Integrating Dublin Core metadata for cultural heritage collections using ontologies**

Constantia Kakali
Department of Archive and Library Sciences, Ionian University, Greece
nkakal@panteion.gr

Irene Lourdi
Department of Archive and Library Sciences, Ionian University, Greece
elourdi@lib.uoa.gr

Thomais Stasinopoulou
Department of Archive and Library Sciences, Ionian University, Greece
stasthomais@yahoo.gr

Lina Bountouri
Department of Archive and Library Sciences, Ionian University, Greece
boudouri@ionio.gr

Christos Papatheodorou
Department of Archive and Library Sciences, Ionian University, Greece
papatheodor@ionio.gr

Martin Doerr
Institute of Computer Science, Foundation for Research and Technology, Greece
martin@ics.forth.gr

Manolis Gergatsoulis
Department of Archive and Library Sciences, Ionian University, Greece
manolis@ionio.gr

Abstract

Metadata interoperability is an active research area, especially for cultural heritage collections, which consist of heterogeneous objects described by a variety of metadata schemas. In this paper we propose an ontology-based metadata interoperability approach, which exploits, in an optimal way, the semantics of metadata schemas. In particular, we propose the use of CIDOC/CRM ontology as a mediating schema and present a methodology for mapping DC Type Vocabulary to CIDOC/CRM, demonstrating a real-world effort for ontology-based metadata integration.

Keywords: ontology-based integration; metadata interoperability; CIDOC/CRM; DC-Type.

1. Introduction

Heterogeneity is one of the main characteristics of cultural heritage collections. Such collections may be composed of text, written on different materials, paintings, photographs, 3D objects, sound recordings, maps or even digital objects. Furthermore, the objects are strongly related with the social and historical events that take place over time. Consequently, it is quite justifiable to expose the composite structure, diverse semantics and multiple kinds of relationships between the objects of these collections. Hence, making accessible cultural heritage resources requires metadata schemas rich in semantics and structure able to cover the material heterogeneity and variety of memory institutions (libraries, archives and museums).

Taking into account the variety of cultural heritage metadata schemas, which are very often semantically related to each other, and the increasing demand by users for global access to highly distributed, heterogeneous, and dynamic collections, emphasis is given to matters of

** The research work of the authors I. Lourdi, L. Bountouri, and M. Gergatsoulis was partially co-funded by the European Social Fund (75%) and National Resources (25%) – Operational Program for Educational and Vocational Training (EPEAEK II) and particularly by the Research Program "PYTHAGORAS II". The research work of the authors C. Kakali, Th. Stasinopoulou, C. Papatheodorou and M. Doerr was supported by DELOS Network of Excellence on Digital Libraries funded by European Commission.

https://doi.org/10.23106/dcmi.952108714

interoperability and integration. Interoperability is required not only at the syntactic and system level but also at the more complex semantic level.

Data Integration has been a dynamic and challenging research area for many years to provide users with a uniform interface to access, relate, and combine data. Recently, research interests are moving from Data Integration to Semantic Integration in many communities and disciplines, such as e-government and cultural heritage. Semantic Integration is the process of using a conceptual representation of the data and of their relationships to eliminate possible heterogeneities (Cruz & Xiao, 2005). One of the main infrastructures of the Semantic Web, related to semantic integration issues, is ontologies (Sure & Studer, 2005). Ontologies provide a shared understanding of a domain of interest to support communication among human and computer agents, typically being represented in a machine-processable representation language. Ontologies offer solutions to the semantic heterogeneity problem (Wache et al., 2001) and can be used in integration architectures as a global schema to which metadata from different sources can be mapped. In comparison to other schemas, they conceptualize particular domains of interest and express their rich semantics.

Dublin Core element set (DCMI, 1998) is a flexible and usable metadata schema enabling information exchange and integration between digital sources. It is widely used by almost all digital libraries since it is simple, small and easily expandable, providing qualifiers enabling the semantic expression. The significant role of DC in data exchange is obvious by the fact that there are mappings from and to it by many widely used metadata schemas (Day, 2002). DCMI Type Vocabulary (DCMI, 2006a) identifies the genre of a resource. However, the semantics that DC elements express depend on the type of the described resource. For instance, the element DC.creator for the type "text" means the author or writer of a text but for the type "image" means the photographer or the painter. In an integration scenario through DC such information might be lost due to the simplicity of DC that cannot express all the required wealth of information semantics. Additionally, the plethora of DC application profiles makes even more difficult the interoperability for which DC has been created. The same elements express different meanings for different cases and types. Therefore, a mediator based on semantics resolution is necessary and the usage of ontologies is considered a suitable solution to this obstacle.

In this paper we present a methodology for mapping DCMI Type vocabulary to CIDOC/CRM. The mapping preserves the semantics of the DC records that correspond to different material types. CIDOC/CRM ontology (CIDOC, 2005) is a conceptual representation of the cultural heritage domain ensuring semantic integration between different cultural metadata schemas and eliminating their possible semantic heterogeneities. In particular, this paper presents the problems arisen when creating semantic mappings from DC metadata to ontological models. In detail, Section 2 presents a semantic integration architecture, which considers CIDOC/CRM as a mediator schema and the mapping of DC to CIDOC as part of this architecture. Based on this context, in Section 3, a methodology for obtaining the desired mapping is proposed, while in Section 4 and 5 the main results of the mapping process are presented. Finally, the paper closes in Section 6 discussing its conclusions.

2. Problem Definition and Related Work

2.1. Mapping Metadata Schemas to Ontologies

Mapping metadata schemas to ontologies is a complicated procedure, since these two constructs present many differences in various levels, as it is clearly explained in many papers (Sowa, 2000; NISO, 2004).

Scope and function: Metadata are used to describe resources in terms of elements, and to facilitate discovery and easy access to information. Ontologies define entities in an abstract level, with the intention of conceptualizing a domain of interest. They do not provide specific elements for the description of a resource, but a general overview of the basic notions of a field of interest and the relations between those notions.

https://doi.org/10.23106/dcmi.952108714

Expression of semantics: Metadata schemas are created for resources' identification and description and they do not express rich semantics. Even though the meaning of metadata information and its relationship to the described resource can be understood and processed by humans, for machine processing this relationship is not obvious, unless declared. In contrast, in ontologies classes are interconnected by specific properties that declare explicitly the semantic relationship between those entities. For example, in DC we write that "a specific poet is the creator of a poem" by assigning a value to DC.creator. In CIDOC we can express general statements about the creation of poems denoting that an Actor (poet) participates in a Creation Event, which produces a Linguistic Object (poem). In this way, the knowledge concerning the poem creation becomes explicit and machine understandable.

Mapping a real metadata schema to an ontology faces many difficulties due to the plethora of conceptual expressions that should be aligned. In our case, the combination of the CIDOC entities and properties generates a large number of conceptual expressions that should be studied in order to select the semantically closest of them to map the DC elements.

2.2. An Ontology-Based Mediator Architecture

In our approach, focus is given to the need to develop information systems able to provide access to heterogeneous data sources. For this purpose, a mediator able to semantically integrate the various schemas is proposed, considering CIDOC ontology as the global schema and defining mappings from the metadata schemas to CIDOC and vice versa.

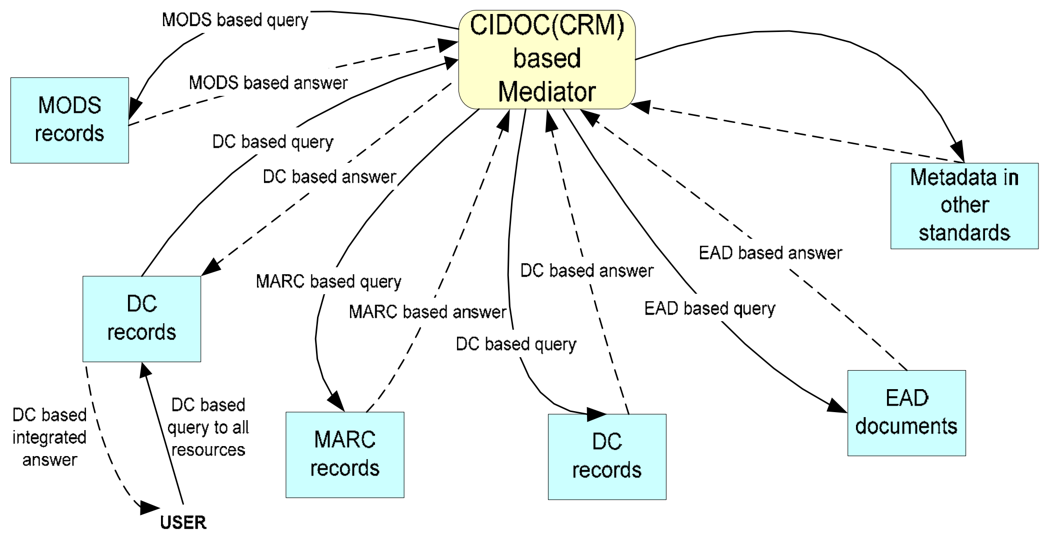

FIG. 1. An ontology-based mediator.

We selected CIDOC as the mediating schema because it is a core ontology designed to be applied to the documentation, integration, mediation and exchange of heterogeneous cultural information. It is a conceptual model, composed of *entities*, which are organized into a hierarchy and semantically related to each other with *properties*. In detail, CIDOC defines the complex interrelationships that exist between objects, actors, events, places and other concepts in the cultural heritage field.

The value of CIDOC/CRM becomes apparent when it is used as the basis for data transfer and exchange between different systems, schemas and semantics (Crofts, Doer & Gill, 2003). In such a scenario, CIDOC acts as a mediated schema to which different metadata can be mapped. Given that it is considered a core ontology, it allows gathering all the necessary cultural information in a

https://doi.org/10.23106/dcmi.952108714

suitable form for further reasoning (Doerr, 2003). FIG. 1 presents an architecture in which a set of data sources exist, each of them following a possibly different metadata schema. All these schemas are mapped to CIDOC. A user can pose his queries to a local data source following the restrictions of the local metadata schema. The local query engine returns the results from its source and promotes the query to CIDOC mediator, which translates the query to suitable forms, using the appropriate mappings, and forwards them to be answered by the other sources. Finally, the results from each source are collected and returned to the user. Note that the queries in the DC sources might be written in a query language such as SPARQL, while the queries in EAD sources might be in XQUERY.

For example, suppose that a user wishes to find metadata records (EAD finding aids, Dublin Core records, etc.) describing documents published by a person whose name is "John Smith". In terms of Dublin Core (DC), the user is looking for records for which DC.publisher="John Smith" and DC.type="text". Suppose that the user poses his (appropriately formed) query to a DC local data source. Then, the DC records from the local source matching the query are returned to the user. The query is then propagated to the ontology mediator where it is transformed, using a set of mapping rules from DC to CIDOC, into an equivalent query in terms of CIDOC/CRM ontology. In this query, the conditions (corresponding to the conditions of the initial query) locate the values that should be checked through appropriately formed CIDOC paths such as[††]: E33 (Linguistic Object) - *P94 (has created/was created)* - E65 (Creation Event) - *P14 (performed) [with subproperty P14.1 (in the role of)* - E55(Type) = "Publisher"] - E39(Actor) - *P131(is identified by/identifies)* - E82(Actor Appellation) = "John Smith". The CIDOC query is then transformed to other formats and propagated to the corresponding sources. For example, a local source keeping its data in EAD receives a query whose condition is applied on the value returned by the EAD path /ead/eadheader/filedesc/publicationstmt/publisher/name/ (Stasinopoulou et. al., 2007), which is compared with the string value "John Smith". If they match, the whole finding aid is returned to CIDOC and then to the user after being transformed into DC format.

Works related to ontology-based integration usually emphasize on element and structure level mappings and transformations (i.e. elements to classes, attributes to properties etc.). In (Cruz, Xiao & Hsu, 2004), authors map the XML data of every local source to an RDFS local ontology, created by transforming the XML elements and attributes to RDFS classes and properties. Additionally, the structure of an XML local source is encoded in the local RDFS ontology and then, local ontologies are merged to a global ontology for unified access and semantic integration of local data sources. Further, in (Lehti & Fankhauser, 2004), an XML data integration approach is presented based on the Web Ontology Language (OWL). More specifically, the proposed architecture maps XML structures (such as elements and attributes) to OWL structural components (such classes, properties, etc.) and thus they convert the XML data to an OWL global ontology.

In (Amann et al., 2001) the intention of the work is to propose a mechanism for the cultural information sources integration. The authors map pieces of information contained in XML fragments to domain specific ontologies, such as CIDOC, defining (1) a mapping language that describes the resources by a set of rules relating XPath location paths to the concepts and roles of an ontology, and (2) a query rewriting algorithm for translating user queries into queries expressed in an XML query language, which are sent for evaluation to XML sources. In (Tousa & Delgado, 2006) a model-mapping approach is applied to represent instances of XML and XML Schema in RDF. The architecture of Contorsion is described, which is a semantic XPath processor that acts over an RDF mapping of XML and is fed with an unlimited set of XML schemas and/or RDFS/OWL ontologies. Even though, all those approaches deal with semantic integration, they are strongly oriented to integrate XML data to RDFS and OWL ontology languages, giving emphasis to define structure mappings or model mappings between them.

[††] The notation Enn, *Pnn* corresponds to CIDOC entities and properties respectively.

https://doi.org/10.23106/dcmi.952108714

However, their effectiveness in mapping really complex semantically data structures, such as metadata schemas, has not yet been tested.

Nevertheless, there are some works referring to mapping data schemas to CIDOC/CRM, which try to enable the exchange and sharing of heterogeneous sources both within and between cultural institutions. A great effort analyses how to combine MPEG-7 with CIDOC entities into a single ontology for describing and managing multimedia in museums (Hunter, 2002). In this paper CIDOC is extended by MPEG-7 components, to add multimedia metadata capabilities. In (Doerr, 2000) a mapping of DC elements to CIDOC classes is presented. Our work extends and refines the mapping presented in this work.

3. Mapping Methodology

DC metadata element set contains an element DC.Type that identifies the genre of the described resource and provides a general, cross-domain list of proposed terms that may be used as values for this element. These are: *collection, dataset, event, image, moving image, physical object, sound, service* and *text*. Especially for collection, the Dublin Core Collection Description Working Group has developed an application profile called Dublin Core Collections Application Profile (DCMI, 2006b), which contains metadata terms drawn from Dublin Core as well as from other metadata vocabularies. The profile, which will be called for abbreviation reasons in the paper DCCAP, describes a collection as a separate digital or physical object.

Mapping one schema to another is the specification of a transformation of each instance of the source schema into a valid instance of the target schema. The first step in mapping DC.Types to CIDOC is to clarify the semantics of a DC record that corresponds to each DC.Type. Then we interpret DC paths to semantic equivalent CIDOC paths (Kondylakis, Doer & Plexousakis, 2006). A CIDOC path is defined as a chain of the form entity-property-entity, such that the entities associated by a property correspond to the property's domain and range. For example, consider the CIDOC path E19(Physical Object)-*P108(has produced)*-E12 (Production Event)-*P4(has time span)*-E52(Time Span). This path denotes that the result of a Production Event (E12), which took place in a specific time period (Time Span, E52), was a Physical Object (E19).

A DC path is created by linking the DC record, belonging to a material type, with a sequence of DC elements, element refinements, encoding schemes and vocabulary terms. For example, the path DC->DC.Date.Created denotes the creation date of a resource. Specifically, this path is part of the Qualified DC element set. Hence, we have to map the DC paths to CIDOC paths in a way that satisfies the semantic equivalence, taking into consideration the characteristics of each schema. It is impossible to map directly some elements to CIDOC entities due to the strictly event-aware character of the second. For instance, in order to map the creation date of a physical object (e.g. painting) to CIDOC terms, the whole DC record is mapped to CIDOC entity Physical Object (E19). The DC element Date could be mapped to the CIDOC entity Time Span (E52), but in CIDOC Time-Spans are used to define the temporal extent of instances of E4 Period, E5 Event and any other phenomena valid for a certain time. Thus, the entities Time Span (E52) and Physical Object (E19) could be interlinked only with the intermediation of an event or an activity, denoting that the event, which took place in a specific date, resulted in the described object. As a result, the semantic equivalent CIDOC path is: E19(Physical Object)-*P108(has produced)*-E12(Production Event)-*P4(has time span)*-E52(Time Span). To conclude, the main philosophy of our work is to view and correlate DC elements under the context of temporal entities. Therefore, in our mappings we apply intermediate activities and events in which the objects (documents, persons, places, structures, etc.) participate.

4. Results

Following our methodology, the types *physical object* and *collection* correspond to the CIDOC entity Physical Object (E19) and Collection (E78) respectively, the type *text* to the entity Linguistic Object (E33), while *image* corresponds to the entity Image (E38), *moving image* to

https://doi.org/10.23106/dcmi.952108714

Visual Item (E36), *service* to Design or Procedure (E29), and *dataset* to Document (E31). The type *Sound* corresponds to Linguistic Object (E33), if it contains speech, and to Information Object (E73) in every other case. The type *event* is mapped directly to the entity "Activity" (E7). The mappings of the DC elements, for every dc.type, to CIDOC paths are as follows:

DC->DC.Title. For the most material types, we can use the CIDOC property *has title/is title of* (P102) to denote the title of a DC record. For example, the title of a *text* is denoted: E33(Text)-*P102(has title/is title of)*-E35(Title). However, if the DC record corresponds to a *physical object* (E19) or an *event* (E7, Activity) the title is given by the property *is identified by/identifies* (P1). For instance, the corresponding CIDOC path of the type physical object is: E19 (Physical Object)-*P1(is identified by/identifies)*-E35(Title) (fig. 2). In case there is alternative title for the described item, the path DC->DC.Title. Alternative for all types results if we add after the entity Title (E35) the sub-path: *P139 (has alternative form)*-E41 (Appellation).

DC->DC.Creator. As mentioned, CIDOC correlates objects with persons performing particular roles through events. A person corresponds to the entity Actor (E39). For physical objects such as the types *physical object* and *collection* the entity Production Event (E12) is used, while for immaterial objects such as all other types - except the type *event* - the entity Creation Event (E65) is used. For example, the corresponding CIDOC path for the type *physical object* is: E19(Physical Object)-*P108(has produced/was produced by)*-E12(Production Event)-*P14(carried out by/performed[with subproperty P14.1 (in the role of)*-E55(Type)="creator"])-E39(Actor)-*P131(is identified by/identifies)*-E82(Actor Appellation) (FIG. 2), while for the type *image* the path is: E38 (Image)-*P94(has created/was created by)*-E65(Creation Event)-*P14(carried out by/performed[with subproperty P14.1 (in the role of)*-E55(Type)="photographer"])-E39(Actor)-*P131(is identified by/identifies)*-E82(Actor Appellation). Concerning the type *event* the mapping is different, since it denotes by itself an activity; therefore the path is: E7(Activity)-*P14(carried out by/performed)*-E39(Actor)-*P131(is identified by/identifies)*-E82(Actor Appellation). The sub-property *in the role of* (P14.1) allows the nature of an Actor's participation to be specified. Even though DCMI does not distinguish the use of roles for element Creator, in our approach the roles are applied to clarify the semantics of the element Creator. The mapping of the element Publisher is obtained in the same way by substituting the value of the entity Type(E55) with the value "publisher".

DC->DC.Date.Created. CIDOC correlates an object with its creation date, either through the entity Creation Event (E65), if the object is immaterial, or through the entity Production Event (E12) otherwise. For example if a DC record concerns a text, the corresponding CIDOC path is: E33(Linguistic Object)-*P94(has created/was created by)*-E65(Creation Event)-*P4(has time-span/is time-span of)*-E52(Time-Span). An exception to this rule is when the DC record is about an *event.* In this case the corresponding CIDOC path is: E7(Activity)-*P4(has Time-Span)*-E52(Time-Span).

DC->DC.Subject.Encoding scheme. This element Subject is mapped to the upper entity CRM-Entity (E1). For the types *text, sound, service* and *dataset* CIDOC provides the property *is about/is subject of* (P129) identifying the subject of an object. As a consequence the corresponding path is: E73(Information Object)-*P129(is about/is subject of)*-E1(CRM Entity)-*P70(documents/is documented to*)-E32(Authority Document)-*P71(lists/is listed in)*-E55 (Type: LC, DDC, LGSH, MESH, UDC). The sub-path E1(CRM Entity)-*P70(documents/is documented to*)-E32(Authority Document)-*P71(lists/is listed in)*-E55 (Type: LC, DDC, LGSH, MESH, UDC) corresponds to the mapping of the encoding schemes used for the subject encoding. For the types *physical object* and *collection,* the concept of subject is correlated with the Production Event of the object instead of the object itself, expressing that in reality the object was created for a purpose. Thus the corresponding CIDOC path is: E78(Collection)-*P108(has produced/was produced by)*- E12(Production Event)-*P17(was motivated by/motivated)*-E1(CRM Entity))-*P70(documents/is documented to*)-E32(Authority Document)-*P71(lists/is listed in)*-E55 (Type: LC, DDC, LGSH, MESH, UDC). The subjects for the types *image* and *moving image* are expressed with the property *represents/has representation* (P138) with domain the entities Image

https://doi.org/10.23106/dcmi.952108714

(E38) or Visual Item (E36) and range the entity CRM Entity (E1). Finally the path for the subject of the type *event* is*:* E7(Activity)-*P17was motivated by/motivated*-E1(CRM Entity).

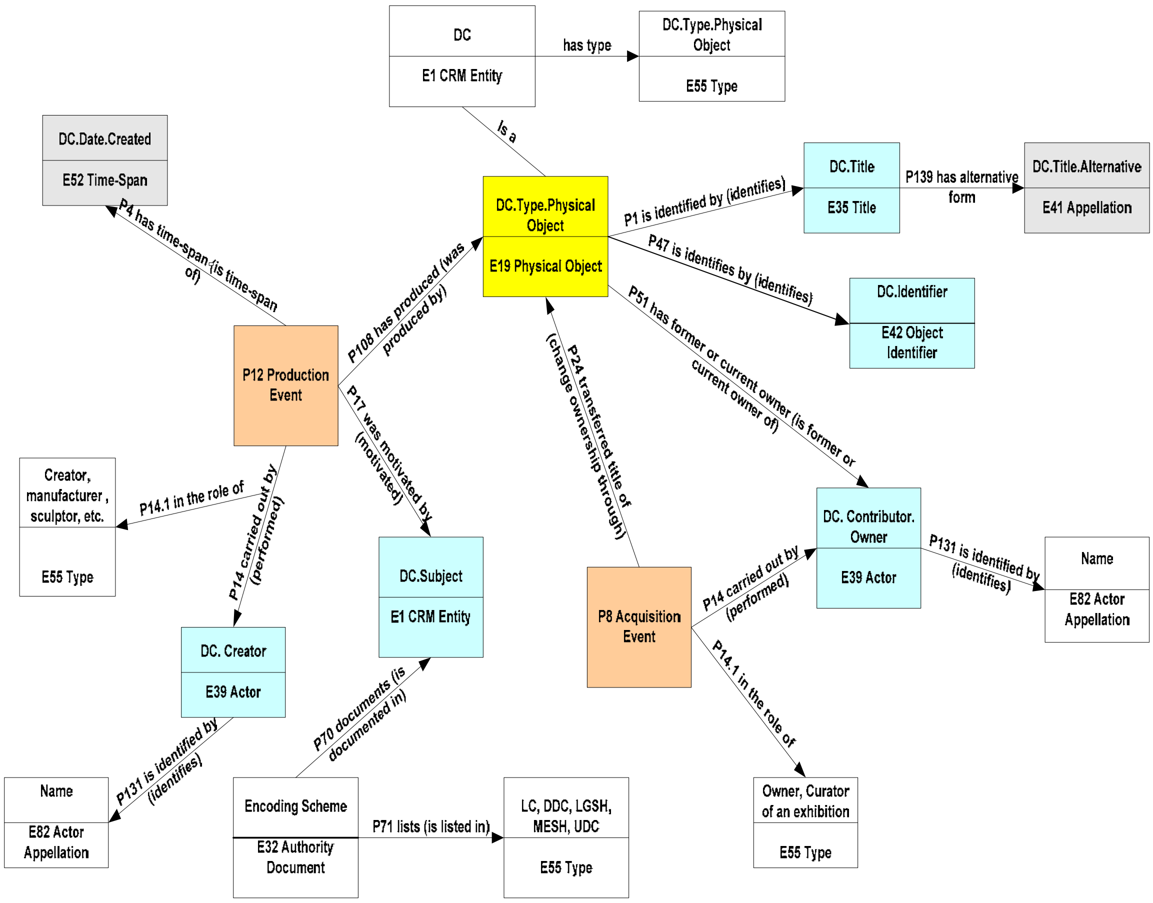

FIG. 2. An indicative part of the mapping for DC.type=Physical Object.

DC->DC.Contributor. When the value of the type is *physical object* the contributor is its Owner's name. When the type is *collection* the contributor is its "Curator", e.g. for *physical object* the CIDOC path is: E19(Physical Object)-*P51(has former or current owner/is former or current owner)*-E39(Actor). Moreover CIDOC provides the possibility to correlate the owner or the curator with the Acquisition Event (E8) of the object through the path: E19(Physical Object)-*P24(transferred title of/change ownership through)*-E8(Acquisition Event)-*P14(carried out by/performed[with subproperty P14.1 in the role*])-E55(Type), where the entity Type (E55) has the value "Owner" in the case of the type *physical object* and in the case of a *collection* the value "Curator". The mapping of the element Contributor for the other material types is similar with the mapping of the element Creator, using the subproperty *in the role of* (P14.1). For example, the corresponding path for the type *moving image* is: E36(Visual Item)-*P94(has created/was created by*-E65(Creation Event)-*P14(carried out byperformed [with subproperty P14.1 (in the role of)*-E55(Type)="contributor"])-E39(Actor). Finally, for the type *event* the contributor participates to the event and thus the corresponding CIDOC path is: E7(Activity) -*P11(had participant/participated in)*-E39(Actor).

DC->DC.Identifier. Identifiers in CIDOC depend on the objects' nature. Hence, for a physical object the corresponding path is E19(Physical Object)-*P47(is identified by/identifies)*-E42(Object Identifier). For the rest of the types the path differentiates to: E73(Information Object)-*P1(is identified by/identifies)*-E75(Conceptual Object Appellation)-*P2(has type)*-E55(Type). The

https://doi.org/10.23106/dcmi.952108714

subpath E75(Conceptual Object Appellation)-*P2(has type)*-E55(Type) specifies further the type of the identifier.

DC->DC.Rights. For all the types the corresponding CIDOC path is the same: E19(Physical Object)-*P104 is subject to (applies to)*-E30 Right. However for the type *event* there are no rights.

DC->DC.Source. For the types *physical object* and *collection* the element source has the meaning of the location. Therefore the corresponding CIDOC path is: E78(Collection)-*P55(has current location/holds)*-E53(Place) (FIG. 3). The remaining types concern digital images of other objects, which correspond to Physical Man-Made Stuff (E24). For instance, the source for a text is mapped as: E33(Linguistic Object)-*P67(refers/is referred to by)*-E24(Physical Man-Made Stuff).

DC->DC.Relation. Although there are many qualifiers for the Relation element, all of them are manipulated uniformly and independently on the material type. Due to space restrictions it is not possible to analyze all of them. Indicatively some mappings for the type *physical object* are presented (FIG. 3): (a) the path DC -> DC.Relation.IsVersionOf/HasVersion corresponds to: E19(Physical Object)-*P130(shows features of/features are also found on)*-E19 Physical Object. (b) DC -> DC.Relation.HasPart/isPartOf corresponds to: E19(Physical Object) -*P46(is composed of/forms part of)*-{E78(Collection) or E18(Physical Thing)} (c) DC -> DC.Relation.Replaces/IsReplacedBy corresponds to: E19(Physical Object)-*124(transformed/was transformed by)* -E81(Transformation) - *P123(resulted in/resulted from)*-E19(Physical Object).

DC->DC.Coverage. The notion of coverage is depicted to the upper entity CRM-Entity (E1) indicating a thematic relation of an object with a place or a time. For the types *collection* and *physical object* the mapping is similar, i.e.: E78(Collection)-*P62(depicts/is depicted)*-E1(CRM Entity). For *image* and *moving image* the property *represents/has representation (P138)* is used instead of P62, while for the rest of the types the property *is about/is subject of* (P129) is used. For the type *event* the refined term "spatial coverage" is strictly used, resulting in the path: E7(Activity)-*P7(took place at/witnessed)*-E53 (Place).

DC->DC.Description. The description of an object is a text and its mapping is the same for all types. For instance, the CIDOC path for an *image* is: E38(Image)-*P3(has note)*-E62(String).

DC->DC.Format. The qualifier Medium describes the material of an object. For the type *physical object* the corresponding CIDOC path is: E19(Physical Object)-*P45(consists of/is incorporated in)*-E57(Material), while for the type *event* the path is: E7(Activity)-*P125(used object of type/was type of object used in)*-E55(Type). All the other types are mapped to the same CIDOC path, composed by the entity corresponding to the specific DC.type followed by the pattern: *P2(has type)*-E55(Type), e.g. the path for *text* is: E33(Linguistic Object)-*P2(has type)*-E55(Type). The path DC->DC.Format.Extent concerns the dimensions of an object. For the type *physical object* the corresponding path is: E19(Physical Object)-*P43(has dimension/is dimension of)* E54 (Dimension)-*P91(has unit/is unit of)*-E58 (Measurement Unit). Moreover, the duration of the type *event* is mapped using the property *had at least duration/was minimum duration of* (P84).

DC -> DC.Language. The element language is applicable only for the type *text,* which is mapped to the entity Linguistic Object(E33) that comprises identifiable expressions in natural language or languages. As a result the corresponding CIDOC path is: E33(Linguistic Object)-*P76(has language/is language of)*-E56(Language).

https://doi.org/10.23106/dcmi.952108714

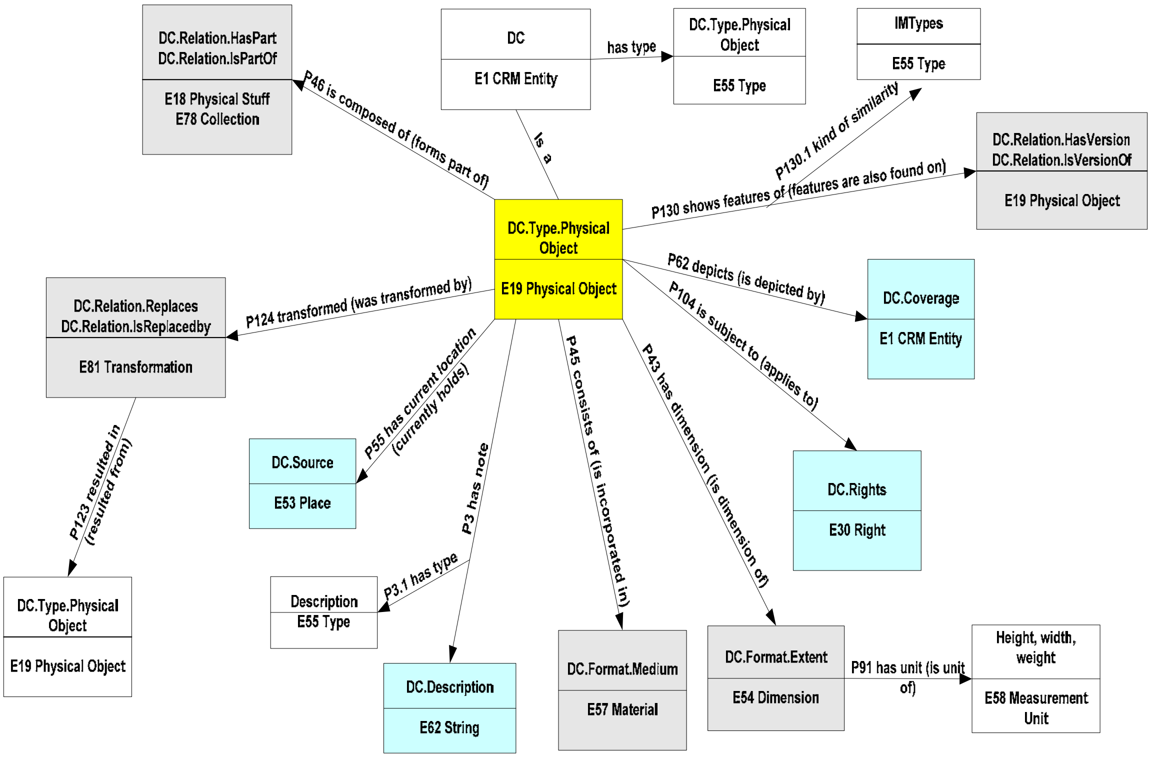

FIG. 3. The DC.relation mapping to CIDOC/CRM.

5. Mapping DCCAP to CIDOC

Dublin Core has developed an application profile for the description of collections (DCCAP). The records of DCCAP are semantically related with the DC records that correspond to the type *collection*. Therefore, the proposed methodology could be followed for the mapping of DCCAP to CIDOC (FIG. 4), demonstrating how CIDOC integrates the metadata for collection-level descriptions. Most of the DCCAP paths correspond to the same CIDOC paths described for the type *collection*, since they contain the same elements. Hence, emphasis is given to elements and concepts that they have not yet introduced.

The mapping of the DCCAP to CIDOC is as follows: The DCCAP is mapped to the CIDOC entity Collection (E78) and is interlinked through an IS_A relation with the DC.type.Collection. The collection development and management process is mapped to the entity Production Event (E12). The elements Subject and Audience are mapped to the entities CRM-Entity (E1) and Person (E21) respectively and thus the paths DCCAP-> Subject, and DCCAP->Audience correspond to the CIDOC paths E78 (Collection)-*P108* (*has produced/was produced by)*-E12 (Production Event)-*P17 (was motivated by/motivated)* - E1(CRM-Entity) and E78 (Collection)-*P108* (*has produced/was produced by)* -E12 (Production Event) -*P15 (was influenced by/influenced)* -E21 (Person) respectively (FIG. 4).

The element Collector/curator corresponds to the entity Actor (E39) and is semantically linked with the Production Event (E12). Hence the path DCCAP->Collector/curator corresponds to the following CIDOC paths, which are semantically interlinked with an IS_A relation E78(Collection)-*P109 (has current or former curator/is current or former curator)* -E39 (Actor) -*P14* (*carried out by/performed)* - E12 (Production Event) and E78 (Collection)-*P108* (*has produced/was produced by)* -E12 (Production Event) -E39 (Actor). The element Accumulation Date range corresponds to the entity Time Span (E52) and the path DCCAP->Accumulation Date

https://doi.org/10.23106/dcmi.952108714

range corresponds to the CIDOC path E78 (Collection) -*P108* (*has produced/was produced by*)- E12 (Production Event) -*P4 (has time span/is time span of)* -E52 (Time Span).

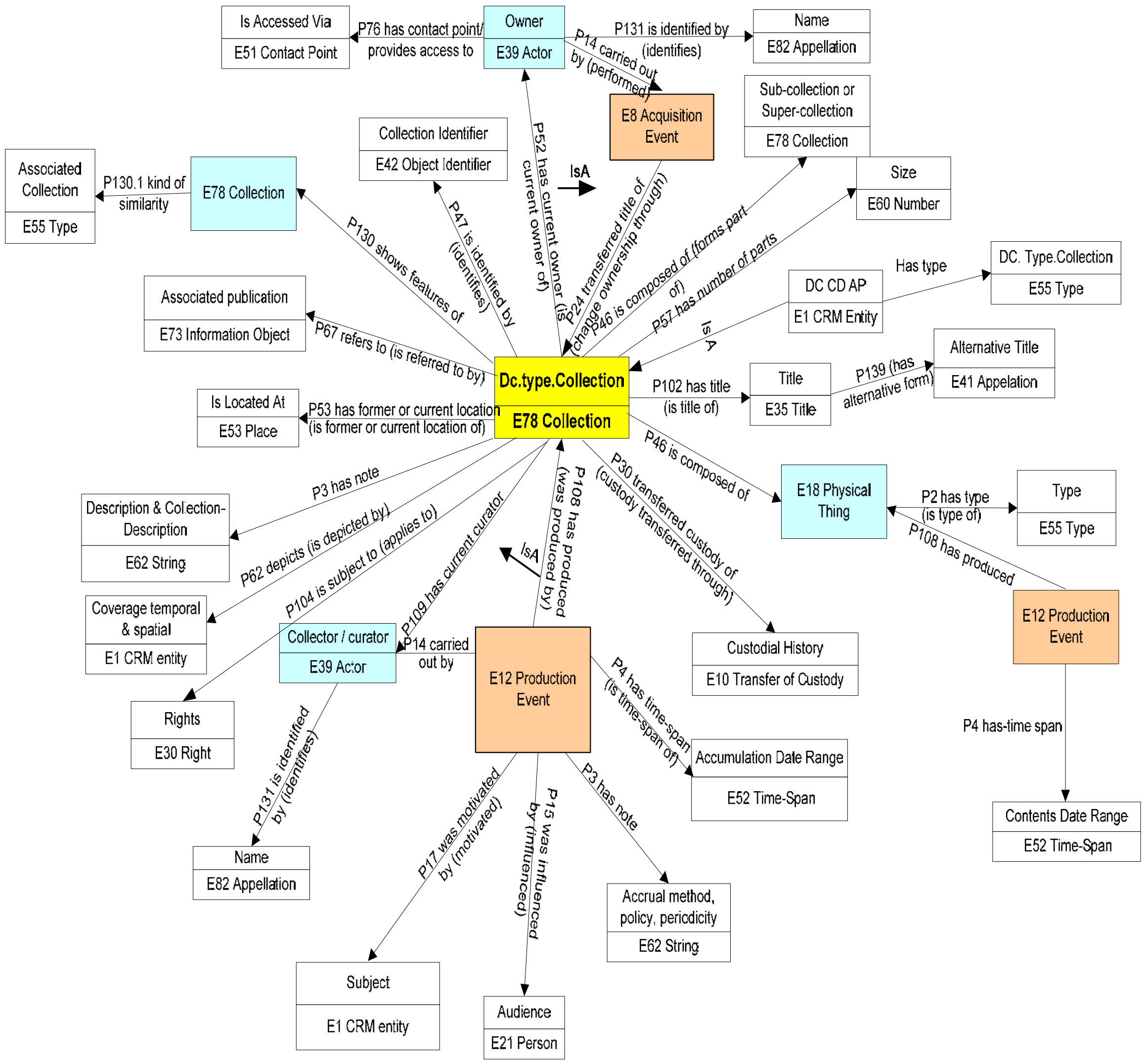

FIG. 4. The mapping of DCCAP to CIDOC/CRM.

A collection consists of particular objects usually of different types. Thus the DCCAP path DCCAP->Type, which describes "the nature or genre of the content of the resource" and depends on the types of material the collection contains, corresponds to the following CIDOC path: E78(Collection)-*P46(is composed of/forms part of)*-E18(Physical Thing)-*P2(has type/is type of)*-E55 (Type). Moreover, the path DCCAP->dateContentsCreated, that expresses the creation dates of the collection's objects, corresponds to the CIDOC path: E78(Collection)-*P46(is composed of/forms part of)*-E18(Physical Thing)-*P108(has produced/was produced by)*-E12(ProductionEvent)-*P4(has time span/is time span of)*-E52(Time-Span). The path DCCAP->Supercollection denoting the super (or the sub) collections of a collection is mapped to the following CIDOC path: E78(Collection)-*P46(is composed of/forms part of)*-E78(Collection), while the associated collections i.e. the path DCCAP->AssociatedCollection is mapped to the CIDOC path: E78(Collection)-*P130(shows features of/features are also found on)*-E78(Collection)- *P130.1(kind of similarity)*-E55(Type).

https://doi.org/10.23106/dcmi.952108714

The Custodial History (i.e the path DCCAP->CustodialHistory) is mapped to the CIDOC path: E78(Collection)-*P30(transferred custody of/custody transferred through)*-E10(Transfer of Custody). The path DCCAP->owner corresponds to the following CIDOC paths, which are semantically interlinked with an IS_A relation: E78(Collection)-*P52(has current owner/is current owner)*-E39(Actor) and E39(Actor)-*P14*(*carried out by/performed)*-E8(Acquisition Event)-*P24(transferred title of/changed ownership through)*- E78(Collection). Finally, the path DCCAP->isAccessedVia describes the service that provides access to the resource, which is the owner of the collection, i.e. a person. The corresponding CIDOC path is: E78(Collection)-*P52(has current owner/is current owner of)*-E39(Actor)-*P76(has contact point/provides access to)*-E51(Contact Point).

Concluding, CIDOC does not provide an entity denoting that a collection is the result of a specific development and management plan. Therefore we propose the addition of a new activity which may be called "Curation Activity (or "Accrual Activity"), and a new property called "*plans/is planned*" with domain the "Curation Activity" and range the entity Collection (E78).

6. Conclusion

Metadata semantic interoperability in the cultural heritage domain is one of the main issues in the digital environment. In our attempt to accomplish that goal, we proposed a semantic integration mechanism, so as to provide unified access to collections of heterogeneous material. In this context, we described an ontology-based integration architecture and addressed the issues of mapping metadata schemas to ontologies. What is more, we presented a significant part of the mapping from DC.Types, as well the mapping from the application profile for the collection description DCCAP to CIDOC/CRM ontology. The presented mapping was complex enough, given that CIDOC follows an event-based approach. Due to that fact, we had to make use of intermediate CIDOC activity and event entities to represent the relationships expressed in DC between objects and persons. In general, the proposed methodology reveals explicitly rich semantic correlations and it is consequently able to integrate a variety of application profiles focusing on the cultural heritage. Our future work is to define the inverse mapping from CIDOC to DC, and implement the DC-CIDOC-DC query engine.

References

Amann, Bernd, Irini Fundulaki, Michel Scholl, Catriel Beeri, and Anne-Marie Vercoustre. (2001). Mapping XML fragments to community Web ontologies. *Fourth International Workshop on the Web and Databases, WebDB 2001, Santa Barbara, California* (pp. 97-102).

CIDOC Documentation Standards Working Group, and CIDOC CRM SIG. (2005). *The CIDOC Conceptual Reference Model*. Retrieved March 10, 2007, from http://cidoc.ics.forth.gr/.

Crofts, Nick, Martin Doerr, and Tony Gill. (2003). The CIDOC Conceptual Reference Model: A standard for communicating cultural contents. *Cultivate Interactive, 9*. Retrieved March 10, 2007, from http://www.cultivate-int.org/issue9/chios/.

Cruz, Isabel. F., and Xiao Huiyong. (2005). The role of ontologies in data integration. *Journal of Engineering Intelligent Systems*, *13*(4). Retrieved March 10, 2007, from http://www.cs.uic.edu/~advis/publications/dataint.

Cruz, Isabel. F., Xiao Huiyong, and Hsu Feihong. (2004). An ontology-based framework for XML semantic integration. *Proceedings of the 8th International Database Engineering and Applications Symposium (IDEAS 2004), Portugal.*

Day, M. (2002). *Metadata: Mapping between metadata formats*. Retrieved June 11, 2007, from http://www.ukoln.ac.uk/metadata/interoperability/.

DCMI. (1998). *Dublin Core Metadata Element Set, version 1.0: Reference description*. Retrieved January 10, 2007, from http://www.dublincore.org/documents/1998/09/dces/.

DCMI. (2006a). *DCMI Type Vocabulary*. Retrieved March 10, 2007, from http://dublincore.org/documents/dcmi-type-vocabulary/.

DCMI. (2006b). *Dublin Core Collection Description Application Profile*. Retrieved January 10, 2007, from http://www.dublincore.org/groups/collections/collection-application-profile/.

https://doi.org/10.23106/dcmi.952108714

Doerr, Martin. (2000). *Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM* (Technical Report 274). Heraklion, Crete: Institute of Computer Science, Foundation of Research and Technology, Greece.

Doerr, Martin. (2003). The CIDOC CRM: An ontological approach to semantic interoperability of metadata. *AI Magazine*, *24,* 75. Retrieved March 10, 2007, from http://cidoc.ics.forth.gr/docs/ontological_approach.pdf.

Hunter, Jane. (2002). *Combining the CIDOC CRM and MPEG-7 to describe multimedia in museums*. Retrieved March 10, 2007, from http://metadata.net/harmony/MW2002_paper.pdf.

Kondylakis, Haridimos, Martin Doerr, and Dimitris Plexousakis. (2006). *Mapping language for information integration* (Technical Report 385). Heraklion, Crete: Institue of Computer Science, Foundation of Research and Technology, Greece.

Lehti, Patrick, and Peter Fankhauser. (2004). XML data integration with OWL: Experiences and challenges. *Proceedings of the SAINT* (pp. 160-170).

NISO. (2004). *Understanding metadata*. Retrieved June 11, 2007, from http://www.niso.org/standards/resources/UnderstandingMetadata.pdf.

Sowa, John F. (2000). Ontology, metadata, and semiotics. *Proceedings of the 8th International Conference on* Conceptual Structures *(ICCS 2000), Germany.*

Stasinopoulou, Thomais, Martin Doerr, Christos Papatheodorou, and Constantia Kakali. (2007). *WP5 - Task 5.5.: EAD mapping to CIDOC/CRM* (Report, 2007). DELOS-WP5 - Task 5.5 Ontology-driven Interoperability.

Sure, York, and Rudi Studer. (2005). *Semantic Web technologies for digital libraries*. Retrieved March 10, 2007, http://www.aifb.uni-karlsruhe.de/WBS/ysu/publications/2005_sw_for_dl.pdf.

Tousa, Rubén, and Jaime Delgado. (2006). Contorsion: A semantic XPath processor. *Proceedings of the International Workshop on Database Interoperability (InterDB 2005)* (pp. 87-102).

Wache, H., T. Voegele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Huebner. (2001). Ontology-based integration of information—A survey of existing approaches. *Proceedings of the IJCAI-01 Workshop on Ontologies and Information Sharing, Seattle, WA* (pp. 108-118).

https://doi.org/10.23106/dcmi.952108714