

A CIDOC CRM-Based Ontology System

Federica Dentamaro¹, Paolo G. De Luca², Ludovico Genco²,
Giulia Perrino³, Chiara Cannito⁴, Mariannunziata A. Stufano⁴, Maria G. Sibilano⁵

¹Gius.Laterza & Figli S.p.A.
Bari, Italy

²Tecnopolis Csata S.c.r.l.
Valenzano, Italy

³Università degli Studi di Bari
Bari, Italy

⁴Università degli Studi di Lecce
Lecce, Italy

⁵Università degli Studi di Foggia
Foggia, Italy
dentamaro@graphiservice.it

Abstract

“Re-usability” is a key factor for increasing the value of knowledge concerning cultural heritage. To achieve this goal, can we combine the different points of view from which dissimilar users observe the same cultural assets? How can their specific needs be safeguarded? This paper describes a knowledge representation approach aimed at preserving both the re-use and the customization of cultural heritage knowledge. The first benefit is pursued through a CIDOC CRM-based *CORE Ontology*, which takes into account the “objective” features of cultural assets. These are described just once, through a model that provides general classes and properties with “high-resolution” semantics. On the other hand, knowledge “customization” is pursued through independent *Domain Ontologies*; these are based on *CORE Ontology* and describe cultural assets by asserting appropriate necessary and sufficient conditions.

1 Introduction: An Approach for Re-using Knowledge

This paper describes the approach we used to build a set of ontologies for storing, sharing, and re-using knowledge about cultural heritage. Such a difficult task, indeed, needs some preliminary remarks. The notion of “cultural heritage” (hereafter CH) indicates a complex and varied world: pictures, buildings, oral traditions, archaeological sites, and so on. In order to acquire full knowledge about CH “assets,” they all have to be analyzed and described in detail, looking at their peculiar distinguishing features. Each of these features, in fact, gives its contribution in qualifying a cultural asset as a unique and “organic” entity. Thus, a CH ontology should constitute, on one hand, a model able to take into account any considerable feature of cultural assets; at the same time, such an ontology should reach a level of representation close enough to satisfy specific needs that different cultural domains and users could point out about the CH knowledge.

The attempt to get a complete, complex representation model of cultural assets implies two main risks. First, a single model for representing several different CH features can grow in an uncontrolled way, especially if that model has to consider all the “subjective” knowledge needs coming from specific disciplines and research fields (like, for instance, archaeology, art history, history of science). Second, starting from such a CH model the task of building a “customized” representation of knowledge can be very difficult.

In order to avoid both of these problems, we separated the modeling aimed at the knowledge base (hereafter KB)

population from the modeling related to the KB access. Looking at the first goal, we tried to obtain an objective conceptualization, i.e., a *CORE Ontology*. In this regard, we have to point out that an ontology is nothing but a social and cultural artifact; it is really impossible, therefore, to look at any domain in an absolutely objective way. However, in the first step of our work we still attempted to reach such an “objective model” for CH. For the second goal, we looked for *subjective/customized conceptualizations*, focused on specific fields and special needs of final users. These are pursued through *Domain Ontologies* (hereafter *DO*): a *DO* has to consider just the narrow set of concepts and relationships needed to deal with specific knowledge goals, and so can ignore the remaining knowledge facets of a domain. As a consequence, *DOs* can better match the cultural background of their users, even from a lexical point of view. Thanks to this approach, different disciplines as well as different users can look at an asset from their specific point of view (subjective modeling), whereas the uniqueness of any asset is kept and safeguarded by the objective modeling.

2 The *CORE Ontology* as an Extended CIDOC CRM

Being focused on CH, our attempt to reach an “objective conceptualization” had to start necessarily from CIDOC CRM, the declared aim of which is “to enable information

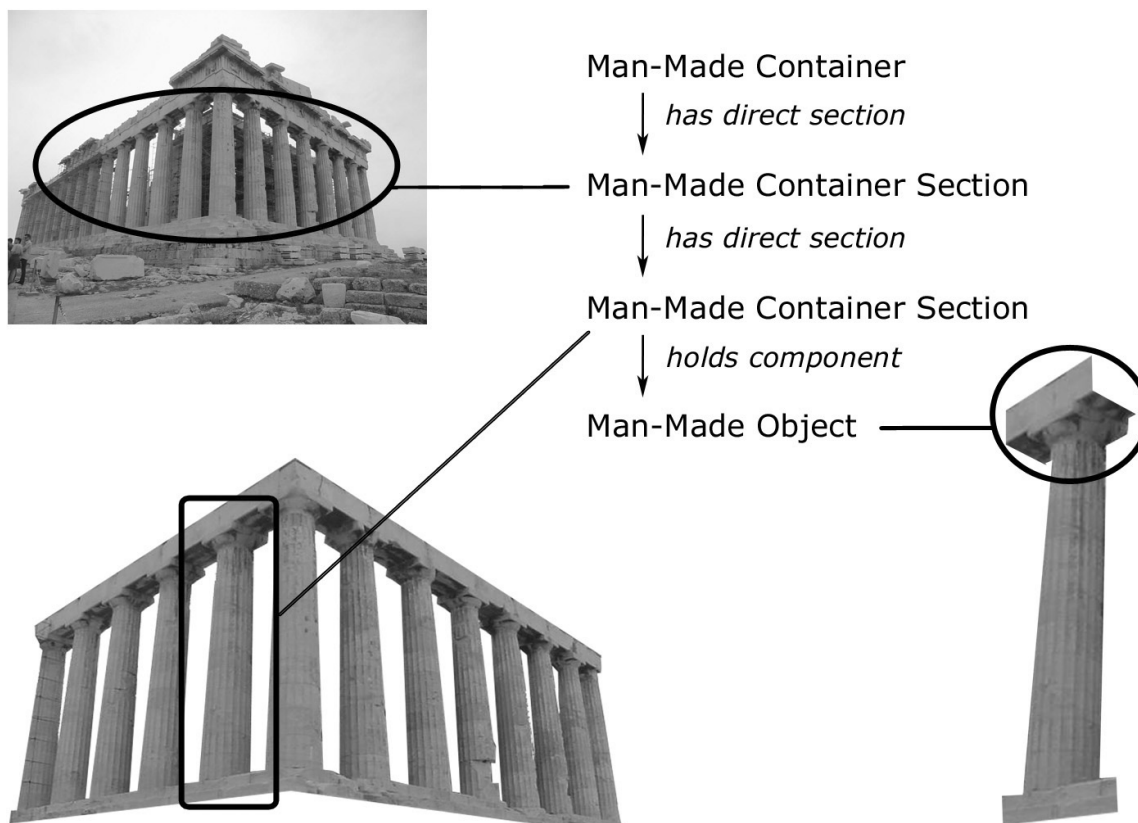


Figure 1. Within the CORE Ontology, a temple (here: the Parthenon, Athens) is represented as a “Man-Made Container,” which can be further analyzed into its relevant “Sections” (e.g., the colonnade, a single column, a column capital).

exchange and integration between heterogeneous sources of cultural heritage information” (Crofts et al. 2005:i). So, we continued testing that model in order to check the way it matches with our tasks. In fact, CRM produced *objective* as required; yet we observed a lack of descriptive power about some considerable CH features. In the following pages, we briefly expose some additions that we made to CIDOC CRM, especially those concerned with the representation of archaeological entities.

2.1 Container and Container Sections

A first set of revisions to CIDOC CRM have been motivated by the need to describe cultural assets in which other mobile assets are placed: we can call them “Containers” (just like, for example, a church containing statues, paintings, etc.). The CRM, indeed, appears to be unconcerned enough with such an eventuality; in fact, using CRM a “Container” can be described alternatively as a “Place” or as a “Physical Thing.” In the first case, however, mobile assets can be placed in it, but the “Container” itself will not be described as an “Asset;” in the second case, a “Container” can be described in detail, but mobile assets will not be placed within it. In one way or another, therefore, several important features of “Containers” have to be left apart.

Among these features, we have first to mention the

different, single parts of which a “Container” can be composed; for instance, a church is typically composed of some aisles, a transept, an apse, etc. These parts, in turn, could contain mobile assets, so even these parts should have to be described both as a “Physical Thing” and as a “Place.” Moreover, in order to represent the whole body of a “Container,” its parts have to be connected to each other by specific spatial relationships (for example, the relationship existing between the first and the second floor of a building). These relationships, actually, allow for the archaeological analysis of buildings; identifying spatial connections between single parts of a structure is the first step in reconstructing the sequence of the building phases, or the “life-cycle” of the structure.

That is why we extended the CIDOC CRM with a new class, named *Man-Made Container*, which is a subclass of both *E53 Place* and *E24 Physical Man-Made Thing*. This hierarchical structure, obviously, allows a “Container” to be described as both a “Physical Thing” and a “Place.” In order to describe a “Container” through its component parts, moreover, we built another new class, *Man-Made Container Section*, which is a subclass of both *E53 Place* and *E18 Physical Thing*; as a consequence, the “Sections” also can be seen at the same time as “Places” and as “Physical Things.” Being “Places,” then, they can be further divided into other parts (e.g., wall, floor, and ceiling in a room), each with its own features, and so on (see Figure 1). In this way, a “Container” can be described as deeply in

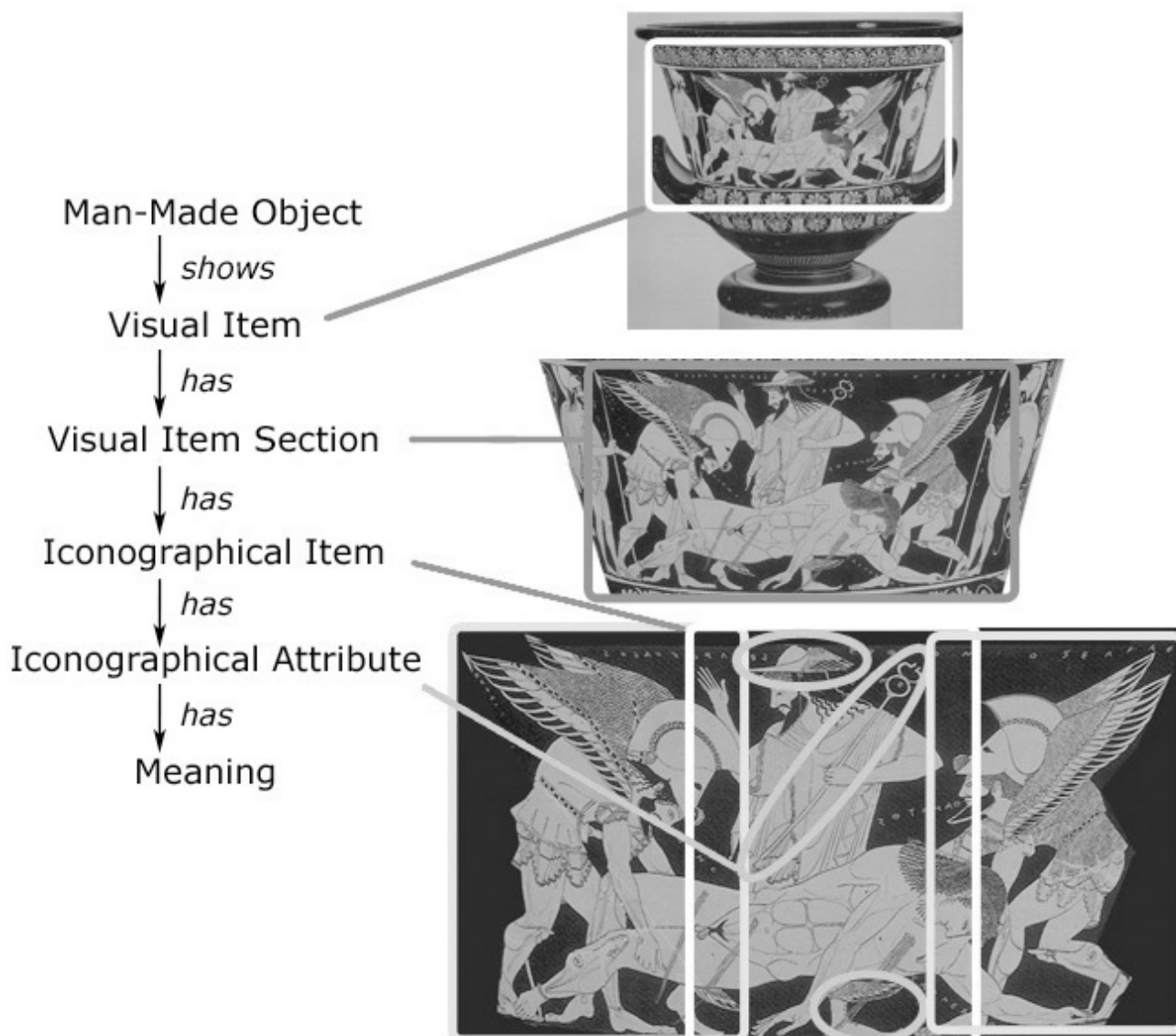


Figure 2. A vase painting (here: *Death of Sarpedon*, calyx-crater by Euphronios, about 515 BC) shows a “Visual Item”, which can be analyzed into its relevant “Iconographical Items” (each of them having “Iconographical Item Attributes”).

detail as needed, whereas every “Section” can be connected to its own “Container” as well as to other “Sections,” to this end new properties (such as, *is direct section of* and *is section of*) have been devised. Finally, the spatial relationships between “Sections” have been represented, through new properties added to the CRM (*rests on, has below, cuts, covers, etc.*). Such a model also allows us to place a cultural asset generically within a “Container” or, more precisely, within one of its “Sections.”

2.2 Iconography

Iconography deals with the identification of subjects represented in a work of art. This task is performed through detection, analysis, description, and classification of all the elements recognizable in a “visual Item;” these elements can be objects, sceneries, animals, or people. In this last case, also, “iconographical Attributes” like physical appearance

(hair, physiognomics), clothing, gesture/motion, and so on become matters of analysis (Panofsky 1995, 1996).

Figure 2. A vase painting (here: *Death of Sarpedon*, calyx-crater by Euphronios, about 515 BC) showing a “Visual Item” that can be analysed into its relevant “Iconographical Items” (each of them having “Iconographical Item Attributes”).

Within the CIDOC Model, “intellectual or conceptual aspects of recognizable marks and images” (Crofts et al. 2005:18) can be represented using the *E36 Visual Item*; this class, however, allows for just the description of a visual item in itself, and this could not be enough to identify the subject of a work of art. In order to achieve such an outcome, therefore, we made another addition to the CRM, with a set of new specific concepts (placed as subclasses of *E28 Conceptual Object*), and new relationships between them; altogether they allow for a structured description of iconographical subjects. Such a description takes into account the *canonical Representation* (i.e., coded conceptual prototype) of a specific subject as opposed to its variants represented in existing works of art.

The above distinction is performed decomposing a *Visual Item* shown by an object (e.g., a vase painting) into *Sections*, each of them eventually containing generic *iconographical Items*; for example, objects or people. The characterization of people in terms of their *iconographical Attributes* (for instance, petasus, caduceus, and winged shoes of Mercurius, as shown in Figure 2) enable us to assert whether a concerned representation is a canonical or a non-canonical one.

Finally, each “iconographical Attribute” can be connected to a *Meaning*. This relation, if necessary, can be qualified in terms of space, time, and other features in order to describe its context of validity.

2.3 Cultural Assets Context

A comprehensive analysis of any cultural asset cannot be carried out apart from considering its geographical and historical *context*, i.e., the specific place and time in which that asset was made. Such a context can be represented through the set of relationships existing between a *territory* (which can be subjected, time by time, to several, successive *sovereignties, managements, and appellations*) and the *ages* it traversed: for instance, universally acknowledged historical periods (like the Roman Early Empire), but also shorter phases and peculiar time-lags, which typically are strictly connected to the *rules* of a place (e.g., the Byzantine Rule of Apulia).

In other words, such an entity—the new concept of *Land*—can be seen as an abstraction of a “Place” in “Time,” a sort of “Place” whose boundaries can change in time but still remain the historical referent for nowadays existing “Places” (Cities, Districts, and so on). In order to be fully described, this extra-spatial entity has to be related to *historical Phases* (for example, the “Land of Apulia” is related to the “Norman Phase,” the “Angevin Phase,” etc.), as well as to the intercepted *Toponyms* and to *political Dominations*.

In turn, each of these characterizations can be further detailed: a *Domination*, for instance, can be described through its *Rulers* and their biographical features, *Titles*, and *Roles*. In other words, representing the “Angevin Domination in South Italy” implies the description of “Charles I Anjou” (*Ruler*), “king” (*Title*) of the “Realm of Sicily” (*Rule*).

We modeled all the above mentioned concepts through new classes, making up another addition to CIDOC CRM, in order to achieve a broad representation of historical and geographic contexts. Special properties were modeled, too, which allow one to place cultural assets within their proper context. Although the resulting model is in some ways complex, nonetheless it could be further extended to include, for example, cultural or religious phases modeling.

2.4 Types

This class comprises arbitrary concepts (universals) and provides a mechanism for organising

them into a hierarchy The class *E55 Type* can be regarded as a metaclass ... used to denote a user-defined specialization of some class or property of the Model, without introducing any additional formal properties for this specialization Ideally, instances of the class *E55 Type* should be organised into thesauri, with scope notes, illustrations, etc. to clarify their meaning.

In general, it is expected that different domains and cultural groups will develop different thesauri in parallel (Crofts et al. 2005:26).

The CORE elaboration tried—to some extent—to produce a *backbone* of “Types,” so that the cited “different domains and cultural groups” can have an organized and shared categorization for their own instances of (sub)class (of) CRM concept *E55 Type*. All those instances can be related to each other through the CRM properties *P127 has broader term (has narrower term)*, whatever the originating thesaurus. We need to point out that in order to reduce the reasoning effort, we moved the class *E55 Type* under *E1 CRM Entity* (while CIDOC CRM established it as a subclass of *E28 Conceptual Object*).

Into the backbone of “Types,” noteworthy sub-taxonomy are starting from *Material, Place Type, Man-Made Object Type, and Intended Use* classes. As an example of the issues we faced, here we briefly discuss just the taxonomy introduced by the *Intended Use* class. Such a class has its grounds based on the remark that “different domains and cultural groups” need to describe a cultural asset (instance of the CRM class *E70 Thing*) also by representing the *intended uses or functions* ascribed to it, both in the past—at the time of its production—and today. Indeed, a cultural asset could be made in a “non-cultural” context; however, even right after its production, this context may change and so, too, the use of the asset, being amplified or differently perceived in the light of current needs.

CIDOC CRM does not foresee such a change, allowing just for the representation of the original use of any cultural asset (with the *P101F had as general use* property). In that model, no taxonomy is developed concerning the intended uses for cultural assets, while the taxonomy in our *CORE* provides for 65 taxa, organized in six branches: *Structural, Social, Scientific, Decorative, Perceptive, and Relational Use*. Moreover, in order to take account of the difference between original, following and current uses of a cultural asset, we built the properties hierarchy seen in Figure 3

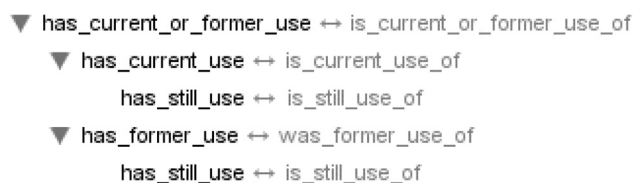


Figure 3. Hierarchy of properties for the *Intended Use* class.

As a further consideration, the whole *Type* taxonomy can be used to define non-primitive concepts within a *Domain Ontology*. Indeed, necessary and sufficient conditions can be asserted through proper restrictions on the range of the CRM property *P2F has type*. For example, such a restricted

Table 1. Measure of CIDOC CRM extensions.

CORE Concepts		CORE Properties	
CRM Concepts	Added Concepts	CRM Properties	Added Properties
79	252 194 sub Type 11 sub Temporal Entity 16 sub Place 29 sub Persistent Item 7 Roles qualifiers	132	72 66 Object Properties 6 Data Properties

range can be set as an “anonymous” class constituted by a *pivot* instance *i* along with all instances that have *i* as their broader term.

Finally, however, a theoretic question arises: as many dictionaries of terms (thesauri) are “open,” may a “close” *backbone* taxonomy of “Types” encompass all of those terms?

2.5 Property Qualifications

Into the above said *backbone* of “Types,” classes are provided concerning specific roles which an actor can hold in participating to events and performing activities (for instance as “orderer,” “architect,” or “worker” in a building construction). The mere statement of classes (and individuals) related to these *Roles*, however, is not enough to achieve a level of conceptualization able to represent the participation of “Actors” (individuals of *E39 Actor* or its subclasses) in “Events” (individuals of *E5 Event* or its subclasses). In fact, a specific conceptual model is required for qualifying the “Role” carried out by an “Actor” in any “Activity.”

CIDOC CRM, indeed, provides the property *P14.1 in the role of* for pointing out the role held by an actor; this, however, cannot be used for a detailed representation with respect to the needs we said just above. On the other hand, CIDOC CRM drafters positively allow for additions to this part of their model. So we attempted another solution by modeling a *Property Qualification* class, having as its own properties *subject*, *object*, and *predicate qualifier*; these allow for representing the needed assertions. In other words, through such modeling we can state, for example, that “Charles I Anjou” (*object*) participated as “orderer” (*predicate qualifier*) in the “building of Castel Nuovo in Naples” (*subject*).

2.6 Measure of CIDOC CRM Extensions

At the end of our work, the *CORE Ontology* extends the CIDOC CRM with 252 new classes and 72 properties. At first sight, the number of added classes seems considerable. Actually, it is rather restricted if we remember that, for the most part (namely 194 among 252), these new concepts are made up to build a taxonomy of thesauri and terms under

the *E55 Type* class, as foreseen by CIDOC CRM drafters. Concerning the 72 added properties, these are dedicated for the most part to establish special relationships between the added concepts, or between a new one and a CIDOC CRM class. In a few cases, they can also establish a new connection between CIDOC CRM concepts; then they start a hierarchy of existing roles.

Table 1 shows how and where the *CORE Ontology* extends the different branches of CIDOC CRM.

3 Domain Ontologies Case Studies

The discussed approach (*CORE + DOs*) has been tested through three case studies. Two of them are aimed at specific disciplines, building archaeology and iconography; so their main task is to deal with narrow and distinct portions of the world represented by the *CORE Ontology*, in order to achieve a deep description of those portions, suitable for scientists and researchers. In this way, people will also be allowed to use their peculiar language and vocabulary in interfacing with the KB.

The third case study, on the other hand, is focused on reaching a CH representation suitable for tourists. In carrying on this task, both the CH features and the background of their final users must be considered; moreover, a detailed description of historical, social, and economic processes is required. For this reason, from a theoretical and methodological point of view this last Case Study seems strictly connected with issues developed, for instance, by archaeological communication (Forte 2000).

3.1 Building Archaeology Domain Ontology

Our first Domain Ontology is focused on specific needs of building archaeology. This discipline (also called archaeology of standing structures) employs stratigraphic methods in order to study ancient architectures; thus, it needs to identify, on the body of any building structure, *Building Layers* (i.e., recognizable signs of single building activities, made into a defined space and at a determined time) as well as *Coating Layers* (plaster layers, wall paintings, mosaics, etc.). Building Layers (and Coating Layers) will be ordered into a chronological sequence, based on all

CORE classes which are considerable for Iconography special aims: *Image*, *Worship*, *Iconographical Item*, *Orderer* and *Ordering Event*, *Man-Made Object*. Restrictions are usually put on properties whose range is in the above said *Type* taxonomy; for example, some of the involved ranges are *Intended Use*, *Design or Procedure Type*, *Actor Role* classes.

As an example, within the *DO* the *Image* concept is defined as a *Physical Man-Made Thing* that shows a *Visual Item*. In so doing, it qualifies the artistic object in a “corporeal” sense, opposite to the “conceptual” sense given to the homonymous concept of the *CORE*. Another example may concern the *Iconographical Item* class. Within the *CORE*, indeed, this class defines the representations of people and things shown by “Visual Item Sections.” This is enough to satisfy the requirements for a first, general approach to the comprehension of images; in the *DO*, on the contrary, it will be needed to specialize the analysis of an image within the iconographic domain. In other words, we need to provide the description of an “Iconographic Subject” through a detailed “de-composition” of a “Visual Item” and its “Sections.” Looking for this aim, in the *DO* we created another class, named *Iconographical Item*, having three subclasses (*Character*, *Scenery* e *Animal*); in such a way, it is possible to qualify different kinds of “Iconographical Items.”

3.3 CH Domain Ontology for Tourism

Unlike the previously mentioned *DOs*, this last one was made in order to answer needs and interests about CH pointed out by tourists. Such interests may concern, for example, historical, stylistic, or technical features of an asset; so a taxonomy of concepts has been created, regarding all those features and in this way providing for a broad representation of cultural assets, suitable for tourism.

Since this *DO* is not actually concerned with a specific discipline, it has to represent cultural assets as belonging to one of these four domains in the CH field: archaeology, history of art, history of science, anthropology. For each of these, a specific taxonomy was developed, including different typologies of Assets (and so providing, for instance, for concepts like Temple, Theatre and Mosaic).

As usual, all these *DO* concepts were defined imposing necessary and sufficient conditions on *Core Ontology* concepts, often making use of the *backbone* of “Types.” An asset, hence, can be seen in different ways, according to the chosen point of view (which of course depends from special interests of users). For example, a mosaic can be seen as a “Mosaic” *tout court* (which is defined as a “Man-Made Object” having a specific “Type”), but also as an “Archaeological Find” (defined as a “Physical Thing” involved into a “Finding” event). Moreover, such definitions are different from those in the Building Archaeology *DO*: there, in fact, a “mosaic” can be represented both, as a “Coating Layer,” or as an “Archaeological Find” (but in this case we are talking about a “Man-Made Object” located into an “Archaeological Context”). The differences mentioned

show how extremely far can be the points of view through which different *DO* look at a same cultural asset.

4 Experience with the Approach

The formal model of CH knowledge system is based on OWL-DL representation (McGuinness and, van Harmelen 2004) and requires the ALCQHIR+ highly expressive logic. The software used for building all the ontologies is the open source *Protégé*, version 3.1 (<http://protege.stanford.edu/index.html>), while the reasoning system used is *Racer Pro* by Racer Systems GmbH & Co. KG (<http://www.racer-systems.com/>).

Both the *CORE* model complexity and the logic expressiveness actually imply reasoning performance problems. Table 2 shows how, for example, a *Man-Made Object* class member could be described using 90 allowed properties (hierarchic roles included).

Table 2. Number of roles for some considerable concepts of *CORE Ontology*.

	CORE Concepts	CORE Roles	
		CRM	Added
CIDOC-CRM Concepts	Event	44	11
	Place	29	12
	Type	22	9
	Man-Made Object	67	23
	Conceptual Object	36	24
Added Concepts	Man-Made Container	71	41
	Iconographical Item	37	31

Hence, at the moment, if extensively populated, such a Cultural Heritage KB represents a tremendous reasoning effort. Many small KBs, dedicated to few, similar cultural assets could reach the best effectiveness. Also, the adopted approach itself (population through a *CORE Ontology*, accessed through *DOs*) prevents any attempt to optimize the model. Reducing the number of inverse properties, for example, is a quite impossible task. Indeed, the way in which the *CORE* model will be traversed by *DOs* cannot be fixed in advance; in other words, we cannot know in advance which necessary and sufficient conditions or role restrictions will be asserted over the *CORE Ontology* concepts by *DOs*. Hence, we cannot restrict the inversion of roles to a narrower set of properties.

The approach mentioned is subject to a second inborn drawback, arising whenever two or more available *DOs* are imported to form a new one. Indeed, as *DOs* can be produced as separate entities from one another, they may likely contain equivalent concepts or domain concepts subsumed by more than one other domain concept. This mostly happens when concepts’ semantics take into account the boundary shared by the involved domains: for this reason, the resulting combined model can potentially suffer a low normalization. Moreover, nominally identical concepts could have distinct conditions, which are asserted as necessary

and sufficient. Vice versa, concepts with identical asserted conditions could have different names.

On the other hand, the approach shows an unquestionable advantage: any set of cultural assets need to be described just once, through the *CORE* model. Different users can then have access to the available knowledge through the *CORE*, as well as by the mediation of a *DO*; this can actually represent the former knowledge in a more natural, immediate way with respect to the background of the user and his (or her) peculiar needs.

As a consequence, the implementation of some *DOs* is an ideal benchmark to test the *CORE* conceptualization completeness, as well as the completeness of cultural assets' description (i.e., the KB population). Indeed, the definition of concepts in *DOs* stresses the ability of the *CORE* model in giving multi-faceted representations of the same cultural asset. In fact, these representations are obtained mostly through restrictions asserted on the *CORE* model properties; hence, if a concept cannot be defined in a *DO* by asserting any restriction, then the *CORE* conceptualization model is inadequate. Moreover, if an inference on the *CORE + DO* system does not give the expected results (or better, if an individual existing in the KB is not inferred as a member of a defined concept as foreseen) then the description of individuals (i.e., the instantiation) is most likely incomplete.

Acknowledgements

The issues here shown have been developed within the CNOSSO Project, partially funded by MIUR – Ministero dell'Istruzione, dell'Università e della Ricerca – Italy.

This paper has been drawn, organized, and revised by all the authors together; so Introduction (par. 1) has been written in common. Single paragraphs have been further developed as follows: 2.1 and 3.1 by Federica Dentamaro, 2.2 and 3.2 by Giulia Perrino, 2.3 by Chiara Cannito, 2.4 and 2.5 by Mariannunziata A. Stufano, 2.6 and 4 by Paolo G. De Luca and Ludovico Genco, 3.3 by Maria G. Sibilano.

References Cited

Barbanera, Marcello. 2000. Iconografia e Iconologia. In, *Dizionario di Archeologia*. R. Francovich, D. Manacorda, eds., pp. 163-168. Roma-Bari: Laterza.

Crofts, Nick, Doerr, Martin, Gill, Tony, Stead, Stephen, and Stiff, Matthew, eds. 2005. Definition of the CIDOC Conceptual Reference Model, version 4.2. Updated July 2005, Electronic document, http://cidoc.ics.forth.gr/official_release_cidoc.html (Accessed 28 June 2006).

Ferrando Cabona, Isabella. 2002. Guida critica all'archeologia dell'architettura. *Archeologia dell'Architettura* 7:7-42.

Forte, Maurizio. 2000. Comunicazione archeologica. In, *Dizionario di Archeologia*, R. Francovich, D. Manacorda, eds., pp. 75-80. Roma-Bari: Laterza.

Harris, Edward C. 2003. The Stratigraphy of Standing Structures. *Archeologia dell'Architettura* 8:9-16.

McGuinness, Deborah L. and van Harmelen, Frank, eds. 2004. W3C Recommendation. OWL Web Ontology Language Overview. Updated 10 February 2004, Electronic document, <http://www.w3.org/TR/2004/REC-owl-features-20040210/> (Accessed 28 June 2006).

Panofsky, Erwin. 1995. Sul problema della descrizione e dell'interpretazione del contenuto di opere d'arte figurativa. In, *La prospettiva come forma simbolica*. G. Neri, ed., pp. 215-232. Milano: Feltrinelli.

Panofsky, Erwin. 1996. Iconografia e iconologia. In, *Il significato nelle arti visive*. E. Panofsky, ed., pp. 29-57. Torino: Einaudi.

Parenti, Roberto. 2000. Archeologia dell'Architettura. In, *Dizionario di Archeologia*. R. Francovich, D. Manacorda, eds., pp. 39-43. Roma-Bari: Laterza.