Mapping VRA Core 4.0 to the CIDOC CRM ontology*

Panorea Gaitanou and Manolis Gergatsoulis

Database & Information Systems Group (DBIS),
Laboratory on Digital Libraries and Electronic Publishing,
Department of Archives and Library Science, Ionian University,
Ioannou Theotoki 72, 49100 Corfu, Greece.
{rgaitanou, manolis}@ionio.gr

Abstract. In this paper, we present an effort to semantically map VRA Core 4.0, a cultural heritage metadata schema describing visual resources, to CIDOC CRM. This work is based on a semantic integration scenario, where CIDOC CRM acts as a mediation schema. More specifically, each element of the schema (along with its subelements and attributes) is mapped to the equivalent CRM path (represented as a sequence of classes and properties). The mapping is formally described using a Mapping Description Language (MDL), which explicitly defines semantic rules from the source schema to the target schema.

1 Introduction

Managing cultural heritage resources is a rather complex process, in which a range of sciences and scientists (computer scientists, information scientists, archives scientists, museologists, historians, etc.) are involved. Cultural heritage institutions are challenged to handle the information and knowledge dissemination in such a way that the needs and demands of various user groups are efficiently met. Within this framework, cultural heritage institutions (otherwise called "memory institutions") use various metadata schemas for the documentation of cultural collections, that facilitate access and retrieval to cultural information via the web. The complexity of the cultural information imposes the development of several different metadata standards (such as DCMI, VRA Core 4.0, EAD, Spectrum etc.), which exhibit significant diversity. This heterogeneity often results in data exchange failure, as the end user cannot access an integrated information system and retrieve the desired information. In order to address all the aforementioned issues and achieve a unified and standard-independent access to the relative information, it is necessary to integrate all these schemas. One

* This research has been co-financed by the European Union (European Social Fund ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.

of the most important and continuously evolving methods implemented in the interoperability domain is the ontology-based integration [7]. Ontologies provide the means for defining common vocabularies, representing the domain knowledge, while at the same time facilitating knowledge sharing and reuse among heterogeneous and distributed application systems.

The basic component of an information integration system is the mapping of the various metadata schemas to a schema or a core ontology, acting as a mediation schema, so that (meta)data integration is successfully accomplished. In the integration scenario proposed by our research group [9], the CIDOC CRM ontology is used as a mediation schema, to which different metadata schemas (such as MODS, DC, MARC, EAD etc.) are mapped.

In this paper, we present a mapping methodology of the VRA Core 4.0 schema to the CIDOC CRM ontology. This methodology, which is based on a path-oriented approach, is formally defined using a Mapping Description Language (MDL), which defines semantic rules from the source schema to the target schema. In the proposed methodology, each element of the schema (with its subelements) is represented as a VRA path (expressed in XPath form) and is then semantically translated to an equivalent path of classes and properties of CIDOC CRM. It is important to note that the mapping procedure focuses on the restricted version of the VRA Core 4.0, which imposes controlled vocabularies and type lists. Thus, each attribute assigned to an element/subelement of the metadata schema may generate different semantic paths on the ontology, depending on the value it takes each time, and produces a plethora of conceptual expressions for the same element/subelement. The use of several global attributes provided by the VRA Core 4.0 schema makes the mapping procedure even more complicated, by generating additional semantic paths on the ontology.

2 Mapping VRA Core 4.0 to CIDOC CRM

2.1 Brief description of the VRA Core 4.0

VRA Core 4.0 [10] is a metadata schema for the cultural heritage community, initially developed by the Visual Resources Association's Data Standards Committee. Currently, it is hosted by the Network Development and MARC Standards Office of the Library of Congress (LC) [5] in partnership with the Visual Resources Association. VRA Core 4.0 provides guidance on describing works of visual culture, collections, as well as images that document them. Therefore, it allows for three broad groups of entities, which are works, images, and collections. A work may represent a painting, sculpture or other artistic product. An image is a visual representation of a work that can come in a wide range of formats, and include various image formats (such as JPEG, GIF, TIFF) or could include physical photographs, slides, etc. Finally, a collection represents a group of works or images.

VRA Core 4.0 contains 19 elements (`work`, `agent`, `culturalContext`, `date`, `description`, `inscription`, `location`, `material`, `measurements`, `relation`, rights, source, stateEdition, stylePeriod, subject, technique, textref, title and worktype) and several optional global attributes (dataDate, extent, href, pref, refid, rules, source, vocab, xml:lang), which are applied additionally to any element or subelement, when necessary. Two XML Schema versions have been proposed for the VRA Core 4.0. An *unrestricted version*, which specifies the basic structure of the schema and imposes no restrictions on the values entered into any of the elements, sub-elements, or attributes, and a *restricted version*, which extends the unrestricted one by imposing controlled type lists and date formats.

Example 1. In this example we present a fragment of a simplified VRA document, describing a textual manuscript of the 18th century, taken from
http://www.vraweb.org/projects/vracore4/example017.html.

```
<?xml version="1.0" encoding="UTF-8" ?>
 <vra>
  <work id="w_4" source="Core 4 Sample Database (VCat)" refid="4">
   <agentSet>
    <agent>
     <name vocab="ULAN" refid="500017255"
           type="personal">Jefferson, Thomas</name>
     <dates type="life">
      <earliestDate>1743</earliestDate>
      <latestDate>1826</latestDate>
     </dates>
     <culture>American</culture>
     <role>author</role>
    </agent>
   </agentSet>
   <measurementsSet>
    <measurements type="height" unit="cm">75.56</measurements>
    <measurements type="width" unit="cm">62.23</measurements>
   </measurementsSet>
   <stylePeriodSet>
    <stylePeriod vocab="LCSAF"
                 refid="85041401">Eighteenth century</stylePeriod>
   </stylePeriodSet>
   <techniqueSet>
    <technique vocab="AAT" refid="300053162">calligraphy(process)</technique>
    <technique vocab="AAT" refid="300054698">writing(process)</technique>
   </techniqueSet>
   <titleSet>
    <title type="popular" xml:lang="en">Declaration of Independence</title>
   </titleSet>
   <worktypeSet>
    <worktype>manuscript (document genre)</worktype>
   </worktypeSet>
  </work>
 </vra>
```

2.2 The CIDOC CRM ontology

The *CIDOC Conceptual Reference Model* (CIDOC CRM) [3], which emerged from the CIDOC Documentation Standards Group in 1999, is a formal extensible ontology, which aims at providing a conceptual representation of cultural heritage domain, promoting semantic interoperability and integration. It is an object-oriented model comprised of a class hierarchy of 86 named classes interlinked by 137 named properties. CIDOC CRM defines the complex interrelationships between objects, actors, events, places, and other concepts used in the cultural heritage domain [2].

A *class* (also called *entity*), identified by a number preceded by the letter "E" (e.g. E1 CRM Entity, E2 Temporal Entity), groups items (called *class instances*) that share common characteristics. A class may be the *domain* or the *range* of *properties*, which are binary relations between classes. Properties are identified by numbers preceded by the letter "P" (e.g. P2 has type (is type of) with domain the class E1 CRM Entity and range the class E55 Type). A property can be interpreted in both directions (active and passive voice), with two distinct but related interpretations. A *subclass* is a class that specializes another class (its *superclass*). A class may have one or more immediate superclasses. When a class A is a subclass of a class B then all instances of A are also instances of B. A subclass inherits the properties of its superclasses without exception (*strict inheritance*) in addition to having none, one or more properties of its own. A *subproperty* is a property that specializes another property. A sample of CIDOC CRM properties is shown in Fig. 1.

Property Id & Name	**Entity - Domain**	**Entity - Range**
P1 is identified by (identifies)	E1 CRM Entity	E41 Appelation
P2 has type (is type of)	E1 CRM Entity	E55 Type
P4 has time-span (is time-span of)	E2 Temporal Entity	E52 Time-Span
P14 carried out by (performed)	E7 Activity	E39 Actor
P58 has section definition (defines section)	E18 Physical Thing	E46 Section Definition
P108 has produced (was produced by)	E12 Production	E24 Physical Man-Made Thing

Fig. 1. A sample of CIDOC CRM properties.

2.3 The Mapping Description Language (MDL)

The proposed mapping method between the metadata schemas and CIDOC CRM is based on a path-oriented approach. A mapping from a source schema to a target schema transforms each instance of the source schema into a valid instance of the target schema. Hence, we interpret the metadata paths to semantically equivalent CIDOC CRM paths. As we are interested in metadata schemas, which are based on XML, the paths in the source schemas are based on XPath [11], in fact they extend the XPath *location paths* with *variables* and stars (meaning *data transfer*). The syntax of the MDL *mapping rules* is given bellow in EBNF:

(R1) R ::= Left '−−' Right
(R2) Left ::= A_{Path} | V_{Path}
(R3) A_{Path} ::= ϵ | '/' R_{Path}
(R4) R_{Path} ::= L | L '*' | L '{' V_l '}' | L '*' '{' V_l '}'
(R5) V_{Path} ::= '$' V_l '/' R_{Path} | '$' V_l '{' V_l '}'
(R6) Right ::= E_t | E_e '→' O | '$' V_c '→' O | '$' V_p '→' P_p '→' E_{t55}
(R7) O ::= P_e '→' E_t | P_e '→' E_e '→' O
(R8) E_e ::= E | E '{' V_c '}'
(R9) E_t ::= E | E '{' V_c '}' | E '{=' String '}'
(R10) E_{t55} ::= E55 | E55 '{' V_c '}' | E55 '{=' String '}'
(R11) P_e ::= P | P '{' V_p '}'

The terminals used in these rules have the following semantics:

- L: it represents an XPath *location path*.
- V_l: it represents the *location variables*, which are used to declare the "branches" of the XML trees (XPath paths).
- V_c: it represents the *class variables*. The class variables are used to declare that a class can be the starting point of one or more CIDOC CRM paths.
- V_p: it represents the *property variables* The property variables are used to declare that a property can be the starting point of a new CIDOC CRM path, which - in this case - it is a property of a property linking the property that the variable represents to an instance of the `E55 Type` class.
- E: it represents the identifier of the class.
- E55: it represents the identifier of the class `E55 Type`.
- P: it represents the identifier of the property.
- P_p: it represents the identifier of the property of a property.
- *String*: it represents a string.

3 Mapping VRA Core 4.0 elements to equivalent CIDOC CRM paths

VRA Core 4.0 is an XML-based standard, therefore we use the XPath to locate VRA elements/attributes. A *VRA path* is a sequence of VRA elements and subelements, starting from the schema root element `vra` separated by the slash symbol (/). For instance, the path `/vra/work/titleSet/title` denotes the title of a work being described. A *CRM path* is defined as a chain in the form entity-property-entity, such that the entities associated with a property correspond to the property's domain and range. VRA Core defines three basic top elements: `work`, `collection` and `image`. In the context of a VRA Core 4.0 record, a work is defined as a physical entity that exists, existed in the past, or may exist in the future. It might be an artistic creation, such as a painting or a sculpture, a performance, a building or other construction, etc. Therefore, we associate each `work` element in a VRA document with an instance of the class `E24 Physical Man-Made Thing`, which comprises all persistent physical items that are purposely created by human activity.

In the following paragraphs, we present the mapping of the agent element of the VRA Core 4.0 schema to CIDOC CRM. The agent (including its subelements and attributes) is a representative element of VRA Core 4.0 and its mapping presents significant diversity and complexity. The methodology applied to this mapping can be used to map the other elements of the VRA Core 4.0 as well.

3.1 Mapping the agent element and its subelements

The agent element denotes a person, group or corporate body that has contributed to the production or creation of the work being described. It contains the following five subelements: name, culture, dates, role and attribution. Each one of them provides a part of the agent element:

- The name subelement specifies the names and appellations, assigned to an individual, group or corporate body. A type attribute is assigned to this subelement, with possible values personal, corporate, family, or other.
- The culture subelement refers to the nationality or culture of the person, group, or corporate body that participated to the work being described.
- The dates subelement, which contains two additional subelements, namely the earliestDate and the latestDate, refers to the dates associated with the agent. A type attribute is also assigned to this subelement (with possible values activity, life and other).
- The role subelement denotes the specific role of the individual, group or corporate.
- The attibution subelement defines a characteristic or a specific attribute related to the agent.

Mapping the agent element: It is easy to see, by examining the semantics of the CIDOC CRM classes, that the appropriate class of CIDOC CRM to map the agent element of VRA is the class E39 Actor. The instances of E39 Actor corresponding to each specific agent need to be related to the instance of E24 Physical Man-Made Thing representing the work being described, in order to express that an agent "contributed to the production or creation of the work being described". However, as CIDOC CRM is event-centric, it does not provide properties to directly relate the instances of these two classes. Instead, these instances can be related indirectly, though an event (instance of the class E12 Production) during which the object was created. In this way, the work being described (i.e. the instance of the class E24 Physical Man-Made Thing) is related through the property P108B was produced by to this event. Additionaly, this event should then be related to the instances of the class E39 Actor (representing the agent), through the property P14 carried out by. In this way, a CIDOC CRM path of the following form is created:

E24 Physical Man-Made Thing → P108B was produced by →
E12 Production → P14 carried out by → E39 Actor

which semantically corresponds to a VRA path of the form:

`/vra/work/agentSet/agent`

We should note that, in case there are more than one agents (i.e. more than one `agent` subelements of the element `agentSet`), different subpaths of the form:

→ `P14 carried out by` → `E39 Actor`

will be rooted to the (same) instance of `E12 Production` to relate it with the different agents (instances of `E39 Actor`) that took part in this production event.

Mapping the `name` subelement of `agent` element: The `name` subelement, which identifies the name of an agent, is mapped to an instance of the class `E82 Actor Appellation` and is linked to the corresponding instance of the class `E39 Actor` through the property `P131 is identified by`. In this way, the CIDOC CRM path, which semantically corresponds to the VRA path:

`/vra/work/agentSet/agent/name`

becomes:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` → `E39 Actor` →
`P131 is identified by` → `E82 Actor Appellation`

Mapping the `type` attribute of the `name` subelement: In VRA an attribute named `type` is assigned to the `name` element. This attribute is quite remarkable given that it determines if an agent is a person (when the value of `type` is `personal`), a corporate or an organization (when the value of `type` is `corporate`), a family (when the value of `type` is `family`), or none of the above (when the value of `type` is `other`). To map the attribute `type` in CIDOC CRM, we have investigated two different approaches:

First approach: A first approach to map the `type` attribute in CIDOC CRM is to employ the class `E55 Type` and link instances of this class (of the values `personal`, `corporate`, `family` or `other` respectively) to the corresponding instances of the class `E39 Actor` through the property `P2 has type`. In this case, the following CIDOC CRM path will be created:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` → `E39 Actor` [→ `P2 has type` → `E55 Type`] → `P131 is identified by` → `E82 Actor Appellation`

which semantically corresponds to the VRA path:

`/vra/work/agentSet/agent/name[@type]`

Notice that in this approach the value of the `type` attribute is given as value of the instance `E55 Type`.

The notation [...] in the CIDOC CRM path is used to denote that a new branch is rooted on the `E39 Actor` class node.

Second approach: A second approach to map the `type` attribute in CIDOC CRM is to refine the mapping of the specific agent by replacing the class `E39 Actor` with an appropriate subclass of this class determined by the value of the `type` attribute. More specifically, if the value of the `type` attribute is `personal`, then the corresponding agent can be considered to be an instance of the class `E21 Person`. In this case, the CIDOC CRM path becomes:

`E24 Physical Man-Made Thing` → `P108B was produced by` →

E12 Production → P14 carried out by → E21 Person →
P131 is identified by → E82 Actor Appellation
which semantically corresponds to the VRA path:
/vra/work/agentSet/agent/name[@type="personal"]

If the value of the type is corporate, the corresponding agent will be denoted as an instance of the class E40 Legal Body, while if the value is family then the corresponding agent will be considered as an instance of the class E74 Group.

Fig. 2 depicts the mapping of the agent element and its subelements when the type attribute has the value personal, while applying the second approach. In this figure, the upper part of each box indicates the VRA path mapped to the CIDOC CRM class shown in the lower part. The boxes are linked with arrows that represent CIDOC CRM properties, which appear as labels to these arrows. In case a property is used according to its inverse property name, it is characterized by the letter "B" as part of its name (e.g. P108B was produced by). The mapping of other subelements of the element agent, appearing also in Fig. 2 (that is the subelements culture, role and dates), will be presented in the following paragraphs. At this point, we should mention that the type attribute assigned to the name subelement exhibits a rather weak point of the VRA Core Schema, as it actually refers to the agent element to which we believe that is should have been assigned and not to the name subelement.

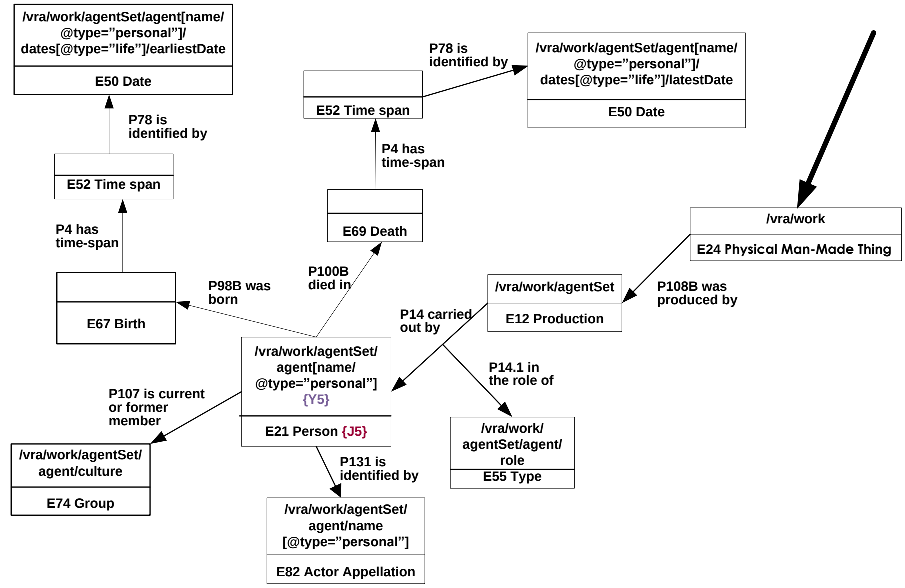

Fig. 2. The mapping of the value personal of the type attribute of the element agent.

Mapping the `role` subelement of `agent`: The `role` subelement, which identifies the role of an agent, is expressed in CIDOC CRM through the subproperty `P14.1 in the role of`, which actually links the property `P14 carried out by` to an instance of the class `E55 Type`. In this way, the CIDOC CRM path becomes:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` [→ `P14.1 in the role of` → `E55 Type`] → `E39 Actor`

which semantically corresponds to the VRA path:

`/vra/work/agentSet/agent/role`

Mapping the `culture` subelement of `agent`: The `culture` subelement, which identifies the nationality or culture of an agent, can be modelled as a membership of the agent to a group. This group is modelled in CIDOC CRM as an instance of the class `E74 Group`, which is related to the corresponding instance of the class `E39 Actor`, through the property `P107B is current or former member of`, resulting in the CIDOC CRM path of the form:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` → `E39 Actor` →
`P107B is current or former member of` → `E74 Group`

which semantically corresponds to the VRA path:

`/vra/work/agentSet/agent/culture`

Mapping the `dates` subelement of `agent`: The `dates` subelement is one of the most complex subelements to map, for three specific reasons: a) it contains a `type` attribute, with possible values `life`, `activity`, and `other`. Thus, it can define either the dates that span the known activity of an individual, group or corporate body, or the birth and death dates of a person (or even none of the above, by implementing the `other` attibute), b) it is strongly related to the `name` subelement, and more specifically to the value of the `type` attribute of the `name` subelement. For instance, if the `type` attribute of the subelement `name` is defined as `corporate`, then the value of the `type` attribute of the `dates` subelement can be either `activity` or `other`, denoting eg. the foundation dates of a corporate body, c) it contains two additional subelements, `earliestDate` and `latestDate`, which also define different semantic mappings. The following mapping of the `dates` refers to the case where the `type` attribute of the `name` subelement has the value `personal`, while the `type` attribute of the `dates` subelement gets the value `life`. The basic idea behind the mapping of the element `dates` (and its subelements) in this case is that the `earliestDate` subelement presents the birth date of an agent, while the `latestDate` subelement represents the date of his/her death.

Mapping the `earliestDate` subelement (when `@type="life"`): In order to map the `earliestDate` subelement, an instance of the class `E67 Birth` is created and related to an instance of the class `E21 Person`, through the property `P98B was born` (denoting the birth event of a person). Then, an instance of the class `E52 Time-Span` is linked to an instance of `E67 Birth`, through the property `P4 has` `time-span`, and finally in order to denote the specific date of the birth event, an instance of the class `E50 Date` is linked to an instance of `E52 Time-Span` through the property `P78 is identified by`. Thus, the following CIDOC CRM path:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` → `E21 Person` →
`P98B was born` → `E76 Birth` → `P4 has time-span` →
`E52 Time-Span` → `P78 is identified by` → `E50 Date`

semantically corresponds to:

`/vra/work/agentSet/agent/name[@type="personal"]`
`/dates[@type="life"]/earliestDate`

Mapping the ***`latestDate`*** *subelement (when* ***`@type="life"`****):* In order to map the `latestDate` subelement, an instance of the class `E69 Death` is created and related to an instance of the class `E21 Person`, through the property `P100B died in` (denoting the death event of a person). Then, adding as before the path → `P4 has time-span` → `E52 Time-Span` → `P78 is identified by` → `E50 Date`, we get the following CIDOC CRM path:

`E24 Physical Man-Made Thing` → `P108B was produced by` →
`E12 Production` → `P14 carried out by` → `E21 Person` →
`P100B died in` → `E69 Death` → `P4 has time-span` →
`E52 Time-Span` → `P78 is identified by` → `E50 Date`

which semantically corresponds to the VRA path:

`/vra/work/agentSet/agent/name[@type="personal"]`
`/dates[@type="life"]/latestDate`

The mappings presented in this section are also shown in Fig. 2.

3.2 The mapping of the `agent` element expressed in MDL

MDL can be used to formally describe the mapping rules of the elements/ attributes of a source schema to equivalent paths of the target schema. Part of the mapping, containing the rules that map the VRA element `agent` and its subelements/attributes, is shown in Table 1, expressed in MDL. In this section, a brief analysis of the rules' semantics is presented. For example, Rule R1 states that the `/vra/work` is mapped to an instance of the class `E24`. R2 states that the `agentSet` corresponds to an instance of the class `E12`, which is linked to `E24` through the binary relation `P108B`. Rules R3, R4, R5 describe the three different versions of the `agent` element, according to the three possible values of the `type` attribute of the `name` subelement, which correspond to the three different subclasses (`E21`, `E40`, `E74`), respectively. It is also important to note here that the variables Y5, Y10 and Y15 on the left part of the rules, as well as the variables J5, J10 and J15 on the right part, denote branching points, that indicate that more than one paths may extend the previous paths (see also Fig. 2). Rules R6, R7, R8, R9 and R10 can be appended to the Rule R3.

RuleNo	VRA paths	CIDOC CRM paths
R1:	/vra/work{X1}	E24{C1}
R2:	$X1/agentSet{Y1}	$C1→P108B→E12{J1}
R3:	$Y1/agent[name/@type="personal"]{Y5}	$J1→P14{S2} →E21{J5}
R4:	$Y1/agent[name/@type="corporate"]{Y10}	$J1→P14{S3} →E40{J10}
R5:	$Y1/agent[name/@type="family"]{Y15}	$J1→P14{S4} →E74{J15}
R6:	$Y5\|$Y10|$Y15/name*	$J5\|$J10|$J15→P131→E82
R7:	$Y5\|$Y10|$Y15/culture*	$J5\|$J10|$J15→P107→E74
R8:	$Y5\|$Y10|$Y15/role*	$S2\|$S3|$S4→P14.1→E55
R9:	$Y5/dates[@type="life"]/earliestDate*	$J5→P98→E67→P4→E52→ P78→E50
R10:	$Y5/dates[@type="life"]/latestDate*	$J5→P100B→E69→P4→E52→ P78→E50

Table 1. Mapping the VRA element `agent` to the CIDOC CRM using MDL.

4 Related work

There is quite an amount of research dealing with ontology-based integration. Amann et al. [1] propose a mechanism for the integration of cultural information resources, by mapping XML fragments to domain specific ontologies, such as CIDOC CRM. In this way, they define a mapping language, which provides a set of rules that describe these resources, relating XPath location paths to the concepts and roles of an ontology. Furthermore, they define a query rewriting algorithm which translates queries executed by users into queries expressed in an XML language and are afterwards sent to XML resources for evaluation. This approach is worth mentioning as it describes a mapping language quite similar to ours and also focuses on the significance of offering mechanisms for representing the semantics of XML data. In [4] XML data are transformed to a global ontology (using the OWL syntax), defining mapping rules that are also based in OWL. In this way, issues of synonymy and structure hierarchy are faced. This work shares common ideas with ours, as it transforms data to a global ontology, although the mapping rules defined in our MDL are not based in OWL syntax.

In [6], an effort is described to integrate the CIDOC CRM ontology in the core model of the BRICKS project. This integration has been accomplished through a mapping scenario applied between the source schemas and the CRM ontology, although a number of issues had to be resolved. Some of them refer to inconsistencies, which mostly originate from the abstractness of some concepts definitions of the CRM [8]. This approach provides mappings that are implemented in spreadsheets, without defining a formal mapping methodology.

5 Conclusions

The mapping methodology presented in this paper is part of an ontology-based metadata integration scenario, where CIDOC CRM acts as a mediating schema among several metadata schemas. More specifically, a semantic mapping from the VRA Core 4.0 standard to the CIDOC CRM ontology is presented.

Mapping VRA elements to CIDOC CRM paths proved to be a rather difficult and time-consuming activity, which required a deep and conceptual work. CIDOC CRM provides very rich structuring mechanisms for metadata descriptions and an abstract but fine-grained conceptualization for events, objects, agents, things, etc. Thus, the combination of this wide range of CRM classes and properties generated a large number of conceptual expressions that should be studied very carefully in order to select the semantically closest one to map to the metadata schemas. Furthermore, the mapping procedure encountered significant obstacles due to the plethora of conceptual expressions that should be aligned. The `type` attribute assigned to several subelements defined different semantic mappings, making mapping even more complex. Finally, it is essentiall to note that the `agent` element and all the related information to the work's production, include the class `E12 Production`, which reveals one of the main characteristics of CIDOC CRM, which is the event-based approach adopted.

Currently, we are investigating the transformation of queries among various cultural heritage metadata schemas and the CIDOC CRM ontology. Our next research steps include the definition of the reverse semantic mappings from the ontology to the VRA Core schema, in order to enrich the mapping procedure proposed by our research group.

References

1. B. Amann, C. Beeri, I. Fundulaki, and M. Scholl. Ontology-Based Integration of XML Web Resources. In *ISWC 2002*, volume 2342 of *LNCS*, pages 117–131. Springer, 2002.
2. C. Kakali and I. Lourdi and T. Stasinopoulou and L. Bountouri and C. Papatheodorou and M. Doerr and M. Gergatsoulis. Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. In *DC-2007*, pages 128–139, 2007.
3. ICOM/CIDOC CRM Special interest Group. Definition of the CIDOC Conceptual Reference Model, Version 5.0.2, January 2010. Available at http://www.cidoc-crm.org.
4. P. Lehti and P. Fankhauser. XML Data Integration with OWL: Experiences and Challenges. In *SAINT 2004*, pages 160–170. IEEE Computer Society, 2004.
5. Library of Congress (LC). VRA Core: a Data Standard for the Description of Works of Visual Culture, 2011. Available at http://www.loc.gov/standards/vracore/.
6. C. Meghini and T. Risse. BRICKS: A Digital Library Management System for Cultural Heritage. *ERCIM News*, (61), April 2005.
7. N. F. Noy. Semantic Integration: a Survey of Ontology-Based Approaches. In *SIGMOD Record*, volume 33, 2004.
8. P. Nussbaumer and B. Haslhofer. CIDOC CRM in Action - Experiences and Challenges. In *ECDL 2007*, volume 4675 of *LNCS*, pages 532–533. Springer, 2007.
9. T. Stasinopoulou and L. Bountouri and I. Lourdi and C. Papatheodorou and M. Doerr and M. Gergatsoulis. Ontology-Based Metadata Integration in the Cultural Heritage Domain. In *ICADL 2007*, volume 4822 of *LNCS*, pages 165–175. Springer, 2007.

10. VRA Core Oversight Committee. VRA Core 4.0 Element Description and Tagging Examples, 2007. Available at http://www.loc.gov/standards/vracore/schemas.html.
11. World Wide Web Consortium (W3C). XML Path Language (XPath) 2.0. Available at http://www.w3.org/TR/xpath20/, 2007.