1<sup>st</sup> Workshop on Digital Information Management March 30-31, 2011

# Mapping VRA Core 4.0 to the CIDOC/CRM ontology

### Panorea Gaitanou, Manolis Gergatsoulis

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





## **Presentation Overview**

- Cultural heritage domain
- Interoperability issues
- Ontology-based integration
- Our integration scenario
- Mapping VRA Core 4 to the CIDOC/CRM ontology
  - Brief presentation of VRA Core 4.0 and CIDOC/CRM
  - Mapping Description Language (MDL)
  - Mapping VRA elements to CIDOC/CRM paths
- Conclusions Future directions
- References



## **Cultural heritage domain**

With the **advent of new digital information technologies**, access to cultural heritage repositories is obviously improved.

The documentation, management and preservation of cultural artifacts and collections still seems to be rather a complex task.

Even in homogeneous environments (relational databases etc), there are problems in the exchange of information, because of different structures and syntaxes used.

In **cultural heritage collections**, the **problem of heterogeneity** is even more complex, as these collections contain **multiple material**.



## **Interoperability issues**

The continuous development of various metadata schemas leads to the need to integrate all these standards, in such a way that data and information exchange and sharing can be fully achieved.

#### Cultural heritage communities explore several methodologies:

 $\rightarrow$  to reduce search time

 $\rightarrow$  to guarantee unified access to these resources (under a global schema)





"An ontology is a formal, explicit specification of a shared conceptualization". (Gruber, 1993)

#### **Ontologies**

 $\rightarrow$  can play a leading role in this area, as they are considered to be an important building block for integration architectures.

 $\rightarrow$  provide the means for defining common vocabularies, representing the domain knowledge, facilitating at the same time knowledge sharing and reuse among heterogeneous and distributed application systems.

 $\rightarrow$  can act as a mediating schema that semantically integrates all meta(data)









## VRA Core 4.0 (1/2)

Visual Resources Association's Data Standards Committee + Network Development and MARC Standards Office of the Library of Congress (LC)

Provides guidance on describing works of visual culture, collections, as well as images that document them.

#### **Two XML Schema versions**

 $\rightarrow$  Unrestricted version (specifies the basic structure and imposes no restrictions on the values entered into any of the elements, sub-elements, or attributes)

→ **Restricted version** (extends the unrestricted one by imposing controlled type lists and date formats)



## VRA Core 4.0 (2/2)

**Record types**: Work, Image, Collection

#### 19 elements (with their subelements)

(*agent*, culturalContext, date, description, *inscription*, location, material, measurements, relation, rights, source, stateEdition, stylePeriod, subject, technique, textref, title, worktype)

#### global attributes

(dataDate, extent, href, pref, refid, rules, source, vocab, xml:lang)



### **A VRA record in XML**

```
<?xml version="1.0" encoding="UTF-8" ?>
```

```
<vra>
```

```
<work id="w_000777" refid="000597" source="History of Art Visual Resources Collection, UCB">
```

- <agentSet>
  - <agent>
    - <name type="personal">Cropsey, Jasper Francis</name>
  - <dates type="life">
  - <earliestDate>1823</earliestDate>
  - <latestDate>1900</latestDate>
  - </dates>
  - </agent>
- </agentSet>
- <dateSet>
- <date type="creation">
- <earliestDate>1860</earliestDate>
- <latestDate>1860</latestDate>
- </date>
- </dateSet>
- <inscriptionSet>
- <inscription>
- <author vocab="ULAN" refid="500012491">Cropsey, Jasper Francis</author>
- <position>lower center</position>
- <text>Autumn-on the Hudson River/J. F Cropsey/London 1860</text>
- </inscription>
- </inscriptionSet>





## CIDOC Conceptual Reference Model (CIDOC/CRM, Version 5.0.2)

#### **CIDOC Documentation Standards Group (1999)**

 $\rightarrow$  formal extensible ontology, which aims at providing a conceptual representation of cultural heritage domain, promoting semantic interoperability and integration.

 $\rightarrow$  object-oriented model composed of entities, organized into a hierarchy, and semantically related to each other properties.

 $\rightarrow$  targets to *cover contextual information*: the historical, geographical, and theoretical background in which individual items are placed and which gives them much of their significance and value

 $\rightarrow$  defines the *complex interrelationships* between objects, actors, events, places, and other concepts used in the cultural heritage domain

→ consists of a hierarchy of *86 classes* and *137 properties*.



# A sample of CIDOC/CRM properties

Property Id and Name	Entity - Domain	Entity - Range
P1 is identified by (identifies)	E1 CRM Entity	E61 Appellation
P2 has type (is type of)	E1 CRM Entity	<b>Е55</b> Туре
P4 has time-span (is time span of)	E2 Temporal Entity	E52 Time-Span
<b>P14</b> carried out by (performed)	E7 Activity	E39 Actor
<b>P53</b> has former or current location (is former or current location of)	E18 Physical Thing	E53 Place
P108 has produced (was produced by)	E12 Production	<b>E24</b> Physical Man-Made Thing



## Mapping methodology

#### 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.

VRA paths based on XPath  $\rightarrow$  equivalent CIDOC/CRM paths.

- <u>VRA path</u>: a sequence of VRA elements and subelements, starting from the schema root element <vra> separated by the slash symbol (/).
- <u>CIDOC/CRM path</u>: a chain in the form entity-property-entity (e-p-e), such that the entities associated by a property correspond to the property's domain and range.



# **Mapping Description Language (MDL)**

### The syntax of MDL (in EBNF notation) is:



# "Agent" element (work level) (1/7)

#### Element <agent>

Subelements:

<name>

<u>Attribute</u>: *type* 

<u>Values</u>: personal, corporate, family, other

<dates>

<u>Attribute</u>: *type* 

<u>Values</u>: *life, activity, other* 

Subelements:

<earliestDate>

<latestDate>

<culture>

<role>



# "Agent" element (work level) (2/7)

#### VRA: work/agentSet/agent

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor

#### VRA: work/agentSet/agent/name[@type]

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type



# "Agent" element (work level) (3/7)

### VRA: work/agentSet/agent/name[@type]

- Possible values for "type" attribute= personal, corporate, family, other.
- The value of the "type" attribute of the subelement "name" in the VRA paths defines different equivalent CIDOC/CRM paths.



### "Agent" element (work level) (4/7)

#### VRA: work/agentSet/agent/name[@type="personal"]

1) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type (instance="personal")

<u>Alternatively</u>,

2) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E21 Person -> P131 is identified by -> E82 Actor Appellation

#### VRA: work/agentSet/agent/name[@type="corporate"]

1) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type (instance="corporate")

Alternatively,

2) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E40 Legal Body -> P131 is identified by -> E82 Actor Appellation

#### VRA: work/agentSet/agent/name[@type="family"]

1) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type (instance="family")

Alternatively,

2) <u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E74 Group -> P131 is identified by -> E82 Actor Appellation



### "Agent" element (work level) (5/7)

### VRA: work/agentSet/agent[name[@type="personal"]]/dates[@type]

 Possible values for the "type" attribute of the "dates" subelement: life, activity, other



### "Agent" element (work level) (6/7)

#### <u>VRA</u>:

#### work/agentSet/agent[name[@type="personal"]]/dates[@type="life"]/earliest Date

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E21 Person -> P98B was born -> E67 Birth -> P4 has time-span -> E52 Time-Span -> P78 is identified by -> E50 Date

#### <u>VRA</u>:

#### work/agentSet/agent[name[@type="personal"]]/dates[@type="life"]/latestD ate

<u>CIDOC/CRM</u>: 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



### "Agent" element (work level) (7/7)

#### VRA: work/agentSet/agent/culture

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E21 Person -> P107B is current or former member of -> E74 Group

#### VRA: work/agentSet/agent/role

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing –> P108B was produced by –> E12 Production –> P14 carried out by [P14.1 in the role of –> E55 Type] –> E21 Person



### "Agent" element (work level) (type=personal)



### "Agent" element (work level) (type=corporate)

#### <u>VRA</u>:

# work/agentSet/agent[name[@type="corporate"]]/dates[@type="activity"]/ earliestDate

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E40 Legal Body -> P92B was brought into existence by -> E63 Beginning of Existence -> P4 has time-span -> E52 Time-Span -> P78 is identified by -> E50 Date

#### <u>VRA</u>:

#### work/agentSet/agent[name[@type="corporate"]]/dates[@type="activity"]/ latestDate

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E40 Legal Body -> P93B was taken out of existence by -> E64 End of Existence -> P4 has time-span -> E52 Time-Span -> P78 is identified by -> E50 Date



### "Agent" element (work level) (type=corporate)



## **MDL rules for the "agent" element**

Rule No	XPath location paths	CIDOC/CRM paths
R1	Work {X1}	E24 <b>{C1}</b>
R2	<pre>\$X1/agentSet {R1}</pre>	<pre>\$C1 -&gt; P108B -&gt; E12 {J1}</pre>
R3	<pre>\$R1/agent[name/@type = "personal"] {R5}</pre>	<b>\$J1</b> -> P14 { <b>S2</b> } -> E21 { <b>J5</b> }
R4	<pre>\$R1/agent[name/@type = "corporate"] {R10}</pre>	<b>\$J1</b> -> P14 <b>{S3}</b> -> E40 <b>{J10}</b>
R5	<pre>\$R1/agent[name/@type = "family"] {R15}</pre>	<b>\$J1</b> -> P14 <b>{S4}</b> -> E74 <b>{J15}</b>
R6	\$R5 \$R10 \$R15/name*	<b>\$J5 \$J10 \$J15</b> -> P131 -> E82
R7	<pre>\$R5 \$R10 \$R15/culture*</pre>	<b>\$J5 \$J10 \$J15</b> -> P107 -> E74
R8	\$R5 \$R10 \$R15/role*	\$S2 \$S3 \$S4 -> P14.1 -> E55
R9	<pre>\$R5/dates[@type= "life"] {T1}</pre>	
R10	<pre>\$T1/earliestDate*</pre>	<b>\$J5</b> -> P98 -> E67 -> P4 -> E52 -> P78 -> E50
R11	<pre>\$T1/latestDate*</pre>	<b>\$J5</b> -> P100B -> E69 -> P4 -> E52 -> P78 -> E50
R12	<pre>\$R10/dates[@type="activity"] {T2}</pre>	
R13	\$T2/earliestDate*	<b>\$J10</b> -> P92B -> E63 -> P4 -> E52 -> P78 -> E50
R14	<pre>\$T2/latestDate*</pre>	<b>\$J10</b> -> P93B -> E64 -> P4 -> E52 -> P78 -> E50

# "Inscription" element (work level)

Element <inscription>

Subelements:

<author>

<position>

<text>

Attribute: type

<u>Values</u>: *signature, mark, caption, date, text, translation, other* 



## "Inscription" element (work level)



### **MDL rules for the "inscription" element**

Rule No	XPath location paths	CIDOC/CRM paths
R1	Work <b>{X1}</b>	E24 <b>{C1}</b>
R2	<pre>\$X1/inscriptionSet/inscription {Y1}</pre>	<b>\$C1</b> -> P128 -> E37 <b>{D1</b> }
R3	\$Y1/author*	<b>\$D1</b> -> P94B-> E65 -> P14 -> E39 -> P131-> E82
R4	<b>\$Y1</b> /text* <b>{Y3}</b>	<b>\$D1</b> -> P138 -> E33 <b>{D3}</b>
R5	\$Y3/@type*	<b>\$D3</b> -> P2 -> E55
R6	<b>\$Y3</b> /@xml:lang*	<b>\$D3</b> -> P72 -> E56
R7	<pre>\$Y1/position*</pre>	<b>\$D1</b> -> P58 -> E46



### **MDL Rules for the "material, measurements,**

### title, rights" elements

Rule No	XPath location paths	CIDOC/CRM paths
R1	Work {X1}	E24 {C1}
R2	\$X1/materialSet/material* {Z1}	\$C1 -> P45 -> E57 {A1}
R3	\$Z1/@type*	\$A1 -> P2 -> E55
R4	\$X1/measurementsSet/measurements* {W1}	\$C1 ->P43 -> E54 {B1} -> P90 -> E60
R5	\$W1/@type*	\$B1 -> P2 -> E55
R6	\$W1/@unit*	\$B1 -> P91 -> E58
R7	\$X1/titleSet/title*{Q1}	\$C1 -> P102 {S1} -> E35
R8	\$Q1/@type*	\$S1 -> P102.1 -> E55
R9	\$X10/rightsSet/rights {F1}	\$C1 -> P104 -> E30 {G1}
R10	\$F1/@type*	\$G1 -> P2 -> E55
R11	\$F1/rightsHolder*	\$G1 -> P75B -> E39
R12	\$F1/notes*	\$G1 ->P3 -> E62
R13	\$F1/text*	\$G1 -> P1 -> E75

## "Agent" element (image level)

#### VRA: work/agentSet/agent/name[@type]

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P108B was produced by -> E12 Production -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type

#### VRA: image/agentSet/agent/name[@type]

<u>CIDOC/CRM</u>: E38 Image -> P94B was created by -> E65 Creation -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation -> P2 has type -> E55 Type

In the image level, all the subelements of the element "agent" are represented accordingly to the mappings of the element "agent" in the work level, with the difference of : E38 Image -> P94B was created by -> E65 Creation ...



# "Inscription" element (image level)

#### <u>VRA</u>: work/inscriptionSet/inscription/author

<u>CIDOC/CRM</u>: E24 Physical Man-Made Thing -> P128 carries -> E37 Mark -> P94B was created by -> E65 Creation -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation

#### VRA: image/inscriptionSet/inscription/author

<u>CIDOC/CRM</u>: E38 Image -> P128 carries -> E37 Mark -> P94B was created by -> E65 Creation -> P14 carried out by -> E39 Actor -> P131 is identified by -> E82 Actor Appellation



### The use of MDL

- Metadata exchange from the VRA paths to equivalent CIDOC/CRM paths
- Transformation of XPath queries posed on XML-based metadata into equivalent queries on the CIDOC/CRM ontology



## **Conclusions**

 Mapping VRA elements to CIDOC/CRM paths proved to be a rather timeconsuming activity, which requires a deep 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.
- the combination of a wide range of classes and properties that generates a large number of conceptual expressions that need to be examined thoroughly (classes and properties hierarchies).
- The use of the classes <u>E12 Production</u> and <u>E65 Creation</u> reveal one of the main characteristics of CIDOC/CRM, event-orientation.
- The <<u>type> attribute</u> assigned to several elements and subelements defines different semantic mappings, making mapping even more complex
- The use of several <u>global attributes</u> (e.g. <xml:lang> attribute) expands further the CIDOC/CRM paths



### **Future Research Directions**

- To validate our approach, we plan to test the proposed mappings by running real life examples and evaluate the results.
- Another direction to explore is the definition of the reverse semantic mappings from the CIDOC/CRM ontology to the VRA Core 4.0, in order to enrich the mapping procedure proposed by our research group



## **References**

- Library of Congress. VRA Core: a data standard for the description of works of visual culture, 2011. <u>http://www.loc.gov/standards/vracore/</u>
- International Council of Museums (ICOM). The CIDOC Conceptual Reference, 2010. <u>http://www.cidoc-crm.org/</u>
- T. R. Gruber (1993). Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In *Guarino, N. and Poli, R. (eds.), Formal Ontology in Conceptual Analysis and Knowledge Representation. Kluwer Academic*
- Th. Stasinopoulou et al. (2007). Ontology-based metadata integration in the cultural heritage domain. In *Proceedings of the 10th International Conference on Asian Digital Libraries, ICADL-2007, Hanoi, Vietnam, Lecture Notes in Computer* Science (LNCS).
- M. Gergatsoulis et al. (2010). Mapping Cultural Metadata Schemas to CIDOC Conceptual Reference Model. In S. Konstantopoulos et al. (eds.), SETN 2010, Athens, Greece, May 4-7, 2010, Proceedings. LNAI, Vol. 6040, pages 321-326, Springer-Verlag



# Thank you for your attention!!!! Questions???



