The aggregation of heterogeneous metadata in Web-based cultural heritage collections. A case study

Silvio Peroni[1], Francesca Tomasi[2], and Fabio Vitali[1]

[1] Department of Computer Science, University of Bologna, Italy
{essepuntato,fabio}@cs.unibo.it
[2] Department of Classical Philology and Italian Studies, University of Bologna, Italy
francesca.tomasi@unibo.it

Abstract. In this paper we discuss some issues related to digital libraries that are born as aggregators of cultural heritage resources harvested from different distributed digital repositories. Whenever such aggregations fall short of providing a real service to users, most of the time this is due to the inadequacy of the data (or conceptual) model proposed for the mapping and exposed to the final user. Our aim is to enhance the formal representation of cultural heritage materials, reasoning about some key concepts to improve the quality of the description of digital resources by refining the conceptual model. The relation between original object and the medium used for its digital representation, the stratification of levels in the description of the object, and the separation between the description of the involved individuals and their roles, are some concepts about which we have reflected in order to propose a new approach to the description of resources. Correctly approaching these three issues, often misunderstood by aggregation platforms, may help in a systemic refinement of the data model of the relevant ontologies, improving the usefulness of the information made available by the aggregation framework. Additionally, even the technologies used to make records of digital libraries available to the public may require a reflection, since it is easy to overlook subtle points in important technologies such as Linked Open Data on the Web when used as one of the publication mechanisms.
In order to discuss about these issues and to proceed towards a redefinition of a data model for cultural heritage, we focus particularly on Europeana, the principal European Digital Library, as the foremost and most paradigmatic example of aggregators, and on the Europeana Data Model (EDM) as the conceptual model on which metadata from different repositories are mapped in Europeana.

Keywords: Digital library, Aggregation, Linked Data, DC, Europeana, EDM.

1 Introduction

The proliferation of standards for metadata on the WWW has largely affected the problem of aggregating descriptive information in distributed Web-accessible platforms.

In the last years, every institution managing a digital library of its cultural heritage has been creating Web-accessible resources of their treasures as complex digital objects (Nelson and Van de Sompel, 2006), often where the data (the digital reproduction of the original object) has been coupled with a variously rich collection of metadata (the description of both the original document and its digital version). Such institutions often use, for the implementation of their own digital libraries, a number of different metadata schemas and different controlled vocabularies. Even though syntaxes such as XML, or formalisms such as RDF, or even ontological languages such as OWL, have been used extensively in the wake of the guidelines of the Semantic Web community, there are still strong differences preventing consistent integration of their metadata element sets and value, or controlled vocabularies[1].

This is due to various reasons, mostly having to do with the various types of institution producing these collections. For example historical archives use different descriptive standards than libraries or museums; secondarily, even the same type of object requires the use of different languages depending on the institutions handling it: for example manuscripts are documents just like archival records, but they must be described according to different metadata models; thirdly, the data format itself of the digital object to be reproduced may be affected by the nature of the original object: a text file, an audio file, a picture, a video, a 3D animation not only require different descriptions but can be interpreted as different levels of digital representation of the same original analogical object, as is the case with a textual transcription, a low-resolution photo, and a high-resolution 3D representation of an ancient codex.

The complexity of the domains being described has led to the adoption of a number of different metadata schemas and models (such as DC Elements (Dublin Core Metadata Initiative, 2012a) and Terms (Dublin Core Metadata Initiative, 2012b), SKOS (Miles and Bechhofer, 2009), CIDOC-CRM (Crofts et al., 2011), EAD/EAC[2], METS[3], TEI[4], FRBR (International Federation of Library Associations and Institution, 2009), BIRO[5] and FaBiO[6] (Peroni and Shotton, 2012) and to the use of a large number of different value vocabularies, i.e., thesauri, authority files and controlled lists, such as Geonames[7], Art & Architecture Thesaurus[8], Iconclass[9], WordNet[10] (Miller, 1995), Dewey Decimal Classification[11], DBPedia[12] (Bizer et al., 2009), Virtual International Authority File[13], Library of Congress/Name Authority File[14], etc.[15].

---

[1] See the Open Metadata Registry (OMR) for a list of vocabularies and element sets http://metadataregistry.org/

2 http://www.loc.gov/ead/

3 http://www.loc.gov/standards/mets/

4 http://www.tei-c.org

5 http://bibliontology.com/

6 http://purl.org/spar/fabio

7 http://www.geonames.org

8 http://www.getty.edu/research/tools/vocabularies/aat/index.html

9 http://www.iconclass.nl/home

10 http://wordnet.princeton.edu/

11 http://www.oclc.org/dewey/

12 http://dbpedia.org/

13 http://viaf.org/

14 http://id.loc.gov/authorities/names.html

This variety of approaches greatly complicates the identification of a model for the integration of metadata collections so heterogeneous in element sets and value vocabularies and it makes it difficult to develop aggregations that do not lose information during the integration of the different metadata sources. The main task of an aggregation platform is in fact to define a "mapping model", i.e., a data model on which individual models, defined and used by the different institutions, are compared and then converted to be exposed to the final user.
Still, aggregation of metadata collections is becoming important: while until a few years ago the focus of the digital library community was mainly directed to the study of technologies and architectures for the creation of digital library and distributed systems of digital objects of individual collection (Arms, 1995; Kahn and Wilensky, 2006), now the focus is rapidly shifting to the syntactic and semantic aspects of interoperability of heterogeneous digital libraries (Candela et al., 2008), and also and particularly to the creation of aggregation frameworks that give access to heterogeneous metadata collections and expose them as integrated datasets (Brogan, 2003).
The OAI-PMH (Open Archive Initiative – Protocol for Metadata Harvesting)[16] has started to reflect on systems for metadata harvesting, so that they can then be enjoyed by a larger community of users on the Web (Shreeves et al., 2003). But the problem of the adequacy of the model to map the collected metadata has not been solved.
For instance, the imposition of Dublin Core as a reference standard for exposing metadata to the end user does make it easy to integrate content, but on the other hand ends up losing much information that oftentimes in the provenance collections is available in a more sophisticated fashion, perhaps because it is based on a much richer schema.
A successful aggregation platform must then face the question not only of the lowest common denominator in schemas and vocabularies being integrated, but also of how to render existing sophisticated properties, and even to create meaningful relationships between descriptions from different collections, for instance when they are connected because they share properties. Whenever the use of the same names, places, or people is a signal of one such connection, the reduction of the original descriptions to a simplified model loses data and prevents the connection to be established, thereby reducing the richness of the information retrieved.

It is essential, therefore, that the mapping model planned for the aggregation identifies not only the syntax, the formalisms and the ontology languages it uses (such as XML, RDF or OWL), but also that certain key concepts are taken into consideration, so that the integrated description is both effective and complete. In this paper, we identify three of these concepts, that we consider among the most relevant ones in the description of digital resources:

1. the awareness that there are different levels of medium through which an object can be translated, researched and described, which means that the rela-

[15] A comprehensive list of metadata and values vocabularies published as Linked Data Sets could be found in: W3C Incubator Group Report 25 October 2011. Library Linked Data Incubator Group: Datasets, Value Vocabularies, and Metadata Element Sets at: http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabdataset-20111025/
[16] http://www.openarchives.org/pmh/

tionship between the original object and its digital representation is not trivial;
2. the awareness of the multiplicity of potential levels of description of the subjects in a resource and
3. the awareness of the distinction between subject and object and between a person and his role, with the relevant space-time considerations of the exercise of that role.

Analyzing these concepts and giving them a formal representation, means also proposing the most appropriate ontological predicates to represent the above challenges.
In order to reflect about these problems and to give a possible solution in a real case scenario, we will use a case study familiar in Europe: the Europeana collection[17], i.e., the European Digital Library, a distributed access point to Europe's multilingual cultural heritage in a digital form, containing an aggregated form of the content of several dozens of the main digital collections of digital heritage resources.
In this paper, we describe some features of Europeana and analyse the query results of the current implementation, proposing some reflections on the project in the current phase (Section 2). We describe some related works and mostly the metadata standards used by Europeana (Section 3). In section 4 we focus on the three important aspects introduced above as a necessity for a richer aggregation data model: the "media type" concept (Section 4.1), the multi-layer description (Section 4.2), and the connection between roles and values (Section 4.3); in the end we analyse the sophistication of the current implementation of European with regard to their Linked Data offering (Section 4.4).

2 A case study: Europeana

The main aim of the project Europeana is to collect metadata from a large number of providers, mainly cultural institutions, across Europe, and to enable search and discovery of cultural items described therein.
The metadata aggregation is based on a mapping between the providers' data description and the Europeana model. The Europeana v1.0 project[18] (Aloia et al., 2011) proposes the Europeana Data Model (EDM)[19] that defines a set of classes and properties to be used in Europeana for describing cultural objects. The EDM (Europeana Project, 2012a) is a clear improvement over the earlier data model, the Europeana Semantic Elements (ESE) (Europeana Project, 2012b). ESE was meant to express the providers'

[17] There are some attempts to create aggregation framework in the field of cultural heritage. In USA from 2002 there is the big project IMLS Digital Collections and Content (http://imlsdcc.grainger.uiuc.edu/). In Italy there are the two portals Michael (http://michael-culture.it/mpf/pub-it/index.html), that is also an European Multilingual Inventory of Cultural Heritage, and CulturaItalia (http://www.culturaitalia.it/). We prefer to work in a European dimension, choosing the Europeana case (http://europeana.eu). It's therefore the most relevant and known project in the field of cultural heritage in which also Michel and CulturaItalia are converged.
[18] Now in: Europeana Professional knowledge-sharing platform http://pro.europeana.eu.
[19] The family of technical documents about EDM (in particular Definition, Primer, Guidelines) could be found at: http://pro.europeana.eu/web/guest/edm-documentation

datasets using the Dublin Core (DC) standard as the "lowest common denominator" for the data model to be used in the conversion. EDM, based on the DCMI Metadata Terms and a number of more advanced metadata models, adopts "an open, cross-domain Semantic Web-based framework" leaving each provider free to use their preferred metadata standard with regard to the element sets and the vocabularies of values (Europeana Project, 2011a).

Given this, an extreme heterogeneity can be observed in the descriptions of cultural objects, a situation determined by the differences in existing collections that Europeana involves (museums, archives, audio-visual collections and libraries), and by the different kinds of objects described in the cultural repositories that were harvested (i.e. manuscripts, documents, paintings, art and architecture objects, photos, videos, etc.). Also, the variety of descriptive situations depends both on the reference models for the metadata element sets (metadata schemas) and from the vocabularies of values (thesauri) used by data providers.

Furthermore, even if the EDM has been introduced, a majority of records in Europeana still seems to follow the old ESE data model and the users' query interface is based on just a few categories of Dublin Core. The main problems in using the current release of Europeana derives in part from the above-mentioned issues: the original object descriptions are often lost and the ESE, the first proposed model, was not sufficient to describe the complexity of many cultural objects, as it was based on just a few DC categories. Providers that had complex and structured descriptions have had to force them into to a much simpler model and providers that created descriptions in a model not compliant with ESE have lost data or have forced them into incorrect properties, leading to aggregations of imprecise information. Consequently, many user queries cannot be satisfied completely. Probably the full integration of EDM in Europeana records will help towards completeness and correctness of the contained information. But some other improvements, on which we will focus on in the next sections, could solve many situations that not even EDM takes into complete account.

The whole point of this effort clearly is to allow users to enact better queries and to obtain better results through the use of a sophisticated and increasingly ontological metadata model on which to let the search engine work. Europeana is now working on improving the quality of the responses to users' queries, but much is still to be done.

For example at the moment there is no possibility to perform a multilingual search, although the vocabulary alignment is a problem being studied currently by the Europeana group, since one of the aims of subproject EuropeanaConnect[20] is in fact to solve this gap.

Some other features are planned to be implemented using Semantic Web technologies[21]. For example, Europeana lacks a semantic network for the subjects, that could help users in finding records by specifying either the exact word or any of its synonyms, hyponyms, hypernyms, related terms, etc. WordNet, for example, has been published as a RDF vocabulary[22] and could be used after a vocabulary alignment. Another issue that Semantic Web technologies could solve is to enrich the existing

[20] http://www.europeanaconnect.eu/

[21] Library Linked Data Incubator Group wiki. Use Case Europeana: http://www.w3.org/2005/Incubator/lld/wiki/Use_Case_Europeana

[22] Wordnet 3.0 in RDF: http://semanticweb.cs.vu.nl/lod/wn30/

metadata sets, contextualising objects and using authoritative sources, including controlled vocabularies, concretely shared in an integrated environment. The "related content", presented in each object description in Europeana, is now mostly limited to other objects from the same data provider. More complex and structured relationships between what Europeana calls the provided "Cultural Heritage Object (CHO)" (Europeana Project, 2012a) and other pertinent resources, internal (other data providers) or external (other digital libraries), could be solved in a Linked Data perspective.
Although some features are now being studied as an experimental increment to the existing feature set, some discrepancies could be noticed between the EDM and the objects described, for which we do not know of any on-going work.
We want therefore to propose a model of resources description that, starting from EDM, uses ontologies in the most appropriate way.

3 Related works

All aggregative data infrastructures in cultural heritage face the problem of defining a conversion model that limits the loss of information. Some projects, such as the Heritage of the People's Europe (HOPE)[23], the Archives Portal Europe network (APEnet)[24], The European Library (TEL)[25], or the European Cultural Heritage Online (ECHO)[26], proposed solutions for such mapping, sometimes in cooperation with Europeana: "the realization of ADIs [Aggregative data infrastructures] for CH [Cultural Heritage] can be particularly complex when compared to other disciplines due to the possibly high heterogeneity of data sources involved" (Bardi, Manghi and Zoppi, 2012). This implies that a deep knowledge of vocabularies and ontologies is required in order to propose a functional and exhaustive model. Many techniques have been developed in order to achieve metadata interoperability (Haslhofer and Klas, 2010) and some applications have been proposed, even software systems for the realization of Aggregative Digital Library Systems (Manghi et al, 2010). But the main issue is to reflect on the different metadata models in use in projects related to cultural heritage: "metadata mapping is the appropriate technique in integration scenarios where an agreement on a certain metadata standard is not possible" (Haslhofer and Klas, 2010).

---

[23] "HOPE presents the heritage of the people of Europe and their history in the process of social change", http://www.peoplesheritage.eu/index.htm. In the section "News" it's possible to find a list of publications. Between them the: The Common HOPE Metadata Structure, including the Harmonisation Specifications, http://www.peoplesheritage.eu/pdf/D2_2_Metadata%20Structure.pdf.

[24] "APEnet (Archives Portal Europe) is a Best Practice Network project supported by the European Commission in the eContent*plus* programme and its objective is to build an *Internet Gateway for Documents and Archives in Europe* where seventeen European National Archives in close cooperation with the EUROPEANA initiative will create a common access point to European archival descriptions and digital collections", http://www.apenet.eu/.

[25] "Designed to meet the needs of the research community worldwide, our online portal [the European Library] offers quick and easy access to the collections of the 48 National Libraries of Europe and leading European Research Libraries", http://www.theeuropeanlibrary.org.

[26] Open Access Infrastructure for a Future Web of Culture and Science**,** http://echo.mpiwg-berlin.mpg.de/home.

In this section we give references to the metadata schemas used by metadata aggregators such as Europeana. In its specifications (EDM), Europeana mentions vocabularies, models and ontologies adopted in its data model (Europeana Project, 2012a). The aim is to represent metadata for cultural heritage objects and to give access to digital representations of these objects. The EDM moves in the context of data aggregation, where objects can be complex, and several data providers may entertain different views on them.
The basis of the metadata description is RDF statements. An XML Schema has been defined for describing classes and properties. Some classes and properties are re-used from public models: DC, DCterms, SKOS, OAI-ORE, CIDOC-CRM, FRBR. Some other classes and properties are specifically created for the EDM and are mostly equivalent to predicates used in the most common ontologies.
In addition to classes and properties, Europeana is defining also controlled vocabularies useful for CHO interoperability (such as AAT, DDC, DBpedia, Iconclass). The main aim of Europeana is to work on Linked Data both exposing record sets (Haslhofer and Isaac, 2011) and using Linked Data resources (Haslhofer et al., 2010) in order to augment Europeana content.
In the following sub-sections we introduce the main external models adopted by EDM, highlighting which part of them are effectively used.

3.1 Dublin Core

The current versions of the Dublin Core (DC) Metadata Elements (Dublin Core Metadata Initiative, 2010b) and of the DC Metadata Terms (Dublin Core Metadata Initiative, 2010a), are the most widely used vocabularies for describing and cataloguing resources. These vocabularies have become particularly important and relevant for sharing metadata about documents among different repositories and digital libraries. While very useful for creating basic metadata that permit bibliographic resource descriptions (e.g., *creator*, *contributor*, *publisher*, *format*), the main limitation of DC is a consequence of the generic nature of its terms. In fact, its classes are organised without a strong hierarchical structure and their properties often lack in clear domain/range definitions. EDM makes extensive use of DC Elements and DC Terms entities, such as the properties *dc:subject*, *dc:contributor*, *dcterms:created* and *dcterms:alternative*.

3.2 SKOS

Data providers, publishers and aggregators, such as Europeana, need to classify the resources they publish according to discipline-specific thesauri and classification schemes. The Simple Knowledge Organization System (SKOS) (Miles and Bechhofer, 2009) is an RDFS ontology to support the use of knowledge organization systems (KOS). A large number of well-known thesauri and classification systems have started to convert their specifications into SKOS documents, such as the "Nuovo Soggettario" of the National Central Library in Florence[27]. This makes SKOS the de facto standard for encoding controlled vocabularies for the Semantic Web. EDM uses the

[27] http://thes.bncf.firenze.sbn.it/

main SKOS class, i.e. *skos:Concept*, defined as a particular kind of *edm:NonInformationResource* for introducing an idea or notion.

3.3 FRBR

The Functional Requirements for Bibliographic Record (FRBR) (International Federation of Library Associations and Institution, 2009) is a general model for describing bibliographic entities, such as documents and artistic works. FRBR specifies four basic concepts – work, expression, manifestation and item – used for characterising a particular bibliographic entity from different perspectives. In particular:

- A *work* is the *abstract essence* of an intellectual or artistic creation, e.g. the ideas in Shakespeare's head concerning the *Macbeth*. A work is realised in one or more expressions;
- An *expression* is the *content* of a particular work at a specific point in time, e.g. the final text of the *Macbeth* written by Shakespeare or its Italian translation made by Andrea Maffei. An expression is embodied in one or more manifestations;
- A *manifestation* is the particular *format* in which an expression is stored, such as a printed object or a digital document, e.g. the 2005 edition of *Macbeth* published by Penguin Books or its HTML Italian version published by. A manifestation is exemplified in one or more item;
- An *item* is a particular *physical or electronic copy* of the *Macbeth* that a person can own, e.g. the printed version of that book you have in your bookcase or the specific HTML document of its Italian version you are visualising in your browser.

Overall, EDM makes only limited use of FRBR concepts, although it declares explicitly their adoption. The only specific references to FRBR are:

- the class *edm:InformationResource*, defined as union of *FRBR Work*, *FRBR Expression* and *FRBR Manifestation* that results in collapsing completely the hierarchy of the FRBR model;
- the classes *edm:Event* and *edm:Place*, defined as equivalent to *FRBR Event* and *FRBR Place* respectively.

3.4 ORE

The Open Reuse and Exchange specification (ORE specification) (Lagoze et al., 2008) is a standard defined by the Open Archives Initiative for describing and exchanging aggregations of Web resources. Europeana uses two terms from this model:

- *Aggregation*, i.e. a particular resource that aggregates, either logically or physically, other resources;
- *Proxy*, used to refer to a specific aggregated resource in a context of a particular aggregation.

EDM uses all the main classes and properties of the ORE specification. For instance, it allows one to describe a "cultural heritage object" (i.e., *edm:providedCHO*) and its digital representations (i.e., *edm:WebResource*) as a particular aggregation

(*ore:Aggregation*) representing the results of the activity of a particular data provider (i.e., *edm:Agent*).

3.5 CIDOC CRM

CIDOC Conceptual Reference Model (CRM) (TC 46/SC 4, 2006) is an ISO standard defining a model for describing and sharing cultural heritage information. It provides entity definitions and a formal multi-level structure to link physical objects to related events and agents (i.e., people and organisations), so as to represent a mediator between different sources of cultural heritage information (e.g., museums, libraries and archives).
EDM aligns some of its classes and properties to the CIDOC CRM specification, for instance the class *edm:Event* as equivalent to *E4 Period*, the class *edm:InformationResource* as subclass of *E73 Information Object*, and the property *edm:wasPresentAt* as equivalent to *P12 occurred in the presence of (was present at)*[28].

4 Using a metadata aggregator: the case of Europeana

In this section we discuss some issues arising when using one of the most relevant metadata aggregators currently available, the Europeana project. The EDM rationale is based on the following principles (Europeana Project, 2012b):

1. distinction between "provided objects" (painting, book, movie, etc.) and their digital representations;
2. distinction between objects and metadata records describing an object;
3. allow for multiple records for a same object, containing potentially contradictory statements about it;
4. support for objects that are composed of other objects;
5. compatibility with different levels of description;
6. standard metadata format that can be specialized;
7. support for contextual resources, including concepts from controlled vocabularies.

Although the Europeana core classes stress the difference between the provided object (*edm:ProvidedCHO*), i.e. the "real object", and its digital representation (*edm:WebResource*), i.e., its Web resource, sometimes this difference is not evident at all in the aggregated metadata exposed to the final user, generating confusion. Sometimes the description seems to be addressed to the electronic version, some other to the original work, without a clear distinction (see 4.1). Additionally, the Europeana contextual classes, which are designed to answer to the four fundamental questions of the *who* (the Agent), the *where* (the Place), the *when* (the Time), and the *what* (the Concept) of the object, sometimes are not correctly represented in the metadata de-

[28]While in the Europeana Data Model the CIDOC CRM property *was present at* has the identifier *P121*, in (TC 46/SC 4, 2006) that property is defined as *P12 occurred in presence of (was present at)*.

scription because of a potential multi-layer representation issue derived from the stratification of object and subject (see 4.2). Therefore, the rationale of the EDM appears not always respected and the application of the listed properties is not totally functional, mostly in the connection between a person and his/her role (see 4.3). Here we describe some examples of the limits in applying classes and properties of EDM.
The EDM represents the best data model proposed in the field of aggregative data infrastructure in the cultural heritage domain, and therefore it represents an excellent starting point for the introduction of support for our three concepts, allowing to elegantly describe them through ontological assertions, leading to a data model refinement.
The concepts we analyze in the following, with real examples from Europeana records, originated from the considerations that EDM shows weaknesses in the conceptualization of these three issues, and that the final user is actually exposed to the drawbacks caused by these limitations when accessing the metadata made available by the Europeana site.

4.1 The media type

The first and most evident source of confusion is the concept of media type found as the topmost choice in the filter section after every query (*edm:type* = text, image, audio, video). The media type is sometimes congruent with the type of the provided CHO, and sometimes to its web resource. Yet, the rationale of Europeana is to distinguish between the description of the CHO and its digital representation. Yet, if we search for any object called "illuminated manuscript" we receive different answers: sometimes it has type "text", sometimes type "image" some other times "physical object". In general, resources classified as IMAGE are in fact image files (regardless of whether they represent pictures or physical objects such as buildings or statues or manuscripts), but resources classified as TEXT are sometimes texts, and sometimes images of texts (e.g., photographs of old volumes or manuscripts). One may wonder which would be more useful for searches, i.e., for the media type to refer to the web resource, providing a description of a computer-specific object, or to the cultural heritage object, which is what the user would be actually searching for. In both cases, it would be quite important to provide subtypes: they could be either subtypes of the relevant Internet MIME type[29] in the first case, or a selection of the values found in the *dc:type* facet of the records as specified in the collections in the other case.

4.2 Multi-layered descriptions

The issue of the separation between web resource and cultural heritage object can be subsumed in the issue of separating objects and subjects in record descriptions. In fact, what does exactly a Europeana record describe? Is it an image, the content of the image, or the object represented in the content of the image? Sometimes this is easy to understand, and sometimes it creates interpretation issues, and the problem of distinguishing between object and subject in a record can go several layers deep. For instance, consider the 1756 publication by Giambattista Piranesi called "Le antichità

[29] http://www.iana.org/assignments/media-types/text/index.html

romane", containing prints of famous Roman monuments, including the Coliseum. A best seller of the time, the volume appears in several of the collections of Europeana. We analysed a few pages from the Bildarchiv Foto Marburg. According to one of them[30], the item described is of *dc:type* druck (print), *dc:creator* Giambattista Piranesi, *dc:date* 1756, and, for some reason, *dc:subject* Lord Charlemont and King Gustaf III of Sweden. According to another[31], the same item is of *dc:type* amphitheatre, and *dc:date* 70/80 a.D. (but no *dc:creator*); it also reports (in the *dc:description* field) that the actual photo was taken between 1960 and 1970 by Konrad Helbig, and that the content is an extract of the Piranesi's book of 1756. Looking for *dc:type* amphitheatre, we can find several color photographs of the Coliseum[32] present in the same collection: besides a *dc:type* amphitheatre, we discover that *dc:date* is also 72/80 a.D., but also *dc:format* is travertine, and *dc:contributor* is Vespasianus (as contractor).

In cases such as Piranesi's, we are witnessing the existence of multiple layers of subjects: the CHO being described is a 1956 b/w photograph by Konrad Helbig, whose subject is a 1756 print whose creator is Giambattista Piranesi, whose subject is a 70 a.D. travertine amphitheatre whose creator (as contractor) was Vespasianus. If we add the issue of the media type of the web resource, as introduced in section 4.1, an additional level becomes manifest: we are describing a 21st century JPEG image of a 20th century photograph of a 18th century print of a 1st century building.

One of the most frequent dilemmas for a provider of metadata about an object is deciding whether interesting information for which no natural facet is available should be omitted, forced into an inappropriate facet (e.g., the *dc:type* or *dc:subject* in the above examples), or dumped into a generic container (e.g., the *dc:description* above). A better solution would be to use a metadata model whose characteristic naturally accommodates the interesting information. As such, a simple solution exists already for the layers of subjects: while in DC the subject is "the topic of the resource [that is] typically […] represented using keywords, key phrases, or classification codes", in FRBR "the «has as subject» relationship indicates that any of the entities in the model, including work itself, may be the subject of a work".

Thus the example of the print by Piranesi could be expressed more precisely and with fewer misunderstanding as a record for a JPEG image whose *frbr:subject* is a 1956 photo whose *frbr:subject* is a 1756 print whose *frbr:subject* is a roman building, for instance as in figure 1[33]:

---

[30] http://www.europeana.eu/portal/record/08501/7B74073B6E9E90F5B572EF6DF20426AF0135202E.html

[31] http://www.europeana.eu/portal/record/08501/84744383AA401E7784B38ED00BD47A276DD1D230.html

[32] for instance, http://www.europeana.eu/portal/record/08501/43E4B1EF54983567EC92DDCDD57B3DBD2D4CC013.html

[33] In this and all subsequent examples, we use the following prefixes (please note that the prefixes with the Europeana domain are fictitious, are present in these examples only as a suggestion and do not correspond to existing ontologies):

@prefix resource: <http://data.europeana.eu/resource/> .

@prefix ontology: <http://data.europeana.eu/ontology/> .

@prefix foaf: <http://xmlns.com/foaf/0.1/> .

@prefix pro: <http://purl.org/spar/pro/> .

```
ontology:Photography rdfs:subClassOf frbr:Work .
ontology:Print rdfs:subClassOf frbr:Work .
ontology:Amphitheatre rdfs:subClassOf frbr:Work .

resource:jpeg-photo a ontology:JPEGImage
       ; frbr:subject resource:photo .

resource:photo a ontology:Photography
       ; dc:date "1956"
       ; frbr:subject resource:antichità-romane .

resource:antichità-romane a ontology:Print
       ; dc:date "1756"
       ; dc:creator "Giovanbattista Piranesi"
       ; frbr:subject resource:colosseum .

resource:colosseum a ontology:Amphitheatre
       ; dc:date "70/80"
       ; dc:creator "Vespasianus [Auftrag]" .
```

Fig. 1. an OWL rendering of the correct relationships between subject layers using FRBR (all other facets are expressed as in the original examples for simplicity)

Using FRBR in its true meaning, so as to distinguish the stratification of layers resulted from describing an object (the idea, the content, the format and the specific item), it is also possible to better distinguish between the different levels (the work, the expression, the manifestation and the item).
EDM, in fact, defines *dc:subject* more precisely than Dublin Core itself, explicitly specifying that its value is either a string or a reference (thus allowing references to other CHOs), and even defines a sub-property of *dc:subject*, called *edm:isRepresentationOf*, to precisely specify the relationship between representations and represented entities (e.g., a statue and a painting of the statue). Yet, the *edm:isRepresentationOf* property is currently greyed out (meaning that it "will not be used in the first implementation so any values provided for them will not be used"), and the current number of subjects specified as strings will make turning them into references a conspicuous and non-trivial job[34].

4.3 Roles and values

Many of DC properties (e.g., *dc:creator* and *dc:contributor*) are so often found insufficient for representing the information available that in most Europeana resources we have checked many of the actual literals were compositions of the relevant name as

---

@prefix dc: <http://purl.org/dc/elements/1.1/> .
@prefix frbr: <http://purl.org/vocab/frbr/core#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

[34] …not to mention the fundamental problem that in OWL a property can either be a data property (i.e., allowing strings) ***or*** an object property (i.e., allowing references) but not both, so that any OWL ontology based on EDM will have to choose one representation for the values of all the properties, including *dc:subject*, that allow either strings or references as their values. This, in and by itself, will be a major exercise in reconversion and qualification of existing data sets.

well as an indication of a role or other contextual information (e.g. as creator and/or contributor). For example consider:

Creator: Morel, Francois (Radierer) [35]
Creator: Cartographer : Ryther, Augustus[36]
Creator: Friedrich, J. C. F. [Production][37]
Subject: AUTN=Piranesi Giovanni Battista; AUTA=1720/ 1778[38]

The universality of this approach is evident, and similarly evident is the need to provide more information than a mere name, although the syntaxes, the metadata models and the provided information differ.
A better solution could be obtained by promoting literals into first-class objects – e.g., converting people names into individuals of classes such as *edm:Agent* or *foaf:Person*[39] – and dealing with people names and roles separately. There exist two alternative ways to address efficiently and effectively this issue.
On the one hand, we can create explicit sub-properties of properties such as *dc:creator* or *dc:contributor*. For instance, by allowing "cartographer" to become an explicit sub-property of *dc:creator* (e.g., property *ontology:cartographer*), the identification of name and roles becomes possible and easy, and consequently the queries become more powerful, as shown in the following excerpt related to *The Cittie of London 31* by Augustus Ryther:

```
resource:cittie-of-london-31 ontology:cartographer resource:ryther .

resource:ryther a foaf:Person
       ; foaf:givenName "Augustus"
       ; foaf:familyName "Ryther" .
```

A problem with this approach is that the TBox of the ontology needs to specify every role as a sub-property of *dc:creator*, and needs to be modified every time a new role is defined, which is not a good design principle in general.
An alternative is to define people's roles as individuals of a class. In theory, CIDOC CRM already implements this behavior by using the meta-property *P14.1 in the role of* (Crofts et al., 2011) (a property of property *P14 carried out by*) so as to specify the role that an agent has in the context of a particular event (e.g., the creation of an artistic work) through an instance of the class *E55 Type*. However, the official RDFS ontology of CIDOC CRM[40] does not implement any meta-property and in reality RDF (as which, ultimately, CIDOC CRM statements are most probably expressed) lacks

[35]http://www.europeana.eu/portal/record/08547/DC2A5E3DB3A0675D12DA7699647D1D8FA1B9293D.html
[36]http://www.europeana.eu/portal/record/92037/25F9104787668C4B5148BE8E5AB8DBEF5BE5FE03.html
[37]http://www.europeana.eu/portal/record/08547/AD78BAEF3D932EF43765BAD78FE8E707EA4AFF85.html
[38]http://www.europeana.eu/portal/record/08504/3F54955AB672A4DA3C5A9C268D659A0075170C7F.html
[39]http://xmlns.com/foaf/spec/#term_Person
[40]http://www.cidoc-crm.org/rdfs/cidoc_crm_v5.0.4_english_label.rdfs

the expressive power needed to define meta-properties altogether. To simulate a meta-property in RDFS/OWL we may define an additional class that associates the person to his/her role. For instance, the *Publishing Roles Ontology* (*PRO*)[41] (Peroni et al., 2012) has this behaviour by means of the class *pro:RoleInTime*,:

```
resource:ryther a foaf:Person
       ; foaf:givenName "Augustus"
       ; foaf:familyName "Ryther"
       ; pro:holdsRoleInTime [ a pro:RoleInTime
              ; pro:withRole resource:cartographer
              ; pro:relatesTo resource:cittie-of-london-31 ] .

resource:cartographer a pro:Role .
```

This approach has the advantage of not requiring any modification to the TBox of the ontology whenever a new role is needed, but only to add a new individual of the class *pro:Role*. Note that this is a crucial aspect that a metadata aggregator should handle, so as to be able to adapt its model to additional and unforeseen roles used by different sources.

In addition, PRO allows temporal extents to be specified to characterise somebody acting in a certain role. PRO implements the ontology pattern *Time-indexed Value in Context* (*TVC*)[42] (Peroni et al., 2012), a general model to describe general scenarios in which a *value* is held by *someone* at a *particular time* within a *particular context*.

Handling time and context can be particularly valuable if we are interested in keeping track and answering questions about the some individual's activity within a time interval and/or a social context. The use of DC properties as shown above is particularly inappropriate to model scenarios that involve time, providing no support for even very simple queries such as "return Tim Hetherington's employers in 2008, as well as all the photos he took for them". The main issue in using DC (and similar vocabularies) is that we do not have a way to define such facts as complex time- and context-based situations. Through PRO, the description of such scenarios become possible, for instance, as in the following excerpt describing some Hetherington's photos and employers from 1996 to 2011:

```
resource:hetherington a foaf:Person
       ; foaf:givenName "Tim"
       ; foaf:familyName "Hetherington"
       ; pro:holdsRoleInTime
              [ a pro:RoleInTime # role: photographer
                     ; pro:withRole resource:photographer
                     # Photos taken by Hetherington #
                     ; pro:relatesToDocument
                            resource:cardiff-royal-infirmary-1-32am
                            , resource:cardiff-royal-infirmary-1-45am
                            , resource:specialist-tad-donoho ] ,
              [ a pro:RoleInTime # role: employee at The Big Issue
                     ; pro:withRole resource:employee
                     ; pro:relatesToOrganization :the-big-issue
                     # Photos taken when he worked at The Big Issue #
                     ; pro:relatesToDocument
```

[41] http://purl.org/spar/pro

[42] http://www.essepuntato.it/2012/04/tvc

```
                        resource:cardiff-royal-infirmary-1-32am
                        , resource:cardiff-royal-infirmary-1-45am
                ; tvc:atTime [ ti:TimeInterval
                        ; ti:hasIntervalStartDate “1996”
                        ; ti:hasIntervalEndDate “2007” ] ] ,
        [ a pro:RoleInTime # role: employee at Vanity Fair
                ; pro:withRole resource:employee
                ; pro:relatesToOrganization :vanity-fair
                # Photo taken when he worked at Vanity Fair #
                ; pro:relatesToDocument
                        resource:specialist-tad-donoho
                ; tvc:atTime [ ti:TimeInterval
                        ; ti:hasIntervalStartDate “2007”
                        ; ti:hasIntervalEndDate “2011” ] ] .

resource:cardiff-royal-infirmary-1-32am a ontology:Photograph
        ; dcterms:title “Cardiff Royal Infirmary - 1:32am” .

...
```

Given a model as above it is now possible to answer the aforementioned query through SPARQL 1.1 (Harris and Seaborne, 2012) as follows:

```
SELECT ?employer ?photo WHERE {
        resource:hetherington pro:holdsRoleInTime
                [ a pro:RoleInTime
                        ; pro:withRole resource:photographer
                        ; pro:relatesToDocument ?photo ] ,
                [ a pro:RoleInTime
                        ; pro:withRole resource:employee
                        ; pro:relatesToDocument ?photo
                        ; pro:relatesToOrganization ?employer
                        ; tvc:atTime [ a ti:TimeInterval
                                ; ti:hasIntervalStartDate ?start
                                ; ti:hasIntervalEndDate ?end ] ]
        FILTER (
                xsd:dateTime(?start) <= "2008-12-31T23:59:59Z" &&
                xsd:dateTime(?end) > "2008-01-01T00:00:00Z")
}
```

4.4 Linked Data and data.europeana.eu

In addition to the issues connected to the correct modeling of metadata within an aggregation context, another issue that needs considering is the technology allowing their fruition by end users, which requires making technological choices that are not trivially adoptable. One obvious wish is to make use of the most recent and advanced technologies, such as the support for Open Linked Data. As such it is natural to evaluate the quality and sophistication of the aggregators of cultural heritage resources also as a function of the correctness and straightforwardness by which such technologies have been adopted and implemented, e.g. to render the collections as RDF triples within a Linked Data store. In the last years, several Web-accessible triplestores have been released and have contributed to the growth of Linked Data. Of course triplestores have to follow precise access rules to be easily and correctly queried by clients and to make their answers understandable in all situations.

Although Linked Data radically simplifies the path towards the publications of resources on the web, there are subtle issues that need to be careful considered in order to make sure of the correctness and usefulness of the published materials.
One example comes again from Europeana: in fact, the current implementation of the Europeana web site (`http://data.europeana.eu`) already contains a first selection of items as RDF statements and they are already queryable through Linked Data aggregators[43]. However some limits can be found in how the site handles non-existing and non-translated resources, which prevents this implementation from being fully compliant with the Linked Data architecture. In fact, while there exist precise guidelines for accessing resources for which data is available, there is no clear good practice in the situation in which *no information* about a resource exists in the triplestore. Does "no information available" mean that the resource does not exist, or that it cannot be returned in the requested format, or that it is unknown whether the resource exists or not?
According to (Heath and Bizer, 2011), one of the most important principles of Linked Data is that all the published resources must be defererenceable. *Content negotiation* is necessary, since information about a resource should be always returned according to the format requested by the user who is navigating the Linked Data, e.g., HTML for humans and Turtle for computer agents. Content negotiation usually has the form of a "303 redirect": the client asks for a resource in a specific format, the server answers with a "303 See Other" HTTP status code indicating the URL where that requested representation is available to the client, and finally the client gets the content from the specified URL.
Europeana does in fact correctly implement the "303 redirect" approach for the resources it makes available in RDF, but it does not behave as correctly for non-existing or non-available resources. Regardless of whether the resource exists or not, in fact, the server always returns a 303 redirect, and then, after the client restates the query to the new URL, it returns an error if the resource is non-existent. Good Linked Data policy, on the other hand, specifies that 303 is only returned on existing resources, and an immediate error is returned for non-existing or non-available resources.
Two different approaches can be adopted for the return code, depending on which perspective is adopted: in an *Open World perspective,* we cannot state whether a resource exists, but we can only say whether we have data about it, while in a *Closed World perspective*, if no data is available about a resource, then the resource itself does not exist.
Depending on which of the above views the server adopts, the client should expect a different reply than a "303 See Other" when its initial request cannot be satisfied, as shown in table 1.

	Open World view	**Closed World view**
Data about the resource is not available in the requested format	*406 Not Acceptable* Information about the resource exists but it cannot be returned in the requested format	*406 Not Acceptable* There is no information about the resource in the requested format
There is no data	*204 No Content* (no body specified)	*404 Not Found*

[43] Although the Europeana Linked Data project is still ongoing, we hope that what we describe in this section can be seen as valuable and meaningful suggestions for future modifications of the Linked Data infrastructure of Europeana.

for a resource with the specified URI	or *303 See Other* and the indication of a new location that contains empty content in the format requested, e.g. an empty RDF/XML string "*<rdf:RDF />*" There is no information about the resource	The resource does not exist

Table 1. HTTP status code for non-existing or non-available resources in Linked Data

5 Conclusion

The final question that, in this paper, we proposed to solve is then: how can we improve user queries in aggregators of cultural heritage collections? How do we work in the direction of an effective enrichment of the exposed metadata, in order to address the information needs of end users? Europeana, our case study, currently imprecise in the object descriptions mainly because of approximate or incomplete mappings of the original metadata sets into the EDM. The variety of metadata vocabularies, ontologies and models that Europeana addresses makes things difficult to manage. Approaches towards a better integration of the different metadata sources that feed Europeana could be helped by existing works on the creation, extension and alignment of OWL ontologies, but most of the work in the mapping of richer models still needs to be dealt with by hand. The EDM Mapping Guidelines (Crofts et al., 2011) should lead content providers to create descriptions compliant to EDM, at the same time leaving them free to use metadata models and value vocabularies most appropriate to their own internal uses. But although by correctly using the Guidelines many of the existing problems would be solved, some aspects of the EDM could be improved. An adequate reflection on the challenges derived from the mentioned concepts could solve some questions: the relation between original object and the medium used for the digital one, the stratification of levels in the resource description, and the person, provided with a role and a value that evolves in time and space. These are some issues that could be solved with an adequate use of classes and properties and, accordingly, of relevant and pertinent ontologies, as described in this paper. A full and correct Linked Data compliancy, furthermore, is the right direction for the future and will help these aggregation platforms in general, and Europeana in particular, in giving more complete and structured descriptions. Yet, these techniques have to be refined further. The authors' involvement in the Europena tech group, started after a first presentation of the mentioned issues during the IRCDL meeting (Peroni, Vitali, Tomasi, 2012), will help to verify the concrete application of our approach to EDM.

References

Aloia, N., Concordia, C., Meghini, C. (2011) 'Europeana v1.0', in Agosti, M., Esposito, F., Meghini, C., Orio, N. (Eds.), *Digital Libraries and Archives - 7th Italian Research Conference (IRCDL 2011)*, Revised Papers, pp. 127-129, Springer, Heidelberg, Germany. DOI=10.1007/978-3-642-27302-5_16.

Arms,W.Y. (1995) 'Key concepts in the architecture of the digital library', *D-Lib Magazine*, Vol. 1(1), available at http://www.dlib.org/dlib/July95/07arms.html.

Bardi A., Manghi P., Zoppi F. (2012) 'Aggregative Data Infrastructures for the Cultural Heritage', in *MTSR 2012 6th Metadata and Semantics Research Conference*, 28-30 November 2012, Cádiz, Spain. Dodero, JuanManuel and PalomoDuarte, Manuel and Karampiperis, Pythagoras (eds.), Communications in Computer and Information Science, Vol. 343, pp.239-251. Springer, Heidelberg, Germany.

Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., Hellmann, S. (2009) 'DBpedia – A Crystallization Point for the Web of Data', in *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, 7 (3), pp.154-165, available at DOI=10.1016/j.websem.2009.07.002.

Brogan, M. L. (2003) *Survey of Digital Library Aggregation Services*, Digital Library Federation, Washington, District Columbia, USA, available at http://old.diglib.org/pubs/dlf101/dlf101.htm.

Candela, L., Castelli, D., Ferro, N., Ioannidis, Y., Koutrika, G., Meghini, C., Pagano, P., Ross, S., Soergel, D., Agosti, M., Dobreva, M., Katifori, V., Schuldt, H. (2008) *The DELOS Digital Library Reference Model - Foundations for Digital Libraries*, Version 0.98, December 2007, available at http://www.delos.info/files/pdf/ReferenceModel/DELOS_DLReferenceModel_0.98.pdf.

Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M (2011) *Definition of the CIDOC Conceptual Reference Model*, Version 5.0.4, November 2011, ICOM/CIDOC CRM Special Interest Group, available at http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf.

Dublin Core Metadata Initiative (2012a) *Dublin Core Metadata Element Set*, DCMI Recommendation, available at http://dublincore.org/documents/2012/06/14/dces/.

Dublin Core Metadata Initiative (2012b) *DCMI Metadata Terms*, DCMI Recommendation, available at http://dublincore.org/documents/2012/06/14/dcmi-terms/.

Europeana Project (2011) *Europeana Data Model Primer*, Version 1.0, 26 October 2011, available at http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5.

Europeana Project (2012) *Europeana Semantic Elements Specification*, Version 3.4.1, 02 February 2012, available at http://pro.europeana.eu/documents/900548/dc80802e-6efb-4127-a98e-c27c95396d57.

Europeana Project (2012) *Definition of the Europeana Data Model*, Version 5.2.3, 24 February 2012, available at http://pro.europeana.eu/documents/900548/bb6b51df-ad11-4a78-8d8a-44cc41810f22.

Europeana Project (2012) *Europeana Data Model Mapping Guidelines*, Version 1.0.1, 24 February 2012, available at http://pro.europeana.eu/documents/900548/ea68f42d-32f6-4900-91e9-ef18006d652e.

Haslhofer, B., Isaac A. (2011) 'data.europeana.eu: The Europeana Linked Open Data Pilot', in *DCMI Proceedings of the 2011 International Conference on Dublin Core and Metadata Applications (DC 2011)*, 21-23 September 2011, pp. 94-104, The Hague, Netherlands.

Haslhofer, B., Roochi, E. M., Gay, M., Simon, R (2010) 'Augmenting Europeana Content with Linked Data Resources', in Paschke, A., Henze, N., Pellegrini, T. (Eds.), *Proceedings of the 6th International Conference on Semantic Systems (i-Semantics 2010)*, ACM, New York, New York, USA, available at DOI=10.1145/1839707.1839757.

Haslhofer, B., Klas, W. (2010) 'A survey of techniques for achieving metadata interoperability', *ACM Computing Surveys*, Vol. 42(2), pp.7:1-7:37.

Harris, S., Seaborne, A. (2012) *SPARQL 1.1 Query Language. W3C Working Draft.* 05 January 2012. World Wide Web Consortium, available at http://www.w3.org/TR/sparql11-query.

Heath, T., Bizer, C. (2011) *Linked Data: Evolving the Web into a Global Data Space*, Morgan & Claypool Publishers, San Francisco, California, USA, available at DOI=10.2200/S00334ED1V01Y201102WBE001.

International Federation of Library Associations and Institutions (2009) *Functional Requirements for Bibliographic Records*. Final Report, available at http://www.cidoc-crm.org/docs/cidoc_crm_version_5.0.4.pdf.

Kahn R, Wilensky R. (2006) 'A framework for Distributed Digital Object Services', in *International Journal on Digital Libraries*, Vol. 6(2), pp.115-123, available at DOI=10.1007/s00799-005-0128-x

Lagoze, C., Van de Sompel, H., Johnston, P., Nelson, M., Sanderson, R., Warner, S. (2008) *ORE Specification - Abstract Data Model*, Open Archives Initiative, Specification document, 17 October 2008, available at http://www.openarchives.org/ore/1.0/datamodel.

Nelson, M. L., Van de Sompel, H. (2006) 'IJDL special issue on complex digital objects: Guest editors' introduction', in *International Journal on Digital Libraries*, Vol. 6(2), pp.113-114, available at DOI=10.1007/s00799-005-0127-y.

Manghi, P., Mikulicic, M., Candela, L., Castelli, D., Pagano, P. (2010) 'Realizing and Maintaining Aggregative Digital Library Systems: D-NET Software Toolkit and OAIster System', *D-Lib Magazine*, Vol. 16(3/4), available at http://www.dlib.org/dlib/march10/manghi/03manghi.html.

Miles, A., Bechhofer, S. (2009) *SKOS Simple Knowledge Organization System Reference*, World Wide Web Consortium, W3C Recommendation, 18 August 2009, available at http://www.w3.org/TR/skos-reference/.

Miller, G. A. (1995) 'WordNet: A Lexical Database for English', in *Communications of the ACM*, Vol. 38(11), pp.39-41, available at DOI=10.1145/219717.219748.

Peroni, S., Shotton, D. (2012) 'FaBiO and CiTO: ontologies for describing bibliographic resources and citations', in *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, available at DOI=10.1016/j.websem.2012.08.001.

Peroni, S., Shotton, D., Vitali, F. (2012) 'Scholarly publishing and the Linked Data: describing roles, statuses, temporal and contextual extents', to appear in *Proceedings of the 8th International Conference on Semantic Systems (i-Semantics 2012)*, pp.9-16, ACM, New York, New York, USA, available at DOI=10.1145/2362499.2362502.

Peroni, S., Tomasi, F., Vitali, F. (2012) 'Reflecting on the Europeana Data Model', in *Proceedings of the 8th Italian Research Conference on Digital Libraries (IRCDL 2012)*, Revised Selected Papers. Communications in Computer and Information Science (CCIS), Vol. 354, pp.45-56. Heidelberg, Germany: Springer.

Shreeves, S. L., Kaczmarek, J. S., Cole, T. W. (2003) 'Harvesting cultural heritage metadata using the OAI Protocol', in *Library Hi Tech*, Vol. 21(2), pp.159-169, available at DOI=10.1108/07378830310479802.

TC 46/SC 4 (2006) *ISO 21127:2006: Information and documentation – A reference ontology for the interchange of cultural heritage information*, International Organization for Standardization, Geneva, Switzerland, available at http://www.iso.org/iso/catalogue_detail.htm?csnumber=34424.