and enrich them or possibly lead to their linking following validation of some expert.

Two approaches are used for answering queries over cultural data. The one is applicable to ontology languages of low expressivity but provides good scalability performance since it results in the use of databases
with their long studied traditional query answering methods \[1\], while the other is able to use more expressive ontologies but cannot currently handle large amounts of data.

In particular, the former is based on the query rewriting technique \[3,8,12\], while the latter is based on the reduction of queries to standard reasoning tasks of OWL reasoners \[9,10,11,6\].

In the following the terminology of description logics is used. In description logics, an ontology \( O \) is represented as a tuple \( \langle T, A \rangle \) where \( T \) is the terminological component (called TBox) containing in the form of axioms general knowledge about the modeled domain and \( A \) is the assertional component (called ABox) containing concrete instance data with their relationships. A conjunctive query \( q \) is of the form \( q : \)
\( Q(x) \leftarrow \bigwedge_{i=1}^{n} C_i(x, y) \), where \( x, y \) are vectors of variables and \( C_i(x, y) \) are predicates, either concept or role atoms. The variables in \( x \) are called distinguished or answer variables and those in \( y \) are called non distinguished or existentially quantified.

According to the first technique, the variables of the query \( q \) we want to answer are substituted by tuples of individual names appearing in the queried ontology \( O \) forming a boolean query \( q' \) and those tuples that result to the entailment of \( q' \) by \( O \) are kept as the answers for \( q \). To avoid performing \( m^n \) entailment checks (where \( m \) is the number of individuals in the ontology and \( n \) is the number of variables in the query) that would be the result of this process, optimizations can be employed to improve the running time of query answering. Such optimizations for OWL DL are described in [6] in the context of the SPARQL query language. The conjuncts of the query can be evaluated sequentially and variables of subsequent conjuncts are mapped only to individuals that have resulted in the entailment of previous instantiated conjuncts.

**Example 1** Let us assume that we want to evaluate the query

\[
Q(x, y) \leftarrow \text{WorkOfArt}(x) \land \text{madeIn}(x, y) \land \text{Period}(y)
\]

over an ontology \( O \). Let us also assume that the conjunct \( \text{WorkOfArt}(x) \) will be evaluated first and a set \( S_{1x} \) of the individuals for the variable \( x \) that satisfy the conjunct is created. Then the variable \( x \) in the second conjunct, \( \text{madeIn}(x, y) \), will be substituted only by the individuals in the set \( S_{1x} \) and not by all individuals appearing in \( O \). In the same way, a set \( S_{1y} \) containing all individuals for the variable \( y \) that satisfy the first two conjuncts is created which contains individuals that can then be used as possible mappings for the variable \( y \) in the conjunct \( \text{Period}(y) \).

Other optimizations refer to the use of more specialized tasks of OWL reasoners such as instance retrieval to retrieve instances of concepts instead of iterating over all individuals of the knowledge base and checking entailment of the instantiated queries obtained by substituting variables with individuals. The use of such methods greatly reduces the running time of queries.

The second technique splits the problem of query answering into two parts, the reasoning part which expands the initial query taking into account terminological knowledge provided by the ontology and the data retrieval part which retrieves the instances of the expanded query from a database. In particular, during the first step (usually called query rewriting) the conjunctive query is analysed and expanded into a set of conjunctive queries (in case the queried ontology is expressed in the OWL QL profile [4]) or a datalog query (in case the queried ontology is expressed in the OWL EL profile [4]), using all the constraints provided by the ontology [12,3]. Then, the resulting queries are processed with traditional query answering methods on (deductive) databases or triple stores, since terminological knowledge is no longer necessary. The query rewriting step is shown via an example below.

**Example 2** Let us assume that an ontology \( O \) consists of the two axioms:

\[
\text{WorkOfArt} \sqsubseteq \exists \text{madeBy.Artist} \quad (1)
\]

\[
\text{Painting} \sqsubseteq \text{WorkOfArt} \quad (2)
\]

and we ask the query

\[
Q(x) \leftarrow \text{madeBy}(x, y) \land \text{Artist}(y) \quad (3)
\]

The rewriting of query (3) w.r.t. \( O \) consists of (3) and the following queries:

\[
Q(x) \leftarrow \text{WorkOfArt}(x) \quad (4)
\]

\[
Q(x) \leftarrow \text{Painting}(x) \quad (5)
\]

Through the decoupling of the data retrieval step from the query rewriting step, users are able to build complex queries without having to know the underlying structure or technical details of the data sources but using only the terminological knowledge expressed in terms of ontologies.

To decide which technique is more appropriate for a specific application scenario we have to take into account the benefits and limitations of each one of them.

The method that uses traditional reasoning services for query answering is applicable to very expressive fragments of OWL such as OWL DL but suffers from the fact that it cannot currently handle large amounts of data. The second technique handles scalability issues well but suffers from the fact that it cannot work with highly expressive languages such as OWL DL which is useful in many practical application scenarios, since in such cases an infinite set of conjunctive or datalog queries can be created. The query rewriting technique is currently only applicable to queries posed over ontologies written in the OWL QL and OWL EL profile of OWL 2.
As far as the cultural domain is concerned, usage of either technique can be advantageous, depending on the context of the query and of user profile, or interests, or on the use case, i.e., for entertainment or education.

4. Query Answering for Improved resource discovery

4.1. Improved resource discovery

The current state of the art in Cultural Heritage implements a model whereby many aggregators, content providers and projects feed their content into a national, thematic, or European portal, and this portal is then used by the end user to find cultural items. There are ongoing efforts to investigate the potential of a semantic layer, which can improve the user search experience, primarily (at present) through the disambiguation of key search query terms.

Figure 4 gives a view of the usage of semantic technologies within the cultural heritage framework. Query answering based on cultural portal content and on the EDM semantic repository is used for enrichment and linking of data in the Web.

Going one step further leads to the involvement of the user in the creation of linked data, by providing services whereby the user explicitly creates links to relevant third party linked-data resources. This innovative approach is expected to yield higher-quality links than a purely automated approach, because it benefits from the expertise of the end user.

Resource linking is the most important function of cultural aggregation and is performed on different levels and between various ‘entities’ in the repository. This may include various cases, such as:

a) Digital resources that point to the same CH object (e.g. two Mona Lisa metadata records originated from different organizations/repositories). b) The same ‘entity’ appearing in different metadata records (e.g. ‘Da Vinci’ and ‘Leonardo Da Vinci’). c) ‘Entities’ in the metadata records and common authority files and vocabularies (e.g. VIAF for people). d) Different objects and ‘entities’ (i.e. people, places, dates) through the use of properties such as occuredAt, wasPresentAt, happenedAt, hasMet etc.

Users will have the ability to create Digital Stories to illustrate their personal narratives that may also be related to CH objects. The process will involve the use of Europeana CH objects together with text and multimedia (from external sources, or uploaded by users themselves). In their digital stories users essentially represent a set of events that link Europeana CH objects to actors (usually starting from themselves) and locations. While some of those links may be of interest only for the storytelling, others can provide important information in the CH domain. For example, a user can create a digital story that involves his visit to the Louvre to see the Mona Lisa, which may include himself (e.g. foaf profile), the Louvre (e.g. name, respective dbpedia resource, location, multimedia or text about it) and Mona Lisa. While the links between him and the museum or the object are only relevant for personalized or community-based retrieval, potential connections between places and items can provide fundamental information that may be missing from the national, thematic or European portal. They can also provide information about previous events that refer to an object and a location other than its current as it is presented in the metadata (e.g. a user participated in the preservation of a CH object that was later exhibited at its current location). Users are allowed to link CH objects together for any reason (e.g. objects that used to belong to a collection, describing a visit to a temporary exhibition or describing a collection of favorite CH objects from the Parthenon); other users can opt to trust this ‘classification’ in their browsing through the repository.

4.2. Query Answering Implementation

In order to experiment with the use of the two query answering techniques described in Section 3 for the collected cultural data, we have used two systems. The one is Requiem developed at the Oxford University Computing Laboratory, which is a prototypical implementation of the query rewriting algorithm presented in [3]. The other system has also been developed at the Oxford University Computing Laboratory and uses SPARQL as a language to express queries over OWL ontologies and evaluate their answers. SPARQL is an RDF based query language which has currently been extended to find answers to queries under the OWL Direct Semantics Entailment relation [5].

5. Experimental Study and Evaluation of the Proposed System

Application of the system proposed in this paper has been and is currently taking place in the framework

\[http://www.comlab.ox.ac.uk/projects/requiem/home.html\]
system in the above frameworks. For this reason, it focuses on a sample of content provided to Europeana mainly by the Hellenic Ministry of Culture and Tourism (HMCT) consisting of about 4200 objects (http://collections.culture.gr), as well as on a sample of about 3600000 objects, provided by the ATHENA EU project to Europeana. In the following, we focus on the Hellenic content, mainly from the museum sector, aggregated to Europeana through the HMCT. The input metadata of this content (as well as that of the ATHENA data set) that was used in the experimental study is in LIDO (Lightweight Information Describing objects, http://www.lido-schema.org) XML format. Using the ingestion tool and the workflow described in Section 2, the LIDO XML records were uploaded in the tool and transformed in EDM OWL/RDF. Figure 5 illustrates the OWL/RDF output of an example record. Similarly, all input records have been transformed in EDM and stored in a Sesame triple store. An amount of about 100000 RDF triples has been generated and used next for query answering and metadata enrichment.

Each object that a museum contributes to the aggregator is described among other by an identifier, a type, a description, the material it is made of, the museum where it can be found, the date it was created. In particular, this cultural content has been aggregated through 34 archaeological museums all over Greece and is classified in 55 categories (such as pottery, jewelry, stamps, wall paintings, engravings, coins) and more than 300 types, within 17 time periods from 35000 b.c.
up to today.

This data is provided by the ingestion tool as data property assertion axioms. In order to be able to perform reasoning over the data, classes and properties were extracted from this data that are used for describing the objects of the collection. In particular, classes were created for the categories of objects that mentioned above and for their types. For example, 18 classes (such as finger jewelry, head jewelry, neck jewelry, cross, zone) were created as subclasses of the jewelry class. As far as the period that each museum object was created is concerned, classes for every century were created that were grouped to form periods that are of particular interest.

We then applied the techniques described in Section 3 for semantic query answering on the above data set. Figure 11 includes a list of queries (Column 1) that can be asked by users, such as researchers, archaeologists, students, in the framework of specific uses and search scenarios (Column 2) and can be answered by the system based on the locations (Column 3) of the objects.

In the first experiment we applied both query answering approaches described in Section 3 to answer the query 'Jewelry of the Late Byzantine period.' Both methods provided 34 results. Some of the related pictures are shown in Figure 6. The query rewriting approach found the answers by expanding the original query in 112 queries. The time required to answer both queries was comparable in this case. Reduction of the computational time, when dealing with larger amounts of data is possible, by reducing the number of e.g. categories of the class jewelry that are being searched, based on information acquired about the scope of use or the user profile. If, for example, the search had to do with a study on religion, objects that are known in the knowledge base to be related to religion, such as crosses, should be first searched for, followed, if necessary, by more searches, in a scalable way. In this case, the three objects shown in the last row of Figure 6 would be computed first, since they better fit the specific context of usage.

Let us now consider the subclass 'apparel jewelry' and let us assume that the archaeologist who is annotating the digitized images of some brooches does not know the material that they are made of. In fact, the records of the two brooches of the Archaeological Museum of Galaxidi shown in Figure 7 do not include anything about the material they are made of. The responsible person in the Museum, therefore, who identifies this, sends a query to the system asking for all objects entitled 'brooch'. Answering his query, the system provides him or her with results and related images of copper brooches shown in Figure 8 from the Archaeological Museums of Xios and the Museum of Byzantine Culture and of golden brooches from the National Archaeological Museum and from the Archaeological Museum of Thessaloniki, shown in Figure 9. The responsible person, possibly assisted by the archaeologist, based on similarity of results, or the information of the other fields of the returned results, enriches their own results, or links their data to some of the objects returned by the query, corresponding, e.g., to the copper images. A similar case is shown in Figure 10 where the two images on the left show objects of type 'mosaic' the material of which is not mentioned, while the other very similar image on the right is declared to be composed of the material lazourite.

Let us now consider a more complex query, where a researcher wants to search the whole knowledge base asking for ‘jewelries composed of rare (uncommon) materials’. Let the knowledge base translate rare materials to all materials except gold, copper, silver and glass. The axiom representing this for the case of jewelries is the following:

\[
\text{JewelRareMaterial} \equiv \text{Jewelry} \sqsubseteq \neg \text{madeOf\{Gold\}} \\
\sqsubseteq \neg \text{madeOf\{Silver\}} \sqsubseteq \neg \text{madeOf\{Copper\}} \\
\sqsubseteq \neg \text{madeOf\{Glass\}}
\]  

(6)

The query then iterates over the individuals of the knowledge base selecting objects from all the 18 jewelry subclasses whose material is not one of the above four. The results include objects from different archaeological museums made of materials such as ivory, bone, amber, pebles, lazourite, gemstone. It should be mentioned that the query rewriting technique cannot be applied to knowledge bases containing axioms like (6).

6. Conclusions and Future Work

Digital Cultural Heritage has been one of the most ambitious and most promising scopes at international level. All over the world, cultural institutions have been digitizing their collections of books, manuscripts, newspapers, maps, museum mobile and immobile objects, archives, audio and visual material, photographs, and are making them available online. Searching for information over all available spaces and semantically interpreting the available cultural content has been one of the main targets of activities performed in national, European and international levels. Different metadata schemas are used to annotate the digitized material.
and make its access feasible for citizens. Europeana, as well as national and thematic content aggregators provide access to the distributed content through collection of contributing metadata schemas. In this framework, semantic interoperability has been identified as one of the main targets of these developments. Recent results in the Semantic Web and the Linked Open Data fields can be used to achieve these goals. Moreover, user engagement and involvement in evaluating and contributing to the aggregated content and the provided services has been recognized as one of the most critical issues for the development of the field in the following years.

In this framework, there is urgent need for systems that can assist both content providers and citizens, users of the field, in generating, disseminating and accessing cultural information in an interoperable, effective and efficient way. To do this, on the one hand, they should provide easy and friendly means for generating mappings between different metadata standards and common models used at national or international levels, as in Europeana, and on the other hand they should provide rich and expressive data representations, that can be shared, used and re-used by content users and providers. Using semantic web and linked open data technologies is a significant way for moving in this direction. Moreover, involving the user and user groups in content tagging, content evaluation and enrichment is a crucial aspect of future developments.

The current paper proposes a system, including a mapping and (Europeana) ingestion tool, for implementing, in a simple, semi-automatic, user-friendly way, the required mappings and data transformations. Using this system, different users’ metadata schemas can be mapped, e.g., to the European Data Model , and expressed in RDF and OWL. As a consequence, they can be used by reasoning and other explorative techniques, in which data from various sources and formats are combined and are appropriately presented to the users so as to cover their needs. In this framework, we propose query answering as the technical approach which can assist content providers and users to enrich their data, to get effective answers meeting the semantics of their queries and to have the ability to validate their selections, creating links between related data and sources of information.

The computational cost of query answering is currently affordable when dealing with normal sized knowledge bases and content sources. Nevertheless, the computational load can become excessive when data and inferences are made at very large scales, e.g., at the Europeana level. The method of query rewriting is proposed in such cases. In particular, future work involves implementation of scalable query-answering services providing the end users with as much as possible and as relevant as possible correct answers to queries expressed in terms of extended vocabularies. Achievement of this will be targeted by overcoming the expressive limitations of existing mapping models, including the scalability and tractability issues. Furthermore, extending the query rewriting approach, the queries obtained by expansion of the initial query can be obtained in an ordered, progressive way, according to significance. Moreover, involving user characteristics, profiles and behaviours in the query answering approaches is another means for reducing their computational load and match their performances to the context of the interaction. Various studies have been produced dealing with cultural content uses and user profiles [21]. Users of Europeana have been classified to categories, including the general user, the schoolchild, the academic user, the expert researcher and the professional user. The types of uses include entertainment, queries about a cultural or historic subject, or person, queries about the current location of cultural heritage, desire for becoming part of a community of interest. Particular personas have also been created in this framework. These user classes can be used in the query answering implementation, so as to rank and prioritise concepts or objects that match the user profiles or uses during the query answering process, or during the presentation of results to users.

Questions of storage, access and analysis are major challenges in the above framework for users and researchers. Cloud infrastructures can offer number of benefits: On the one hand, they can more easily deal with the enormous and constantly increasing amounts of cultural data that have come into existence through digitization initiatives. On the other hand, cloud structures can enable collaborations between researchers working on similar problems and being separated by physical distance, by integrating them into virtual research communities and allowing them to securely share data between trusted sources. Interweaving the semantic query answering techniques with cloud computing is another direction of future work that we currently investigate as a means for providing users with effective and efficient access to the cultural heritage field.

Acknowledgments The authors wish to thank the Hellenic Ministry of Culture and specifically Ms
Metaxia Tsipopoulou, Director of the HMCT Directorate of National Archive and Monuments and Mr Kostas Chatzixristos, Director of the Informatics Division of HMCT for their assistance in working with the cultural content of the www.collections.culture.gr. We also thank Miss Effie Pasatzie for assisting with the mapping of the HMCT metadata schema to EDM through the NTUA ingestion tool.

References

Fig. 4. Semantic technologies within the cultural heritage framework
Fig. 5. Example
<table>
<thead>
<tr>
<th>Query</th>
<th>Scope</th>
<th>Location of objects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pottery of Mycenaean period found in museums of Peloponnese, Crete, Aegean islands.</td>
<td>Research for findings while designing organization of an archaeological (physical and virtual) demonstration</td>
<td>Such items can be found in the HMCT portal and in Europeana coming from the archaeological museums of Kalamata Peloponnese, Heraklion, Crete, Irakleia Crete, Sitia Crete, Kea and Chios in Aegean.</td>
</tr>
<tr>
<td>Minoan pottery with sea pace decoration</td>
<td>Research for publishing findings from excavation</td>
<td>Items from the archaeological museums of Heraklion and Sitia, Crete.</td>
</tr>
<tr>
<td>Jewellery of Hellenistic period</td>
<td>Collection of content for museological educational programs</td>
<td>Items from the archaeological museums of Thessaloniki, Kalamata, Larissa, Athens, Pala.</td>
</tr>
<tr>
<td>Molyvos(sub: king’s stamps) of the Middle and Late Byzantine period</td>
<td>Presentation of characteristic archaeological objects in a University course</td>
<td>Items from the Museum of Byzantine Culture and the Numismatic Museum Athens.</td>
</tr>
<tr>
<td>Minoan and Mycenaean Wall Paintings</td>
<td>Organisation of content for archaeological tours</td>
<td>Items from the Archaeological Museums of Thiva and Heraklion.</td>
</tr>
<tr>
<td>Figurines from the Geometric to the Early Classical period</td>
<td>Electronic aggregation of findings, from a single excavation, that are scattered in different locations or Departments</td>
<td>Items from the National Archaeological Museums and the Museums of Thiva and Samos.</td>
</tr>
<tr>
<td>Engravings and paintings of the 19th century</td>
<td>Search for materials in order to create a thematic portal of archaeological content.</td>
<td>Items from the Museum of Byzantine Culture, the Byzantine and Christian Museum, the Redynno Preveli Monastery and the Pyrgos Picoouli Museum in Atenos.</td>
</tr>
<tr>
<td>Coins of the Late Byzantine period</td>
<td>Preparation of a publication or organization of an exhibition</td>
<td>Items from the Museum of Byzantine Culture and the Numismatic Museum Athens. Adam.</td>
</tr>
<tr>
<td>Individual inscriptions of the Roman period</td>
<td>Providing additional educational content to courses (e.g., history) of the primary or secondary education.</td>
<td>Items from the Epigraphic Museum.</td>
</tr>
<tr>
<td>Copies of Byzantine paintings of the 20th century</td>
<td>Organising touristic visits for educational or training purposes.</td>
<td>Items from the Byzantine and Christian Museum.</td>
</tr>
</tbody>
</table>

Fig. 11. User queries and associated context