The CIDOC CRM – an Ontological Approach to Semantic Interoperability of Metadata

M. Doerr

Institute of Computer Science,
Foundation for Research and Technology – Hellas,
Science and Technology Park of Crete,
Vassilika Vouton, P.O. Box 1385, GR 711 10, Heraklion, Crete, Greece
martin@ics.forth.gr

Abstract: This paper presents the methodology that has been successfully employed over the past 7 years by an interdisciplinary team to create the CIDOC Conceptual Reference Model (CRM), a high-level ontology to enable information integration for cultural heritage data and their correlation with library and archive information. The CIDOC CRM is now in the process to become an ISO standard. This paper justifies in detail the methodology and design by functional requirements and gives examples of its contents. The CIDOC CRM analyses the common conceptualizations behind data and metadata structures to support data transformation, mediation and merging. It is argued that such ontologies are property-centric, in contrast to terminological systems, and should be built with different methodologies. It is demonstrated that ontological and epistemological arguments are equally important for an effective design, in particular when dealing with knowledge from the past in any domain. It is assumed that the presented methodology and the upper level of the ontology are applicable in a far wider domain.

1 Introduction

The creation of the World Wide Web has had a profound impact on the ease with which information can be distributed and presented. This has spurred an increasing interest from professionals, the general public, and consequently politicians to make publicly available the tremendous wealth of information kept in museums, archives and libraries - the so-called "memory organisations". Quite naturally, their development has focused on presentation, such as websites and interfaces to their local databases. Now with more and more information becoming available, there is an increasing demand for targeted global search, comparative studies, data transfer and data migration between heterogeneous sources of cultural contents. This requires interoperability not only at the encoding level - a task solved well by XML for instance - but also at the more complex semantics level, where lie the characteristics of the domain.

Formal methods are very helpful to deal effectively with the large amounts of information coming together on the Internet. Information about cultural heritage poses particular challenges for formal handling – rather not, as often assumed, because it is ill defined, but because of its high *diversity*, and the intrinsic *incompleteness* of information about the past. So far most attempts for semantic interoperability have concentrated on the development and standardization of a shared core data structure (e.g. the CIDOC Relational Model) and a terminology system. In cases a common data structure seemed to be impossible, at least a common metadata schema as "finding aids" have been attempted, the most prominent example being the Dublin Core Element Set. On the terminology side, the Library of Congress Subject Headings and the Art & Architecture Thesaurus are characteristic examples of standards in the US and beyond, for which equivalents in several other languages have been created.

In the meanwhile, the reality of semantic interoperability is getting frustrating. In the cultural area alone, dozens of "standard" and hundreds of proprietary metadata and data structures exist, as well as hundreds of terminology systems. Core systems like the Dublin Core represent a common denominator by far too small to fulfill advanced requirements. Overstretching its already limited semantics in order to capture complex contents leads to further loss of meaning ("metadata pidgin"[1]), even though most of the contents encoded in the various structures seem to be pretty comprehensive in common sense terms, and are often inter-related. We make the hypothesis that much of the diversity of data and metadata structures is due to the fact that they are designed for data capturing – as a guide for good practice of what should be documented, and to optimise coding and storage costs for a specific application – far more than for interpreting data. Necessarily, these data structures are relatively flat (in order to suggest a workflow of entering data to the user) and full of application-specific hidden constants and simplifications.

Since 1996, we have taken part in the development of the CIDOC Conceptual Reference Model (CRM) [2], [3],[4], an attempt of the CIDOC Committee of the International Council of Museums (ICOM) to achieve semantic interoperability for museum data. Work has started in the beginning on a more intuitive base, from a knowledge representation model [5],[6], based on the consensus of a varying team of different domain experts and based on strict intellectual principles. It has got a wide acceptance in CIDOC and by other relevant stakeholders in the domain and in September, 2000, the CIDOC CRM was successfully submitted to ISO TC46 as new work item. It is now registered as ISO/CD 21127 and is expected to become an ISO standard in 2003. It is now in a very stable form, and contains 80 classes and 130 properties, both arranged in multiple isA hierarchies. In the meanwhile, several applications [7] and comparison with related work improves our theoretical understanding of the work done and still ongoing.

Instead of seeking a common schema as a prescription for data capturing, which would be supposed to ensure semantic compatibility of the produced data, we have followed with the CIDOC CRM what Bergamaschi et.al [8] call a semantic approach to integrated access. Being convinced that collecting information is already done well by the existing data structures, possible improvements not withstanding, we aim simply for read-only integration, as e.g. in the DWQ project [9][10]. This comprises data migration, data merging (materialized data integration as in data warehouse applications), and virtual integration via query mediation [11]. Recently more and more projects and theoreticians support the use of formal ontologies as common conceptual schema for information integration [12], [13], [8],[16], [14],

- to provide a conceptual basis for understanding and analyzing existing (meta)data structures and instances;
- to give guidance to communities beginning to examine and develop descriptive vocabularies;
- to develop a conceptual basis for automated mapping amongst metadata structures and their instances[14].

Its seems that the semantics behind a large set of diverse (meta)data structures from a domain with many subdisciplines can be expressed by a coherent formal ontology based on the common conceptualisations of the respective domain experts, whereas the data entry structures themselves often seem to resist merging. We have followed a pragmatic approach to separate a kind of top-level ontology [12], which represents knowledge extracted from schemata and data structures, from pure terminology. This was partially done in order to keep the basic ontology in a manageable size. The semantics of the data structures are richer in n-ary relationships (attributes, properties etc.) than in fine distinctions between classes, whereas thesauri are just the opposite – they build rich isA (BT/NT) hierarchies but typically employ only one ("RT") relationship for any other internal conceptual relation. We turned this observation into a rule: classes were introduced in the CRM only for the domains or ranges of the relevant relationships, such that any other ontological refinement of the classes can be done as additional "terminological distinction" without interfering with the system of relationships (see also [2]). Such a conceptual model seems to cover the ontological top-level automatically, and provides an integrating framework for the often isolated hierarchies found in terminological systems (Fig. 2). We call such a model a "property centric ontology" to stress the specific character and functionality.

This paper presents the CIDOC CRM from a methodological point of view. It relates the intended scope and functionality to the ontological principles that governed its design, presents its key concepts and positions the model with respect to relevant related work. We expect this methodology to be applicable to other domains, even though there is no experience yet to support this claim. The significant contribution of this work are considerations about the specific nature of cultural – historical knowledge and reasoning, which aims at the reconstruction of possible past worlds from loosely correlated records rather than at control and prediction of systems, as in engineering knowledge. "Historical" must be understood in the widest sense, be it cultural, political, archaeological, medical records, managerial records of enterprises, records of scientific experiments or criminalistic data.

2 The Problem

2.1 The Necessity of Data Structure Diversity

Let us regard here data structures for long-term storing of data, as database schemata, but also tagging schemes like SGML/XML DTD, RDF Schema, and data structures designed as fill-in forms guiding users to a complete and consistent documentation, be it as primary data or as metadata about another information source. These structures are always a compromise between the complexity of the information one would like to make accessible by formal queries, the complexity the user can handle, the complexity of the system the user can afford to implement or to pay for, and the cost to learn those structures and fill them with contents. As most applications run in a relatively uniform environment – a library, a museum of modern art, a historical archive of administrational records, a paleontological museum etc., much of the complexity of the one application is negligible for another. This allows for a variety of simplifications, which are required to create efficient applications. For example, documentation in modern art needs no other dating than Julian dates, whereas for archaeology dating is a process of multiple measurements, evaluation of sources, inferring and justification, including different dating systems. It makes no sense to ask the modern art curator about carbon 14 measurements, nor to reserve dozens of special fields and storage space never used. Nevertheless, and this is the crucial point, the notion of date and dating of both experts are completely compatible; there is no difference in conceptualisation, at least from a scientific point of view. The complexity typical for archaeology hardly ever occurs in modern art, however. Similarly, the documentation of a historical building implies a complex history with various phases and persons implied, whereas paintings are typically created in a relatively compact process. Consequently, art documentation schemes like AMICO [15] do not capture multiple creators for multiple phases of objects, in contrast to the architectural descriptions of the Greek National Archive of Monuments [16],[17]. The rare exceptions to such simplifications are normally handled by free text comments and are not used as reasons to change the schema.

Another criterion of simplification appears in the "finding aids". Subtle differences in association, like John and Mary collaborating in the design of one building, but John overseeing the design and Mary the construction of another building, cannot be captured by a scheme listing all involved persons without their individual roles. The question is, how much noise will the replacement of the query "John and Mary designing" by "John and Mary involved" create – probably not much. This is the justification for the use of "flat" metadata records like Dublin Core, that sum up relevant persons and relevant dates etc. without interrelations. These simplifications actually violate our conceptualisation, the sources retrieved on this basis can only be sorted out by reading them. How long this will work is a question of scale. If we have the resources and the requirements for more complex searches and processing, we must find a way to "recover" the common conceptualisation behind these simplifications, if at all possible, or improve our data structures (see section 4.4).

2.2 The Yalta Conference – a demonstration case

Let us regard an artificial, but realistic demonstration case about information objects related to the Yalta Conference in February, 1945. This was the event officially designating the end of WWII. One can hardly find a better documented event in history. We owe the association to the photo below to the Microsoft Encarta Encyclopedia 2000. The text and the TGN record we found on the Internet. The titles are as we have found them. We have created the demonstration metadata below from the information we found associated with the objects:

The State Department of the United States holds a copy of the Yalta Agreement. One paragraph begins, "The following declaration has been approved: The Premier of the Union of Soviet Socialist Republics, the Prime Minister of the United Kingdom and the President of the United States of America have consulted with each other in the common interests of the people of their countries and those of liberated Europe. They jointly declare their mutual agreement to concert …" [http://www.fordham.edu/halsall/mod/1945YALTA.html]. A Dublin Core record may be:

Type:	Text
Title:	Protocol of Proceedings of Crimea Conference
Title.Subtitle:	II. Declaration of Liberated Europe
Date:	February 11, 1945.
Creator:	The Premier of the Union of Soviet Socialist Republics
	The Prime Minister of the United Kingdom

	The President of the United States of America
Publisher:	State Department
Subject:	Postwar division of Europe and Japan

The Bettmann Archive in New York holds a world-famous photo of this event (Fig. 1). A Dublin Core record for this photo might be:

Type:	Image
Title:	Allied Leaders at Yalta
Date:	1945
Publisher:	United Press International (UPI)
Source:	The Bettmann Archive
Copyright:	Corbis
References:	Churchill, Roosevelt, Stalin

Figure 1: Allied Leaders at Yalta

The striking point is that both metadata records have nothing more in common than "1945", hardly a distinctive attribute. An "integrating" piece of information comes from the Thesaurus of Geographic Names (TGN, [http://www.getty.edu/research/tools/vocabulary/tgn/index.html]), which may be captured by the following metadata:

TGN Id:	7012124
Names:	Yalta (C,V), Jalta (C,V)
Types:	inhabited place(C), city (C)
Position:	Lat: 44 30 N,Long: 034 10 E
Hierarchy:	Europe (continent) <– Ukrayina (nation) <– Krym (autonomous republic)
Note:	Located on S shore of Crimean Peninsula; site of conference between Allied powers in WW II in 1945; is a vacation resort noted for pleasant climate, & coastal & mountain scenery; produces wine, canned fruit & tobacco products.
Source:	TGN, Thesaurus of Geographic Names

The keyword "Crimea" can finally be found under the foreign names for "Krym", i.e. via another record (id=1003381). This example demonstrates a fundamental problem: In order to retrieve information related to **one** specific subject, information from multiple sources must be **integrated.** Vocabulary and data structure unification only does not solve the problem.

2.3 Requirements

One problem in this example is to be able to relate Crimea to Krym and then to Yalta, the Premier of the Union of Soviet Socialist Republics to Joseph Stalin and to the Allied Leaders etc. A deeper problem is the fact that the artifacts do not fit our question: People document persistent items like images, texts, and places, but our question was about an event, here the Yalta Conference, something that is only indirectly preserved in those items. The data structures express certain relationships between items, which may or may not be globally identified. Other relations are hidden, like *UPI taking pictures*, which can be either guessed from the context or must be recovered from secondary sources or from background knowledge (at these times, the press photographers were not documented). Having argued above that data structures are full of simplifications and hidden constants, we see that the main problem is to recover this information during data integration. This is where ontologies are most valuable.

When we started work on the CIDOC CRM in 1996, the CIDOC working groups had virtually given up on creating one standard data structure for all museums. We have assumed (and still assume) that a fairly small set of good practice guides (e.g. mda SPECTRUM [18], CIDOC International Guidelines for Museum Object Information [19]) and standard data structures already express well what museum professionals want and should say about their objects in the various disciplines (an extended list can be found in [20]), albeit these can still be improved. In particular, the necessary constraints to improve data integrity are typically applied locally at data entry time already.

The global interoperability between disciplines is clearly needed for the following functions, after data has been created:

- The mediation of global queries to local structures [11];
- the extraction of individual statements from larger units of documentation and
- their comparison for alternative opinions;
- the transformation of data for migration to other systems and
- for merging into more informative units (like data warehouses).

The CIDOC CRM working group wanted to provide one key element: the encoding of the key domain conceptualizations by an interdisciplinary group in a form that enables the above functionality and is extensible enough to ensure a long life-cycle and increasing coverage of details and disciplines. In addition, the ontology is also thought as an intellectual guide in the requirements analysis and conceptual modeling phase of cultural information systems as proposed by [12]. Parallel to the on-going work, more and more methodological principles have been elaborated and applied on the basis of these fundamental requirements. We are now in the position to give a consistent account of this methodology, that has been documented so far only in presentation slides and minutes of the working group. (e.g. [21]).

3   About the CRM Methodology

The problems computer scientists and system implementers have in comprehending the logic of cultural concepts seems to be equally notorious as the inability of the cultural professionals to communicate those to computer scientists. The CIDOC CRM working group is therefore interdisciplinary, aiming at closing that gap. People with background in museology, history of arts, archaeology, natural history, physics, computer science, philosophy and others where involved. We have achieved a functional compromise between the complexity of the conceptualisations and the complexity of formalism the participants would appreciate. Therefore the AI reader may miss in this work some obvious formalization, which is due to future work by the appropriate specialists. On the other side we could convey in a series of targeted seminars more KR principles to non-experts than in any other related standardization work.

Given the limited resources of a project that had no funding at all until recently, and the interdisciplinary character of the group, the concern has always been to concentrate the resources on the most effective task for such a group: to achieve consensus about the ontological commitment of a set of formally defined core concepts of the domain in a way that can guide implementers and computer scientists, and can later be refined by domain specialists. Therefore, many good contributions (e.g. modelling beliefs) were excluded just because they could be taken over either by specialists in a later stage or because they could be dealt with separately. Several of them are mentioned in this paper. Also, strict neutrality with respect to commercial interest groups is the declared policy of CIDOC. All this contributed to create a construct of high intellectual quality and coherence. Starting from an initial formulation of the scope of the CIDOC CRM [21], we have independently developed intellectual principles similar to those in [13].

The CIDOC CRM is a formal conceptual model, however CIDOC CRM instantiation data (factual knowledge) is allowed to be contradictory. Historical data - any description made in the past about the past, be it scientific, medical or cultural – is normally unique and cannot be verified, falsified nor completed in an absolute sense. In history, any conflict resolution of contradictory records is nothing more than yet another opinion. So, for an ontology capable of supporting the collection of knowledge from historical data, ontological principles about how we perceive and express things **must** be taken into account, and epistemological principles about how knowledge **can** be acquired must be respected. There can be huge differences in the credibility of propositions. Typically claims about the existence of material particulars (e.g. El Greco, Mona Lisa, Niniveh) are by far more stable than the reported relationships and attributes. A bit more frequent than absolute doubts about existence of

material particulars are doubts of whether two individuals have actually been one (two pieces of pottery may have been from the same pot, or two different pots), or similarly if one has been two (see e.g. the Union List of Artist Names, [23]). Without going into more detail here, we want to

- provide an **ontologically correct** conceptual model that is
- **compatible** with the granularity of knowledge we typically get from our sources, and
- allows for **compiling** gracefully **contradictory** information.

3.1 Integration of Context-Free Propositions

In the following we mean by conceptual model or ontology a description of categorical knowledge about "possible states of affairs" rather than about one state of affairs [12], and regard both as a special kind of knowledge base [24]. We prefer the term "conceptual model" when talking about actual instantiation and constructs dictated more by the representation formalism than the intended meaning. Categorical knowledge may come from the analysis of data structures, hidden constants or terminology used in the data. We have the vision of a global semantic network model, a **fusion** of relevant knowledge from all museum sources, **abstracted** from their context of creation and **units of documentation** under a common conceptual model. The network should, however, not replace the qualities of good scholarly text. Rather it should maintain links to related primary textual sources to enable their discovery under relevant criteria.
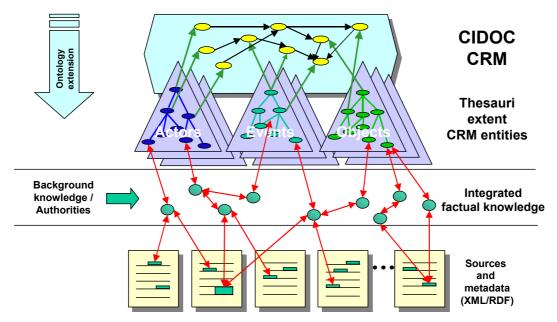

Figure 2: An information integration architecture

Figure 2 shows a possible architecture integrating a property-centric top ontology (here the CIDOC CRM), which provides the semantics for properties of subordinate terminological systems and an integrated factual knowledge layer constructed from source data, metadata and background knowledge like the TGN and other authorities.

The CIDOC CRM plays the role of an "enterprise model" [10], in the following called "common model". We assume for all sources the existence of a conceptual model ("source model"), and that source data can be expressed without loss of meaning in terms of a source model, which is based on the same formalism as the enterprise model. The source model may be restricted to the semantics falling within the scope of the common model. As a representation formalism we have selected the TELOS data model [25],[26] without its assertional language. TELOS, as many other knowledge representation languages, decomposes knowledge into elementary propositions – declarations of individuals, classes and binary relations.

The properties of TELOS relevant for the purpose of this paper are similar to those of RDF, RDFS [27]. As RDF (and OWL) are now on the way to becoming standards for the applications we target, we shall use here the terminology of RDFS as it may be more familiar than that of TELOS, and talk about classes and properties. As our primary interest is ontological, we intend to edit the CRM in various representations, but the primary source

for the CRM is a complete implementation in TELOS on the SIS knowledge management system [28]. Logical assertions are omitted because they can be added in a later stage, once the ontological commitment of the primitive classes, properties and isA relations are set up satisfactorily.

The process of instantiating the common model with factual knowledge can be broken into 2 steps: (1) The creation of global identifiers for the instances of classes (individuals) taken from an interpretation of the source data and their classification in the common model. "Global" is understood with respect to the declared scope of the application. (2) The instantiation of the properties (roles, relationships) of the common model with relations that connect those individuals and that are compatible with the intended meaning of the source data. The mechanisms of creating the global identifiers themselves are out of the scope of the CRM work. Relevant for the design of the ontology are the following properties we would like the represented factual knowledge to satisfy:

1. **Context-free interpretation:** The ontological commitment of each proposition should be **interpretable without** any other **contextual** data. This is achieved on one side by the global identification of individuals; on the other side it requires appropriate design of the ontology. E.g. an instance of a property "creator_birth_date" with domain "man-made object" cannot be interpreted without another property "creator"; a proposition "Martin.has role: buyer" cannot be understood without a sales event etc.. These would be bad models. The advantage of context-free propositions as intermediate step for data transformation and merging should be obvious: The global identifiers are the "fix points" around which directly related information can be compiled without other processing.

2. **Alternative views:** The model should be able to capture **multiple alternative propositions** about any fact, e.g. alternative birth dates for people whose actual birth date is debatable. This is mandatory for historical data. The example demonstrates also that this a design principle: The birth event must be explicitly modeled to render the intended meaning by context free propositions. The compilation of alternative propositions at well-defined points is a great help for subsequent reasoning. One of the more expressive examples of reasoning about historical contradictions is the Union List of Artist Names (ULAN) [23], which has tried to consolidate life data of more than 100.000 artists by compiling all alternative data and expert opinions, opinions on opinions etc.

3. **Appropriate granularity**: The model should make hidden concepts explicit to allow for extensions. If we want to integrate documentation about works of an artist with a report about his birth, the usual properties "birth_date","birth_place" are inappropriate. The hidden, intermediate concept "birth" should have been made explicit beforehand. Under this view, instantiation of a property "birth_date" is no more an elementary proposition. Indeed, as shown later, this notion of "elementary propositions" is not completely application independent for property instances (binary relations).

4. An ontology should require the **minimal ontological commitment** sufficient to support the intended knowledge sharing activities. This overlaps and competes with principle 2. and 3. above, however, we strictly avoid underspecification for that purpose, like the Dublin Core concept of "resource".

Finally it should be noted that virtually all metadata structures violate the above principles for reasons referred to in section 2.1.

3.2 Monotonicity

From the epistemological side, the addition of knowledge that is not in contradiction to existing knowledge is needed on both the categorical and the factual level, otherwise the integration of facts as they come in over time becomes a non-scalable task. Maintaining monotonicity is an important practical consideration in the design of ontologies, because the notion of what is in contradiction and what is not is grounded in the domain conceptualisation. For our purposes, the Open World Assumption is mandatory, because our knowledge changes and this must be taken into account. We view the impact of monotonicity on ontology design in three ways: classification, attribution (properties), and modelling constructs. We regard all those changes in a conceptual model as monotonic that do not invalidate previous instances of it.

3.2.1 Classification and Specialization

No complements: We have chosen to avoid defining any classes to be the complement of others, as such a class would change meaning with each new subclass found. During all our work, we could not find fundamental cultural concepts for which the complement is obvious. Even "male" is not clearly the complement of "female" (there are hermaphrodites, sexless, what else?). We don't know however, if this always holds. "Siblings" B_i, B_j are understood as not mutually exclusive, if not explicitly stated otherwise.

Preservation of classification: If an individual is once correctly classified according to a certain state of knowledge, additional non-contradictory knowledge in the sense of the experts' conceptualization should not invalidate its instantiation of this class, but may add an additional class. The use of multiple instantiation, e.g. to classify a willful destruction event with *E7Intentional Activity* and *E6 Destruction*, is essential to the CRM and supports preservation of existing classifications. Staying in this example: An event may be first recognized as a destruction. The willfulness of the event may be recognized at a later stage by other evidence. Or vice versa. Intentional activity does not imply destruction, nor does destruction imply an intentional activity. The creation of a class "Willful Destruction" does not offer any additional understanding.

An example of non-monotonic change of classification are the large Minoan terracotta vessels in Crete that Evans took for bath tubs – due to their striking similarity with modern ones. After enough had been found with bones in them, they were recognized as sarcophaguses. Had he classified them as container-like - the property he could really recognize - the additional knowledge would not have invalidated the previous classification. This argument is epistemological. It may come into conflict with ontological arguments, however we can design our ontology in a general enough way to support it. The principle presented here has worked well in the way we have used it, though obviously can not always hold sway.

3.2.2 Attribution

Whereas object-oriented design has provided us with an understanding of extension via specialization, the extension of the granularity of attribution seems to be rarely regarded. By that we mean the replacement of one property by a chain of properties and intermediate entities. The inverse operation, to reduce a path to a single property corresponds to the *join* operation in Relational Algebra and is well-defined and well understood. As pointed out in [2], this variable indirection or granularity of attribution is another major source of incompatibility between semantically overlapping descriptions. Such property paths are potentially infinite. One system may refer to the condition of an object as an assessment of the outcome of a number of measurements carried out by a number of people over a period of time. A 'poorer' system may not even refer to the assessment date and diagnosis, but simply register a term such as 'good' 'bad, or 'indifferent'. Such differences may be entirely justified by the intended use of the information in a given context. We have encountered numerous cases where radical differences in the granularity of information are justified by the intended purpose of the documentation.

In such cases, the CIDOC CRM models two paths, a direct and an indirect one, and characterises the "poorer", direct property as a ***short cut*** of the intermediate entity it "bypasses". The resulting CRM model thus appears to be redundant [2]. The idea is that collected factual knowledge would instantiate either the one or the other path. In order to be monotonic, a model must foresee a disciplined way to increase the indirection in data paths without losing the relationship to the coarser information. The intuitive ***short cut*** constructs introduced in the interdisciplinary CIDOC working group should be formalized in the future. In particular we are as yet unsure under which conditions reasoning described in section 4 is preserved by extending attribution paths.

3.2.3 Alternative Models

Finally, the monotonicity that can be achieved in practice may vary depending on modeling alternatives chosen. Our practical experience has not yet given us much guidance here, however we can present some examples instead:

Avoiding unconfirmed states: Many phenomena in history can be perceived as a chain of states and state transition events. There are sound logical theories dealing with such systems. This view is the basis of the ABC model [14], aiming like the CIDOC CRM to capture cultural contents. From an ontological point of view, the transitions can be produced out of the descriptions of the states and vice-versa. From an epistemological point of view, there is a huge difference: (1) If the information is incomplete, states and transitions cannot be transformed into each other. (2) States are difficult to be observed. That a property was valid over an interval of time and neither before nor after needs continuous complete observation. One can observe more easily a status, i.e. the validity of some properties at a point in time, or a transition event.

Under these considerations, the CIDOC CRM gives preference to modeling e.g. ownership changes rather than ownership states. It would result in a non-monotonic model to construct a set of states from any list of events, be they directly observed or not, as in the examples given by [14], because information about additional events may require deletion of existing states. The CIDOC CRM cannot claim to deal with the issue completely, mainly because it tries to restrict itself to the semantics found in a definite set of data structures. We so far propose to transform even a true (rare) observation of a state into transition events for normalization, which results in a slight loss of information. Nevertheless, the issue of introducing more elaborate models of states is under ongoing discussion.

View-neutrality: This principle has been described in detail in [2]. E.g. museums register accession (acquisition) and deaccession events. A transfer from one museum to another is an accession event for the one museum and a deaccession event for the other. Classification as "deaccession" or "accession" may be regarded as non-monotonic, if one allows for the respective change of context. In the CIDOC CRM we replace these notions by symmetric ones, like *Acquisition*, *Change of Custody.*

3.3 Global Coverage

When producing a standard, some attribute of validity is sought. With an extensible model in an open domain it is *a priori* difficult to say what a model covers, and if it has reached any definable stage of maturity. The approach proposed by Calvanese et.al [10] and others is open-ended. The enterprise model is incrementally improved to comprise more and more source model semantics, and we have basically followed the same procedure. In the process of taking more and more data structures into the scope of the CIDOC CRM however we have observed that the upper level becomes very stable, and new data structures typically introduce specializations "covered" by the model rather than "horizontal" extensions. This observation allows two things: (1) The definition of a compatibility attribute; (2) the definition of a standard.

We have designed CIDOC-CRM as a **common model** that contains or covers the intended meaning of all data structures used to encode "*information required for the scientific documentation of cultural heritage collections*", under certain semantic restrictions defined in [20]. For that purpose, the CRM group maintains a list a representative data structures [20], for which the coverage will be identified, some of which are actually contained in the CRM (e.g. [29], [30], [31]). With this claim, the CIDOC CRM is proposed as a standard reference model for the description of cultural heritage collections including the necessary concepts to communicate with library and archives contents. Note that by data structures we mean any database schema or formal document structure, be it Relational, object-oriented, XML-DTD or even an RDF Schema, used to describe primary data or metadata.

3.4 Designing a Manageable Unit

The creation of a standard ontology with limited resources in a reasonable time frame needs strict rules to partition the total of work one could do into functionally complete and manageable units. Such restrictions have been applied to (1) what meanings the contents should cover,(2) the modelling constructs and (3) the explicit rules formulation. In 1997 we identified the following intellectual aspects suitable to restrict the ontology without hampering its utility:

1. The **conceptual framework** (viewpoints) of the intended users: scholars, professionals in cultural heritage management, educators.
2. The **activities** intended to be supported : scientific documentation, research and the exchange of information with libraries and archives relevant to the documentation of cultural heritage collections.
3. The kinds of **objects** targeted at : objects in museums, libraries and archives.
4. The level of **detail** and **precision** required to provide an adequate level of quality of service.
5. Considerations of the necessary and manageable **technical complexity**.

By these criteria an **intended scope** has been formulated [20]. It excludes e.g. data only relevant for the internal management of a museum and not relevant for the exchange of knowledge between organisations. Still, these definitions are fairly fuzzy in practice. Therefore a practical scope is defined based on the semantics that can be identified in a list of existing data structures and are necessary for their coherent interpretation. This list is updated as the progress of work allows.

3.5 A property-driven design process

We have stressed properties (i.e. relations) in our methodology and ontology, and have required that, for the most part, classes are required to be either the domain or range of some. This is motivated by the fact that traditional museum data structures basically do the same. Fine granularity terminology is kept as variables in data fields. It seems not to be as relevant to the propositions data structures render as the properties themselves. This point may deserve further study.

Our property-centric approach has led to an empirical design process that we proposed to the CIDOC Documentation Standards Working Group in 1996. It was based on modeling experience with semantic network applications [5][38], and successfully applied to create the first version of the CIDOC CRM from the CIDOC Relational Model. Since then it has been loosely followed by the CRM group, but it has not yet been verified by independent groups. The driving force behind our process and ontology are the properties rather than the classes. This is contrary to the well-established Booch, Rational Rose [35] and other o-o design methodologies, yet we have found it far more productive for our purposes.

In the past, the CIDOC CRM had been extended several times. Some of these extensions are explicitly documented [41]. During extension, more general domains or ranges were sometimes assigned to preexisting properties. Such a change is monotonic as most of the last extensions have been. This observed behavior confirms the utility of the presented methodology, in particular of seeking minimal domains and ranges within the scope of the model.

4   About the CIDOC CRM contents

The CIDOC CRM contains classes and logical of groups of properties [43]. Those groups have to do with notions of participation, parthood and structure, location, assessment and identification, purpose, motivation, use etc. These properties have put *Temporal Entities* and with it events in a central place, as symbolically shown in Fig.3.
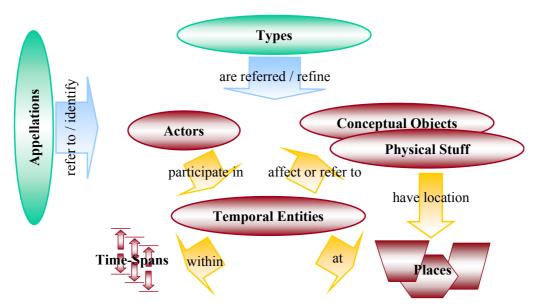

Figure 3: A qualitative metaschema of the CIDOC CRM

All property paths to dates go through temporal entities. Property paths to places that bypass temporal entities are understood as *short cuts* of temporal entities. Similarly, *Actors* are thought to relate to material and immaterial things (Physical Stuff, Conceptual Objects) only via temporal entities. Any instance of a class may be *identified* by *Appellations*, the names, labels, titles or whatever used in the historical context. We model the relation to names and its ambiguity as part of the historical knowledge acquisition process. This should not be confused with database identifiers in implementations of the Model, which are not part of the ontology. All class

instances can be classified in more detail by *Types*, for the additional terminological distinction, as described above. Frequently *Types* serve as the range of properties which refer in general to things of a certain kind, like "a dress made for a wedding" in contrast to the "dress made for my wedding". We present here some prominent logical groups of CRM properties.

4.1 Participation and Spatiotemporal Reasoning

As pointed out in [38], [2], [14], [44] and motivated by examples in this paper, the explicit modelling of events leads to models of cultural contents which can be better integrated. The participation or presence of several non-temporal entities in an event *e1* allows for a most important conclusion: They have been in the same time-interval and in the same space, even without knowledge of the particular time or space. They must have existed at that time. They have not been somewhere else at that time (with electronic communication, the space volume in which events occur can become very large, e.g. Earth to Moon). Culturally seen, the participants may have influenced each other, or, in the case of people, exchanged information. The events $e0_i$ of creation of each participant *i* have happened before or at the time of *e1*. The events $e2_i$ of destruction (or vanishing) of each participant have happened after or at the time of *e1*. These are nothing more than the well known *termini postquem* and *termini antequem* of chronological reasoning in historical research. Often this knowledge is more reliable than sequencing based on explicit date information. Therefore we try carefully to preserve such knowledge if it is primary (i.e. referred as such in a historical record or based on physical evidence).

The property *P11 had participants* denotes active or passive involvement of *Actors*, whereas *P12 occurred in the presence of* ranges from objects just being there (e.g. a desk where a treaty was signed) to use of tools, weapons, consumption of raw products, being produced. Specialization clarifies the more concrete senses modelled in the CIDOC CRM. Table 1 shows the full subproperty hierarchies as indented lists, each dash denoting another specialization level. By such generalization the normally implicit properties that enable temporal ordering of events become explicit and can be used in rules independent from further extension of the model.

Pid	Property Name	Domain	Range
P11	had participants (participated in)	E5 Event	E39 Actor
P14	- carried out by (performed)	E7 Activity	E39 Actor
P22	- - transferred title to (acquired title of)	E8 Acquisition	E39 Actor
P23	- - transferred title from (surrendered title of)	E8 Acquisition	E39 Actor
P28	- - custody surrendered by (surrendered custody)	E10 Transfer of Custody	E39 Actor
P29	- - custody received by (received custody)	E10 Transfer of Custody	E39 Actor
P95	- - has formed (was formed by)	E66 Formation	E74 Group
P96	- by mother (gave birth)	E67 Birth	E21 Person
P98	- brought into life (was born)	E67 Birth	E21 Person
P99	- dissolved (was dissolved by)	E68 Dissolution	E74 Group
P100	- was death of (died in)	E69 Death	E21 Person
P12	occurred in the presence of (was present at)	E5 Event	E70 Stuff
P13	- destroyed (was destroyed by)	E6 Destruction	E19 Physical Object
P16	- used object (was used for)	E7 Activity	E19 Physical Object
P24	- transferred title of (changed ownership by)	E8 Acquisition	E19 Physical Object
P25	- moved (moved by)	E9 Move	E19 Physical Object
P30	- transferred custody of (custody changed by)	E10 Transfer of Custody	E19 Physical Object
P31	- has modified (was modified by)	E11 Modification	E24 Physical Man-Made Stuff
P108	- - has produced (was produced by)	E12 Production	E24 Physical Man-Made Stuff
P34	- concerned (was assessed by)	E14 Condition Assessment	E18 Physical Stuff
P36	- registered (was registered by)	E15 Identifier Assignment	E19 Physical Object
P39	- measured (was measured)	E16 Measurement	E18 Physical Stuff
P94	- has created (was created by)	E65 Conceptual Creation	E28 Conceptual Object

Table 1: The CIDOC CRM property hierarchies P11 and P12.

The next notion relevant in this context are the properties *brought into existence, took out of existence* limiting the existence of things which have a persistent existence, i.e. which can be identified at different, separate times, as in the sentence: "I have seen him again after two years". These properties and their specializations connect the world lines of things with their terminating events. Even those events can be useful for temporal reasoning without explicit time: via participation of other things in the same event one can derive further *termini*. As we perceive events as continuous processes with non-zero extent and infinitely divisible, we argue that each item

participates partially in its creation. Therefore the respective specializations like **has created** etc. appear in both hierarchies:

Pid	Property Name	Domain	Range
P92	brought into existence (was brought into existence by)	E63 Beginning of Existence	E77 Existence
P94	- has created (was created by)	E65 Conceptual Creation	E28 Conceptual Object
P95	- has formed (was formed by)	E66 Formation	E74 Group
P98	- brought into life (was born)	E67 Birth	E21 Person
P108	- has produced (was produced by)	E12 Production	E24 Physical Man-Made Stuff
P93	took out of existence (was taken out of existence by)	E64 End of Existence	E77 Existence
P13	- destroyed (was destroyed by)	E6 Destruction	E19 Physical Object
P99	- dissolved (was dissolved by)	E68 Dissolution	E74 Group
P100	- was death of (died in)	E69 Death	E21 Person

Table 2: The CIDOC CRM property hierarchies P92 and P93.

The properties in tables 1 and 2 characterize the semantics of data structures in the cultural area. Fig. 4 shows an example of instantiating some of these properties, the legendary meeting of Pope Leo the Great with Attila the Hun, in Mantua. Even if the three dates may be wrong, the 4 deductions are true if the meeting has happened at all. Each death date constrains the meeting and both birth dates, the meeting date constrains both death and birth dates, etc. A maximum life-span assumed, any date constrains all others. Note that the CRM does not recognize points in time, only time-intervals. The deductions are not part of the model. They do not contribute to the compilation and integration of the primary data. They can be done by any other system at any other time.
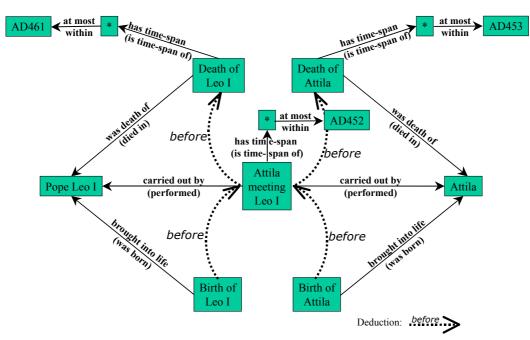

Figure 4: Pope Leo I meeting Attila

Note that any extension of the Model with another property that implies participation, e.g. "was injured in", would not be captured by the above reasoning in some implementations, unless it is an explicitly declared subproperty of P11. As subproperties are not supported by OMG models, it is not possible to implement such a feature in a simple way that is not affected by extension. Note also that the preservation of such a reasoning capability puts further constraints on "compatible extensions", which need more exploration.

4.2 Properties of Locating

The question "where is it" can be answered in natural language by relation to two different kinds of entities: geometric areas or objects. Examples of areas are: *in France, in Athens, 39N 124E*. Points given by spatial coordinates are typically understood as the centre of a wider, extended area. Objects can be in the proper sense ("bona fide objects",[45]), as: *on Queen Elizabeth (the ship), in my suitcase, at home,* or they can be landscape

and other features ("fiat objects",[45]), as: *on mount St Helens, at the Rhine river.* Following the CIDOC CRM, geometric areas (**E53 Place**) can only be defined relative to larger objects, including the surface of earth. Those objects in turn may be located at different times at different places (relative to a larger object). The cultural interest is in the relation to other things and not to an abstract absolute space. Absolute coordinates seem to make no sense when the reference objects move. As historical information is incomplete and sparse, and many reference objects move, normalization of place information in cultural databases to absolute coordinates should not replace the primary information, which is typically relative.

Any direct relation of an object to a place is seen as result of a move or a construction *in situ*, as with buildings. This view is a result of a longer discussion. The notion "place" is ambiguous in English, and gives rise to endless confusions in database design. In particular we take the position, that there is no image of a place, as it is not a material entity.
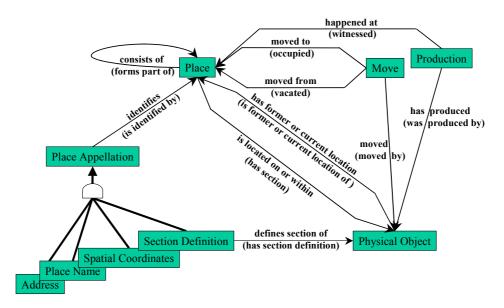
Figure 5: Properties of locating items

Places are identified by proper names or names referring to topological characteristics of object types, so-called "segments" [46] or **E46 Section Definition** in the CRM, like bow, head, neck, bottom etc.
Addresses in general need not be places, their function is often that of a contact point for some person or organisation (**P76 has contact points (provides access to)**), be they physical letterboxes or P.O. boxes. Fig. 5 shows the part of the CIDOC CRM dealing with location. The property **P88 consists of (forms part of)** from **Place** to **Place** is the normal part-of relation for areas. There are no minimal nor maximal area elements.

4.3 Notions of influence

The knowledge of what influenced or motivated a human activity and in turn the persistent things that have come upon us are culturally most relevant. We have not yet developed a systematic understanding of the different forms of influence and their mutual relations. Some are more physical, like using a mould or a tool. The influence of a mould on a produced object is strong and can often be verified on the object afterwards. The influence of a hammer is less specific. Similarly, making a copy of a painting has a strong influence on the product, copying the idea of a painting, a weak one. The latter is more an intellectual influence than a physical one. Further, activities are influenced by other activities, like orders, or just by the emotions they raise. If a real influence existed, a temporal sequence can be deduced. In contrast to "hard facts" as described in section 4.1, the notions described here vary over a continuum of stronger and weaker influence, which can be verified more or less easily afterwards. So far, the CRM contains the following properties of influence:

Pid	Property Name	Domain	Range
P15	took into account (was taken into account by)	E7 Activity	E28 Conceptual Object
P33	- used specific technique (was used by)	E11 Modification	E29 Design or Procedure
P16	used object (was used for) (mode of use : String)	E7 Activity	E19 Physical Object

Pid	Property Name	Domain	Range
P62	depicts object (is depicted by) (mode of depiction : Type)	E24 Physical Man-Made Stuff	E18 Physical Stuff
P63	depicts event (is depicted by) (mode of depiction : Type)	E24 Physical Man-Made Stuff	E5 Event
P65	shows visual item (is shown by)	E24 Physical Man-Made Stuff	E36 Visual Item
P67	refers to (is referred to by) (has type : Type)	E28 Conceptual Object	E1 CRM Entity
P70	- documents (is documented in)	E31 Document	E1 CRM Entity
P17	was motivation for (motivated)	E7 Activity	E19 Physical Object
P18	motivated the creation of (was created for)	E7 Activity	E28 Conceptual Object
P20	had specific purpose (was purpose of)	E7 Activity	E7 Activity

The properties P15,P33,P16 describe plans, prototypes and physical tools (moulds, hammers etc.) that assisted in or influenced an activity, and preexisted. These properties are used in particular in connection with *Modification*, *Production* and *Conceptual Creation* to model not only the influence on the process but also on the product, as with copies, prints etc.

The properties P62,P63,P65,P67,P70 describe an influence which can be manifested in the product without knowledge of the process. They can be seen as *short cuts* of the respective activities. Intended depictions and documentation of identifiable persons, objects, events, periods, ideas etc. play an extraordinary role in historical studies. All range values of these properties must have existed before the respective process which manifested them in the product.

The properties P17,P18,P20 describe an influence that originates in the activity itself, like orders, impressions, or emotions. P20 in particular captures sequences of planned activities. E.g. in the sentence, "George of Kyriaze orders a commemoration cross for donation to the Metropolitan Church of Ankara" [47], there is a *had specific purpose* relation between the order and the donation. All these notions deserve deeper analysis. Only for P15-P33 and P67-P70 could we establish subproperty hierarchies, an indication that the matter is relatively unexplored. Nevertheless they can normally be objectified and play a basic role in historical (as well as jurisdictional) reasoning. (at the time of publication of this paper, the properties of influence have been revised and a final form was decided, see [50] and respective minutes of the Working Group on the CIDOC CRM home page).

4.4 Applications

It would go beyond this paper to describe applications in detail. Several installations based on the CIDOC CRM have already been made [48]. A recent test together with CIMI [49] aimed at demonstrating that the semantics of heterogeneous museum records are preserved under the CIDOC CRM. Two examples were interesting: The Clayton collection of the Natural History Museum in London and the Australian Museums On-line (AMOL) initiative both use flat records in ACCESS databases. The Clayton collection describes a complex relation between plant specimens, initial and current classification events and classification documents. These records can be automatically transformed to CIDOC CRM instances because of the clear semantics of their fields. The AMOL data were easily transformed to CIDOC CRM instances by hand, but not automatically, because their fields are designed more for formatting the presentation. The examples demonstrate two things: data structures (like the Clayton data) need not implement the complexity of an ontology for information integration in order to be interpretable; however, an ontology can help to create interpretable data structures. More such tests will be carried out in the near future.

5   Conclusion

In this paper we have presented an ontology for information integration in culture, and we have tried to justify by the intended functionality a methodology and design. We assume that the applied methodology and the more abstract levels of the model have a wider validity. The presented ontology is a result of ongoing work, and future work will also address more advanced formalisations.

The CIDOC CRM has achieved a relatively high degree of maturity and completeness in capturing the conceptualisations behind the data structures in its envisaged scope, as recent extensions of scope and data transformation tests confirm. The purpose is information integration, but not the further reasoning like

reconstruction of a possible truth. It intends however to allow gathering all necessary information in a suitable form for such further reasoning. It is sufficiently comprehensive for the domain expert, so that a broad consensus on the correct ontological commitment could be achieved, and the ontology was accepted by the ISO as a candidate international standard for cultural heritage information.

The methodology presented here has proven to be applicable in an interdisciplinary group, and our experience in training non-experts in basic KR principles has been very encouraging. The complexity of the domain is intriguing. Philosophical considerations and long discussions were necessary to clarify the role of the modelled knowledge with respect to the working concepts of the domain experts. Without such clarifications, no consensus on the relevant concepts could be achieved. This thinking was new for both sides, the computer scientists and the domain experts, as it is not needed for either work in isolation. It was interesting to learn that not all domain concepts are equally suited as basis for information integration.

The methodology presented here appears to be contrary to well-known o-o methodologies for designing the controlling software of information systems. We wish to point out that there may be a qualitative difference, even though some researchers take ontologies for software products (see for example the WebODE article in this issue). Analysis of the semantics behind data structures for information integration is an ontological problem. This paper tries to illustrate ontology from a point of view seldom taken: the *relationships* between entities as a driving force for the logical structure, more than the nature of involved individuals. This seems to be appropriate to analyse data structures in contrast to terminological systems. A coherent analysis of (non-unary) properties is mandatory for information integration, even more than detailed entity analysis, in particular if one separates the epistemological issue of correcting erroneous input data from the ontological issue of classifying already correct information.

Historical knowledge, to our understanding independent of the specific domain, seems to reveal in this work a character, which is quite distinct from engineering knowledge in a rather subtle way. Even though in our conceptualisation of reality we do not distinguish between past, present and future, the way knowledge is acquired, its quantity and quality, is completely different for the past. We argue therefore that the design of conceptual models to capture the past must be governed far more by epistemological arguments than engineering models. The nature of historical knowledge, the relation between reality, a perceived historical reality and the form of knowledge we can acquire seems to be an interesting topic for further investigation.

The CIDOC CRM is envisaged to become an ISO standard in 2004. In parallel to the standardization work, we intend to engage in more validation experiments and in research on the open theoretical and intellectual issues. A general theory of extensibility for such an ontology under the preservation of certain reasoning capabilities would be very helpful. As discussed in section 4.1, subproperties play a crucial role for that. From the point of contents, the CIDOC CRM still touches only very fundamental concepts, and many extensions will be useful to allow for more reasoning, like temporality of properties, phases of objects, a coherent model of influence, modelling performing arts etc. We see also a need to clarify philosophical questions of foundational character about the nature of the knowledge we describe.

Acknowledgement:

6 References

[1] T.Baker, "A Grammar of Dublin Core", D-Lib Magazine, 6(10), 2000.

[2] Martin Doerr and Nicholas Crofts "Electronic Esperanto: The Role of the Object Oriented CIDOC Reference Model", Proc. of the ICHIM'99, Washington, DC, September 22-26, 1999.

[3] The CIDOC CRM Home Page, 2001, http://cidoc.ics.forth.gr/what_is_crm.html

[4] Nick Crofts, Ifigenia Dionissiadou, Martin Doerr, Matthew Stiff, "Definition of the CIDOC object-oriented Conceptual Reference Model", Version 3.2.1, ISO Working Document ISO/TC46/SC4/WG9/N2, July 2001, http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3.2.1.rtf

[5] I.Dionissiadou, M.Doerr, "Mapping of material culture to a semantic network", in : Automating Museums in the Americas and Beyond, Sourcebook, ICOM-MCN Joint Annual Meeting, August 28-September 3, 1994

[6] Panos Constantopoulos, "Cultural Documentation: The CLIO System", Technical Report FORTH-ICS/TR-115, 12 pages, January 1994.

[7] Martin Doerr, "CIDOC Conceptual Reference Model, Correlation Test Project , Results", June 2001, http://cidoc.ics.forth.gr/testproject_results.html

[8] S. Bergamaschi, S. Castano, S. De Capitani De Vimercati, S. Montanari and M. Vincini, "An Intelligent Approach to Information Integration", International Conference on Formal Ontology in Information Systems (FOIS98), Trento 1998. IOS-Press (Amsterdam)

[9] Foundations of Data Warehouse Quality (DWQ) European ESPRIT IV Long Term Research (LTR) Project 22469, 1996-1999, http://www.dbnet.ece.ntua.gr/~dwq/

[10] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, and R. Rosati, "Description Logic Framework for Information Integration";, In Proc. of the 6th Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'98), 1998, pages 2-13

[11] Gio Wiederhold, "Mediators in the Architecture of Future Information Systems", in : IEEE Computer, March 1992.

[12] Guarino N. "Formal Ontology and Information Systems". In N. Guarino (ed.), Formal Ontology in Information Systems. Proc. of the 1st International Conference, Trento, Italy, 6-8 June 1998. IOS Press

[13] Thomas R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", in: Formal Ontology in Conceptual Analysis and Knowledge Representation, edited by Nicola Guarino, Roberto Poli, Kluwer, 1994

[14] Carl Lagoze, Jane Hunter, "The ABC Ontology and Model", DC-2001, International Conference on Dublin Core and Metadata, Tokyo, October 2001, http://metadata.net/harmony/lagoze_hunter_dc2001.pdf

[15] The Art Museum Image Consortium AMICO, http://amico.org, AMICO data dictionary version 1.3, http://amico.org/AMICOlibrary/dataDictionary.html

[16] Chryssoula Bekiari, Christina Gritzapi, Dimitrios Kalomoirakis "POLEMON: A Federated Database Management System for the Documentation, Management and Promotion of Cultural Heritage", Proc. of the 26th Conference on Computer Applications in Archaeology, Barchelona, March 24-28, 1998.

[17] Chryssoula Bekiari, Panos Constantopoulos & Theodosia Bitzou " DELTOS : A Documentation System for the Antiquities and Preserved Buildings of Crete, Requirements Analysis",   Technical Report FORTH-ICS/TR-60, October 1992. Available in Greek.

[18] "SPECTRUM: The UK Museum Documentation Standard,", second Edition, Museum Documentation Association mda(Europe), Cambridge, United Kingdom,1997-1998, http://www.mda.org.uk/spectrum.htm

[19] "International Guidelines for Museum Object Information: The CIDOC Information Categories", published by CIDOC in June 1995. http://www.cidoc.icom.org/guide/guide.htm

[20] Nicholas Crofts et.al, "CRM Scope Definition", Proposal of the Steering Committee of the CIDOC CRM SIG, July 7, 2001. http://cidoc.ics.forth.gr/crm_scope_definition.html

[21] Nicholas Crofts et.al., "Notes on the data modelling meeting in Crete July 1997", July 1997, http://cidoc.ics.forth.gr/docs/notes_data_modelling_1997_crete.doc

[22] Fernández, M. "Overview of Methodologies for Building Ontologies". Workshop on Ontologies and Problem-Solving Methods: Lessons Learned and Future Trends. (IJCAI99). August. 1999.

[23] J.M.Bower,M.Baca et al., "Union List of Artist Names - A User's Guide to the Authority Reference Tool", Version 1.0, Getty Art Information Program, G.K.Hall, New York, 1994

[24] N. Guarino,P. Giaretta, "Ontologies and Knowledge Bases, Towards a Terminological Clarification", in N.J.I.Mars (ed.), Towards Very Large Knowledge Bases, IOS Press 1995, Amsterdam.

[25] J. Mylopoulos, A. Borgida, M. Jarke, M. Koubarakis, "Telos: Representing Knowledge about Information Systems", ACM Transactions on Information Systems, October 1990.

[26] Anastasia Analyti, Nicolas Spyratos & Panos Constantopoulos, "On the Semantics of a Semantic Network", Fundamenta Informaticae 36 (1998), pp. 109-144, IOS Press.

[27] G. Karvounarakis, V. Christophides, D. Plexousakis, S.Alexaki, "Querying Querying RDF Descriptions for Community Web Portals". In Proc. of the 17 France National Conf. on Databases BDA, 29 Octobre - 2 Novembre 2001, Agadir, Maroc., http://139.91.183.30:9090/RDF/publications/sigmod2000.html

[28] ICS FORTH, Information Systems Laboratory, Heraklion, Crete, Greece, "The Semantic Index System - SIS", http://www.ics.forth.gr/proj/isst/Systems/sis.html

[29] M. Theodoridou, M. Doerr, Mapping of the Encoded Archival Description DTD Element Set to the CIDOC CRM , Technical Report FORTH-ICS/TR-289, June 2001, http://www.ics.forth.gr/proj/isst/Publications/paperlink/ead.pdf

[30] Martin Doerr, Mapping of the AMICO data dictionary to the CIDOC CRM , Technical Report FORTH-ICS/TR-288, June 2001, http://cidoc.ics.forth.gr/docs/mappingamicotocrm.rtf

[31] Martin Doerr, Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM , Technical Report FORTH-ICS/TR-274, July 2000, http://cidoc.ics.forth.gr/docs/dc_to_crm_mapping.rtf

[32] S.R. Ranganathan, "A descriptive account of Colon Classification", Bangalore: Sarada Ranganathan Endowment for Library Science. 1965.

[33] Martin Doerr et.al., "Notes on the transformation of the CIDOC relational data model", July 1996, http://cidoc.ics.forth.gr/docs/notes_trans_cidoc.rtf.

[34] Paul Feyerabend, "Three Dialogues on Knowledge", Basil Blackwell 1991

[35] T. Quatrani, Visual Modeling with Rational Rose and UML, Addison-Wesley,1998.

[36] John L. Schnase, "Semantic Data Modelling of Hypermedia Associations", in: ACM Transactions on Information Systems, Vol.11,No.1, January 1993, p 45

[37] Michael Erdmann, Rudi Studer, "Ontologies as Conceptual Models for XML Documents", research report, Institute AIFB, University of Karlsruhe, 1999.

[38] Maria Christoforaki, Panos Constantopoulos, Martin Doerr, "Modelling occurences in cultural documentation", Proc. of the III Convegno Internazionale di Archeologia e Informatica, Roma, November 22-25, 1995. http://www.ics.forth.gr/proj/isst/Publications/paperlink/Model_occur_in_cultur_doc.ps.gz

[39] B. Tversky, Kathleen Hemenway, "Objects, Parts, and Categories", in Journal of Experimental Psychology: General, Vol.113,No2, June 1984, pp169-193.

[40] "EAD Tag Library for Version 1.0, Encoded Archival Description (EAD)Document Type Definition (DTD)", Version 1.0, Technical Document No. 2, June 1998. Published by the Society of American Archivists and theLibrary of Congress (http://lcweb.loc.gov/ead/tglib/tlhome.html)

[41] Martin Doerr (ed.), "Agios Pavlos Extensions - Add-ons for the Completion of the CIDOC CRM", July 2000, http://cidoc.ics.forth.gr/docs/agios_pavlos_extensions.rtf

[42] Nick Crofts, Ifigenia Dionissiadou, Martin Doerr, Matthew Stiff (ed.), "Definition of the CIDOC object-oriented Conceptual Reference Model", Version 3.2, July 2001, ISO/TC46/SC4/WG9/3, http://cidoc.ics.forth.gr/docs/cidoc_crm_version_3.2.rtf

[43] Nick Crofts, Ifigenia Dionissiadou, Martin Doerr, Pat Reed (editors), "CIDOC Conceptual Reference Model - Information groups", ICOM/CIDOC Documentation Standards Group, September 1998, http://cidoc.ics.forth.gr/docs/info_groups.rtf

[44] INDECS Home Page: Interoperability of Data in E-Commerce Systems, http://www.indecs.org

[45] B. Smith, A. Varzi, "Fiat and Bona Fide Boundaries: Towards on Ontology of Spatially Extended Objects." International Conference COSIT'97, October 15-18, 1997, Proceedings. Lecture Notes in Computer Science, Vol. 1329, Springer, 1997, pp103-119

[46] P.Gerstl, S.Pribbenow, "A conceptual theory of part – whole relations and its applications", Data & Knowledge Engineering 20 305-322, 1996, North Holland- Elsevier

[47] M. Doerr, I. Dionissiadou, "Data Example of the CIDOC Reference Model - Epitaphios GE34604 –" October 2, 1998, http://cidoc.ics.forth.gr/docs/crm_example_1.doc

[48] N. Crofts "Implementing the CIDOC CRM with a relational database" in MCN Spectra. 24 (1) Spring, 1999

[49] John Perkins, "ABC/Harmony CIMI Collaboration Project", September 30th, 2000, http://www.cimi.org/public_docs/Harmony_long_desc.html

[50] Nick Crofts, Martin Doerr, Tony Gill, Stephen Stead, Matthew Stiff (editors), **Definition of the CIDOC object-oriented Conceptual Reference Model, Version 3.4** , November 2002 http://zeus.ics.forth.gr/cidoc/docs/cidoc_crm_version_3.4.rtf