The CIDOC Conceptual Reference Module

An Ontological Approach to Semantic Interoperability of Metadata

Martin Doerr

■ This article presents the methodology that has been successfully used over the past seven years by an interdisciplinary team to create the International Committee for Documentation of the International Council of Museums (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 International Organization for Standardization (ISO) standard. This article justifies in detail the methodology and design by functional requirements and gives examples of its contents. The CIDOC CRM analyzes 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.

The creation of the World Wide Web has had a profound impact on the ease with which information can be distributed and presented. This ease 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 organizations*. Quite naturally, their development has focused on presentation, such as web sites 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 helpful in dealing effectively with the large amounts of information coming together on the internet. Information about cultural heritage poses particular challenges for formal handling—not, as often assumed, because it is ill defined but because it is highly diverse, and the incompleteness of information about the past is intrinsic. To date, most attempts for semantic interoperability have concentrated on the development and standardization of a shared core data structure (for example, the International Committee for Documentation of the International Council of Museums [CIDOC] RELATIONAL MODEL) and a terminology system. In case a common data structure seemed to be impossible, at least a common metadata schema such as "finding aids" has 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 and Architecture Thesaurus* are characteristic examples of standards in the United States and beyond, for which equivalents in several other languages have been created.

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 such as the DUBLIN CORE represent a common denominator by far too small to fulfill advanced requirements. Overstretching its already limited semantics to capture complex contents leads to further loss of meaning ("metadata pidgin" [Baker 2000]), even though most of the contents encoded in the various structures seem to be pretty comprehensive in commonsense terms and are often interrelated. I make the hypothesis that much of the diversity of data and metadata structures is because they are designed for data capturing—to guide good practice of what should be documented and to optimize coding and storage costs for a specific application—far more than for interpreting data. Necessarily, these data structures are relatively flat (to suggest a work flow 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) (Crofts 2001; Doerr and Crofts 1999),[1] an attempt by the CIDOC Committee of the International Council of Museums (ICOM) to achieve semantic interoperability of museum data. Work started in the beginning on a more intuitive base, from a knowledge representation model (Dionissiadou and Doerr 1994; Constantopoulos 1994), working from the consensus of a varying team of domain experts and based on strict intellectual principles. This work received wide acceptance by CIDOC and other relevant stakeholders in the domain, and in September 2000, the CIDOC CRM was successfully submitted as a new work item to ISO (TC46). It is now registered as ISO/CD 21127 and is expected to become an ISO standard in 2004. It is now in a very stable form and contains 80 classes and 130 properties, both arranged in multiple is-a hierarchies. Several applications (Doerr 2001a) and comparison with related work improves the theoretical understanding of the work that has been done and is still ongoing.

Instead of seeking a common schema as a prescription for data capturing, which would supposedly ensure semantic compatibility of the produced data, I agree with the CIDOC CRM what Bergamaschi et al. (1998) call a semantic approach to integrated access. Convinced that information collection is already done well by the existing data structures, possible improvements not withstanding, I aim simply for read-only integration, such as in the Foundations of Data Warehouse Quality Project;[2] also see Calvanese et al. 1998. Read-only integration comprises data migration, data merging (materialized data integration as in data warehouse applications), and virtual integration using query mediation (Wiederhold 1992). Recently, more and more projects and theoreticians support the use of formal ontologies as common conceptual schema for information integration (Bekiari, Gritzapi, and Kalomoirakis 1998; Bergamaschi et al. 1998; Gruber 1993; Guarino 1998; Lagoze and Hunter 2001) to provide a conceptual basis for understanding and analyzing existing (meta)data structures and instances, give guidance to communities beginning to examine and develop descriptive vocabularies, and develop a conceptual basis for automated mapping among metadata structures and their instances (Lagoze and Hunter 2001).

It 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 conceptualizations of the respective domain experts, whereas the data-entry structures themselves often seem to resist merging. I have followed a pragmatic approach to separate a kind of top-level ontology (Guarino 1998), which represents knowledge extracted from schemata and data structures, from pure terminology. This approach was utilized to keep the basic ontology in a manageable size. The semantics of the data structures are richer in n-ary relationships (attributes, properties, and so on) than in fine distinctions between classes, whereas thesauri are just the opposite —they build rich is-a (BT/NT) hierarchies but typically use only one (RT) relationship for any other internal conceptual relation. I 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 Doerr and Crofts [1999]). 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. We call such a model a *property-centric ontology* to stress the specific character and function.

This article presents the CIDOC CRM from a methodological point of view. It relates the intended scope and functions to the ontological principles that governed its design, presents its key concepts, and positions the model with respect to relevant related work. I expect this methodology to be applicable to other domains, even though there is no experience yet to support this claim. The significant contributions 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 the 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.

The Problem

Museum information virtually describes the whole world as manifested in material objects from the past. In this section, I explain that this information is so diverse that no single data structure can even capture "core data" so that one can access the relevant knowledge relating different information assets and objects.

The Necessity of Data Structure Diversity

Let us regard here both data structures for long-term storing of data, as database schemata, but also tagging schemes such as 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 through formal queries, the complexity the user can handle, the complexity of the system the user can afford to implement or pay for, and the cost of learning these structures and filling them with contents. Because most applications run in a relatively uniform environment—a library, a museum of modern art, a historical archive of administrative records, a paleontological museum—much of the complexity of the one application is negligible for another, allowing 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, but 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 or the reservation of dozens of special fields and storage space that will never be used. Nevertheless—a nd this is the crucial point—the notion of date and dating for both experts is completely compatible; there is no difference in conceptualization, at least from a scientific point of view. The complexity typical for archaeology hardly ever occurs in modern art, however. Similarly, the documentation of an historical building implies a complex history with various phases and persons, whereas paintings are typically created in a relatively compact process. Consequently, art documentation schemes such as the Art Museum Image Consortium (AMICO)[3] do not capture multiple creators for multiple phases of objects, in contrast to the architectural descriptions set up by the Greek National Archive of Monuments (Bekiari, Constantopoulos, and Bitzou 1992; Bekiari, Gritzapi, and Kalomoirakis 1998). The rare exceptions to such simplifications are typically 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, such as 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 that lists all involved persons but not 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 such as DUBLIN CORE that add up relevant persons, relevant dates, and so on, without interrelations. These simplifications actually violate the conceptualization; the sources retrieved on this basis can only be sorted out by reading them. How long this approach 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 conceptualization behind these simplifications, if at all possible, or improve our data structures (see the Applications subsection later).

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, the event officially designating the end of World War II. One can hardly find a better documented event in history. I created the demonstration metadata in figure 2 from the information I found associated with the objects:

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

Figure 1. A Possible DUBLIN CORE Record for the Yalta Agreement.

The United States State Department 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…."[4] A DUBLIN CORE record might be as shown in figure 1.

The Bettmann Archive in New York held a world-famous photograph of the Yalta conference that features Winston Churchill, Franklin Delano Roosevelt, and Josef Stalin. A DUBLIN CORE record for this photo might be as shown in figure 2.

The striking point is that both metadata records have nothing more in common than the date *1945*, hardly a distinctive attribute. An "integrating" piece of information comes from the *Thesaurus of Geographic Names* (TGN),[5] which can be captured by the metadata in figure 3.

The keyword *Crimea* can be found under the foreign names for *Krym*, that is, using another record (id = 1003381). Figure 2 demonstrates a fundamental problem: To retrieve information related to one specific subject, information from multiple sources must be integrated. Vocabulary and data structure unification alone do not solve the problem.

Requirements

One problem with the Yalta example is the ability 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, and so on. A deeper problem is the fact that the artifacts do not fit our question: People document persistent items such as images, texts, and places, but our question was about an event, here the Yalta Conference, something that is only indirectly preserved in these items. The data structures express certain relationships between items, which might or might not be identified globally. Other relations are hidden, such as UPI taking pictures, which can either be guessed from the context or must be recovered from secondary sources or background knowledge (at these times, the press photographers were not documented). Having argued earlier that data structures are full of simplifications and hidden constants, we see that the main problem is to recover this information during data integration, which is where ontologies are most valuable.

When work was started on the CIDOC CRM in 1996, the CIDOC working groups had virtually given up on creating one standard data structure for all museums. The working groups

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 2. Allied Leaders at Yalta.

assumed (and still assume) that a fairly small set of good practice guides (for example, Museum Documentation Association [MDA] SPECTRUM [MDA 1998] and CIDOC International Guidelines for Museum Object Information and standard data structures already express well what museum professionals want and should say about their objects in various disciplines (an extended list can be found in Crofts et al. [2001]),[6] albeit these documents can still be improved. In particular, the necessary constraints to improve data integrity are typically applied locally at data entry time.

The global interoperability between disciplines is clearly needed for the following functions after data have been created: the mediation of global queries to local structures (Wiederhold 1992), the extraction of individual statements from larger units of documentation and their comparison for alternative opinions, and the transformation of data for migration to other systems and merging into more informative units (such as 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 previously defined functions 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 Guarino (1998). Parallel to the ongoing work, more and more methodological principles were elaborated and applied on the basis of these fundamental requirements. The working group is now in the position to give a consistent account of this methodology, which to date has been documented only in presentation slides and minutes of the working group.[7]

About the CRM Methodology

The problems computer scientists and system implementers have in comprehending the logic of cultural concepts seems to be equally as notorious as the inability of the cultural professionals to communicate these concepts to computer scientists. The CIDOC CRM working group is therefore interdisciplinary, aiming at closing this gap. People with a background in museology, history of arts, archaeology, natural history,

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 the south shore of the Crimean Peninsula, site of the conference between Allied powers during World War II in 1945. It is a vacation resort noted for pleasant climate and coastal and mountain scenery; it produces wine, canned fruit, and tobacco products.
Source: *Thesaurus of Geographic Names.*

Figure 3. Parts of the TGN Record for Yalta.

physics, computer science, philosophy, and others were involved. We have achieved a functional compromise between the complexity of the conceptualizations and the complexity of formalism the participants would appreciate. Therefore, the AI reader might miss in this work some obvious formalization. We could also convey in a series of targeted seminars more knowledge representation principles to nonexperts 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 a domain in a way that can guide implementers and computer scientists and can later be refined by domain specialists. Therefore, many good contributions (for example, modeling 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 these contributions are mentioned in this article. Also, strict neutrality with respect to commercial interest groups is the declared policy of CIDOC. Thus, CIDOC was able to create a construct of high intellectual quality and coherence. Starting from an initial formulation of the scope of the CIDOC CRM,[8] the working groups independently developed intellectual principles similar to those in Gruber (1993).

The CIDOC CRM is a formal conceptual model; however CIDOC CRM instantiation data (factual knowledge) are allowed to be contradictory. Historical data—any description made in the past about the past, be it scientific, medical or cultural—are typically unique and cannot be verified, falsified, or completed in an absolute sense. In history, any conflict resolution of contradictory records is nothing more than another opinion. Thus, 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 (for example, 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 might have been from the same pot or two different pots) or similarly if one has been two (see, for example, the *Union List of Artist Names* [Bower and Baca 1994]). 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.

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 (Guarino 1998), and regard both as a special kind of knowledge base (Guarino and Giaretta 1995). I prefer the term *conceptual model* when talking about actual instantiation and constructs dictated more by the representation formalism than the intended meaning. Categorical knowledge can come from the analysis of data structures, hidden constants or terminology used in the data. I 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 4 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 such as the *Thesaurus of Geographic Names* and other authorities.

The CIDOC CRM plays the role of an *enterprise model* (Wiederhold 1992); in the following discussion, it is called a *common model*. We assume for all sources that a conceptual model exists (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 can be restricted to the semantics falling within the scope of the common model. As a representation formalism, the working group selected the TELOS data model (Analyti, Spyratos, and Constantopoulos 1998; Mylopoulos et al. 1990) without its assertional language. TELOS, like many other knowledge representation languages, decomposes knowledge into elementary propositions—declarations of individuals, classes, and binary relations.

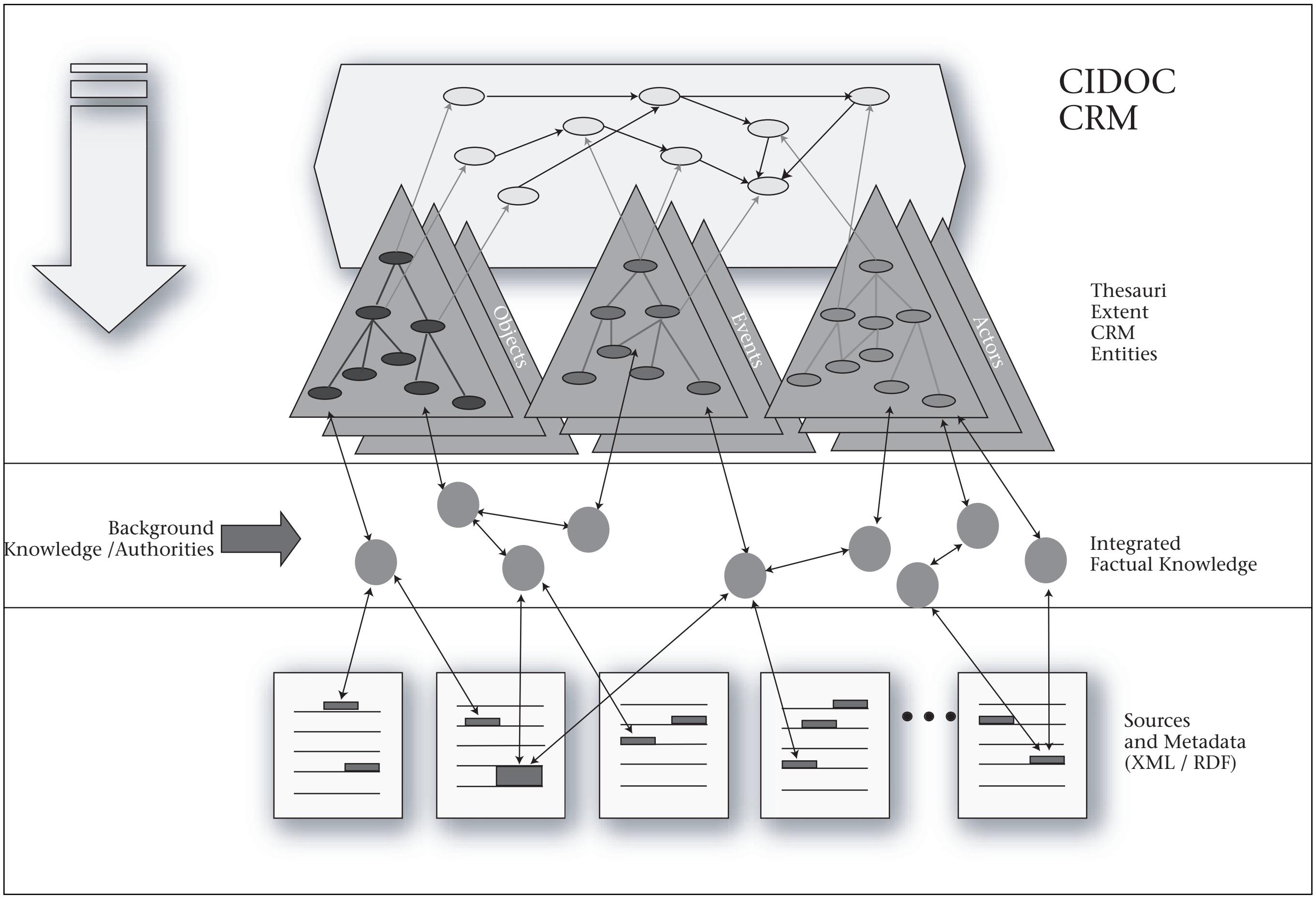

Figure 4. An Information Integration Architecture.

The properties of TELOS relevant for the purpose of this article are similar to those of RDF and RDFS (Karvounarakis et al. 2001). Because RDF (and OWL) are now on the way to becoming standards for the applications our group targets, we here use the terminology of RDFS because it might be more familiar than that of TELOS and talk about classes and properties. Because 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.[9] Logical assertions are omitted because they can be added at a later stage, once the ontological commitment of the primitive classes, properties, and is-a relations are set up satisfactorily.

The process of instantiating the common model with factual knowledge can be broken into two 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) and (2) the instantiation of the properties (roles, relationships) of the common model with relations that connect those individuals 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.

Context-free interpretation: The ontological commitment of each proposition should be interpretable without any other contextual data. This is achieved on the one hand by the global identification of individuals and on the other side by appropriate design of the ontology. For example, an instance of a property *creator_birth_date* with domain *manmade object* cannot be interpreted without another property *creator*; a proposition *Martin.has role: buyer* cannot be understood without a sales event, and so on. These models would be bad. The advantage of context-free propositions as an 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.

Alternative views: The model should be able to capture multiple alternative propositions about any fact, for example, alternative birth dates for people whose actual birth date is debatable. This is mandatory for historical data. The example also demonstrates that this is a design principle: The birth event must be modeled explicitly 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* (Bower et al. 1994), which has tried to consolidate the life data of more than 100,000 artists by compiling all alternative data and expert opinions, opinions on opinions, and so on.

Appropriate granularity: The model should make hidden concepts explicit to allow for extensions. To integrate documentation about works of an artist with a report about his/her birth, the usual properties of birth_date and 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 not an elementary proposition. Indeed, as shown later, this notion of *elementary propositions* is not completely application independent for property instances (binary relations).

An ontology should require the minimal ontological commitment sufficient to support the intended knowledge-sharing activities (which overlaps and competes with the previous two principles); however, I strictly avoid underspecification for this purpose, such as 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 The Necessity of Data Structure Diversity.

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 levels; otherwise, the integration of facts as they come in over time becomes a nonscalable 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 conceptualization. For our purposes, the open-world assumption is mandatory because our knowledge changes, which must be taken into account. I view the impact of monotonicity on ontology design in three ways: (1) classification, (2) attribution (properties), and (3) modeling constructs. I regard all these changes in a conceptual model that do not invalidate previous instances of it as monotonic.

Classification and Specialization

No complements: I chose to avoid defining any classes as the complement of others because such a class would change meaning with each new subclass found. During all the efforts of the working group, 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, and so on). We don't know, however, if this situation 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 correctly classified according to a certain state of knowledge once, additional non-contradictory knowledge in the sense of the experts' conceptualization should not invalidate its instantiation of this class but can add an additional class. The use of multiple instantiation, for example, to classify a willful destruction event with E7 Intentional Activity and E6 Destruction, is essential to the CRM and supports preservation of existing classifications. With this same example, an event can first be recognized as a destruction. The willfulness of the event can be recognized at a later stage by other evidence, or vice versa. Intentional activity does not imply destruction; destruction does not imply an intentional activity. The creation of a class Willful Destruction does not offer any additional understanding.

An example of nonmonotonic change of classification is the large Minoan terra-cotta vessels in Crete that Sir Arthur Evans (excavator of Knossos) took for bath tubs—because of their striking similarity with modern ones. After enough vessels 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 can come into conflict with ontological arguments, but we can design our ontology in a general enough way to support it. The principle presented here has worked well in the way the working group has used it, although it obviously cannot always hold sway.

Attribution

Whereas object-oriented design has provided us with an understanding of extension using specialization, the extension of the granularity of attribution seems to rarely be regarded. By that, I 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 Doerr and Crofts (1999), 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 can 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 might not even refer to the assessment date and diagnosis but simply register a term such as *good, bad,* or *indifferent*. Such differences can entirely be justified by the intended use of the information in a given context. I 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 characterizes the "poorer" direct property as a shortcut of the intermediate entity it bypasses. The resulting CRM model thus appears to be redundant (Doerr and Crofts 1999). The idea is that collected factual knowledge would instantiate either one or the other path. 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 shortcut constructs introduced in the interdisciplinary CIDOC working group should be formalized in the future. In particular, working group members are as yet unsure under which conditions reasoning, as described in About the CIDOC CRM Contents is preserved by extending attribution paths.

Alternative Models

Finally, the monotonicity that can be achieved in practice can vary depending on the modeling alternatives chosen. Our practical experience has not yet given us much guidance, but I can present some examples.

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 (Lagoze, and Hunter 2001), which like the CIDOC CRM, aims to capture cultural contents. From an ontological point of view, the transitions can be produced from the descriptions of the states and vice versa. From an epistemological point of view, there is a huge difference: First, if the information is incomplete, states and transitions cannot be transformed into each other. Second, states are difficult to observe. That a property was valid over an interval of time and neither before nor after needs continuous complete observation. One can more easily observe a status, that is, the validity of some properties at a point in time, or a transition event.

Under these considerations, the CIDOC CRM gives preference to modeling, for example, ownership changes rather than ownership states. It would result in a nonmonotonic model to construct a set of states from any list of events, be they directly observed or not, as in the examples given by Lagoze and Hunter (2001), because information about additional events can 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. To date, working group members 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 undergoing further discussion.

View neutrality: This principle has been described in detail in Doerr and Crofts (1999). For example, 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 can be regarded as nonmonotonic if one allows for the respective change of context. In the CIDOC CRM, we replace these notions by symmetric ones, such as acquisition and change of custody.

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. (1998) 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, working group members have observed that the upper level becomes 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 and (2) the definition of a standard.

The working group 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 Crofts et al.[10] For this purpose, the CRM group maintains a list a representative data structures,[11] for which the coverage will be identified, some of which are actually contained in CRM (Doerr 2001b, 2000; Theodoridou and Doerr 2001). 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 content. Note that by data structures, the working group means any database schema or formal document structure, be it relational, object-oriented, XML-DTD, or even RDF SCHEMA used to describe primary data or metadata.

Designing a Manageable Unit

The creation of a standard ontology with limited resources in a reasonable timeframe 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) the meanings the contents should cover, (2) the modeling constructs, and (3) the explicit rules formulation. In 1997, the working group 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: objects in museums, libraries, and archives; (4) the level of detail and precision required to provide an adequate level of quality of service; and (5) considerations of the necessary and manageable technical complexity.

With these criteria, an intended scope was formulated (Crofts et al. 2001). It excludes, for example, data that are only relevant for the internal management of a museum and not relevant for the exchange of knowledge between organizations. 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.

A Property-Driven Design Process

The working group stresses the primary role of properties (that is, relations) in its methodology and ontology and has required that, for the most part, classes are to be either the domain or range of some property. This is motivated by the fact that traditional museum data structures basically do the same. They encode classes only if they are needed to encode an attribute or relationship. Fine-granularity terminology is kept as variables in data fields. It seems to not be as relevant to the propositions data structures render as to the properties themselves. This point might deserve further study.

This property-centric approach has led to an empirical design process that was proposed to the CIDOC Documentation Standards Working Group in 1996. It was based on modeling experience with semantic network applications (Christoforaki, Constantopoulos, and Doerr 1995; Dionissiadou and Doerr 1994) and successfully applied to create the first version of the CIDOC CRM from the CIDOC relational model. Since then, this approach has loosely been followed by the CRM group, but it has not yet been verified by independent groups. The driving force behind the process and ontology are the properties rather than the classes, which is contrary to the well-established BOOCH, RATIONAL ROSE (Quatrani 1998), and other object-oriented design methodologies, but we have found it far more productive for our purposes.

The CIDOC CRM has been extended several times. Some of these extensions are explicitly documented.[12] 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.

About the CIDOC CRM Contents

The CIDOC CRM contains classes and logical groups of properties.[13] These groups have to do with notions of participation, parthood and structure, location, assessment and identification, purpose, motivation, use, and so on. These properties have put temporal entities and, with it, events in a central place, as sym-

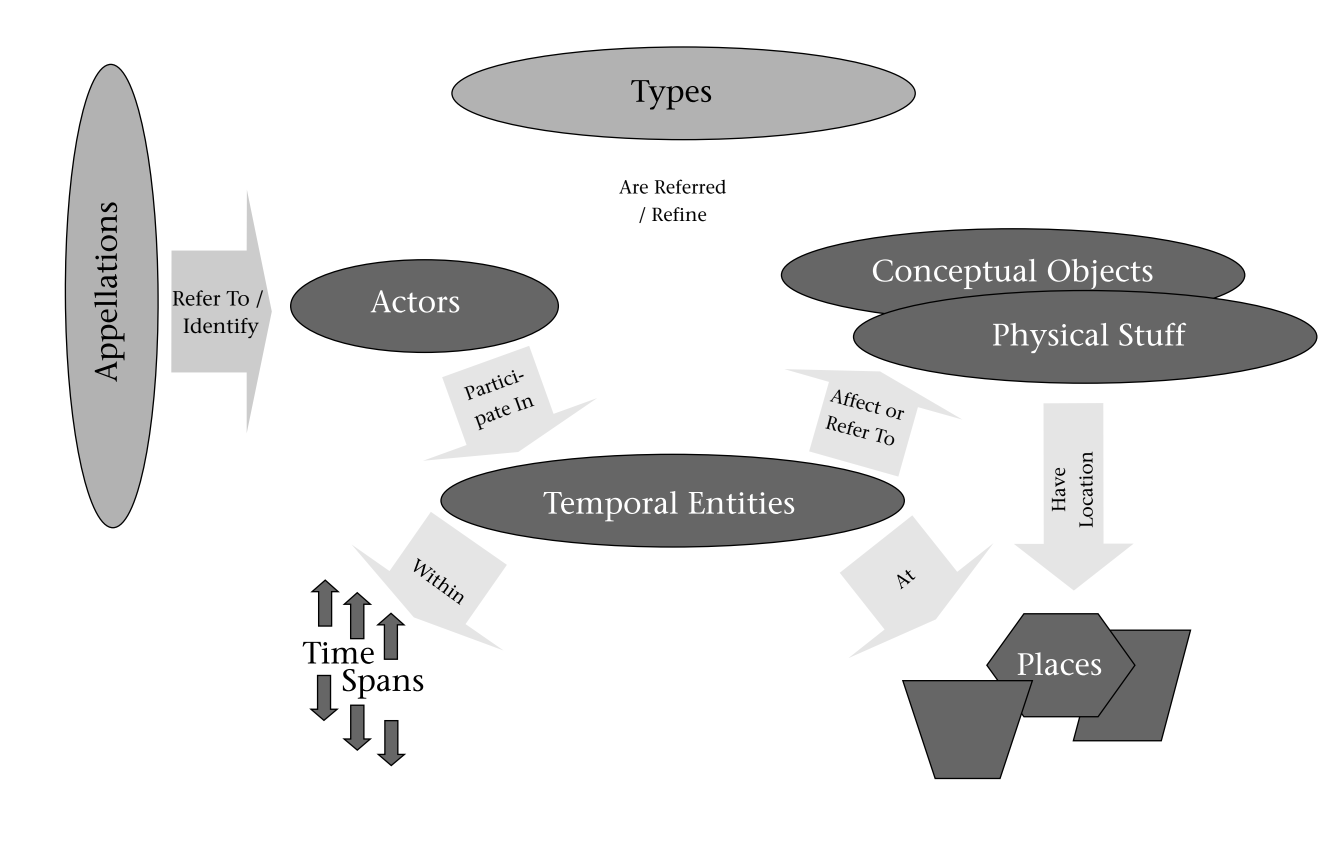

Figure 5. A Qualitative Metaschema of the CIDOC CRM.

bolically shown in figure 5. The base classes in figure 5 turned out to be similar to Ranganathan's (1965) "fundamental categories."

All property paths to dates go through temporal entities. Property paths to places that bypass temporal entities are understood as shortcuts of temporal entities. Similarly, Actors are thought to relate to material and immaterial things (Physical Stuff, Conceptual Objects) only by way of temporal entities. Any instance of a class can be identified by Appellations, the names, labels, titles or whatever used in the historical context. You model the relation to names and its ambiguity as part of the historical knowledge-acquisition process. These names 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 earlier. Frequently, Types serve as the range of properties that refer in general to things of a certain kind, such as "a dress made for a wedding" in contrast to the "dress made for my wedding." I present here some prominent logical groups of CRM properties.

Participation and Spatiotemporal Reasoning

As pointed out in Christoforaki, Constantopoulos, and Doerr (1995), Doerr and Crofts (1999), and Lagoze and Hunter (2001) and motivated by examples in this article, the explicit modeling of events leads to models of cultural contents that can better be integrated. The participation or presence of several nontemporal entities in an event *e*1 allows for an important conclusion: The entities 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 were not somewhere else at the time (with electronic communication, the space volume in which events occur can become very large, for example, Earth to Moon). Culturally, the participants might have influenced each other or, in the case of people, exchanged information. The events $e0_i$ of creation of each participant i happened before or at the time of *e1*. The events $e2_i$ of destruction (or vanishing) of each participant 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

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 Manmade Stuff
P108	- - has produced (was produced by)	E12 Production	E24 Physical Manmade 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.

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 Manmade 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.

knowledge is more reliable than sequencing based on explicit date information. Therefore, we carefully try to preserve such knowledge if it is primary (that is, 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 (for example, a desk where a treaty was signed) to use of tools and weapons and consumption of raw products being produced. Specialization clarifies the more concrete senses modeled 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

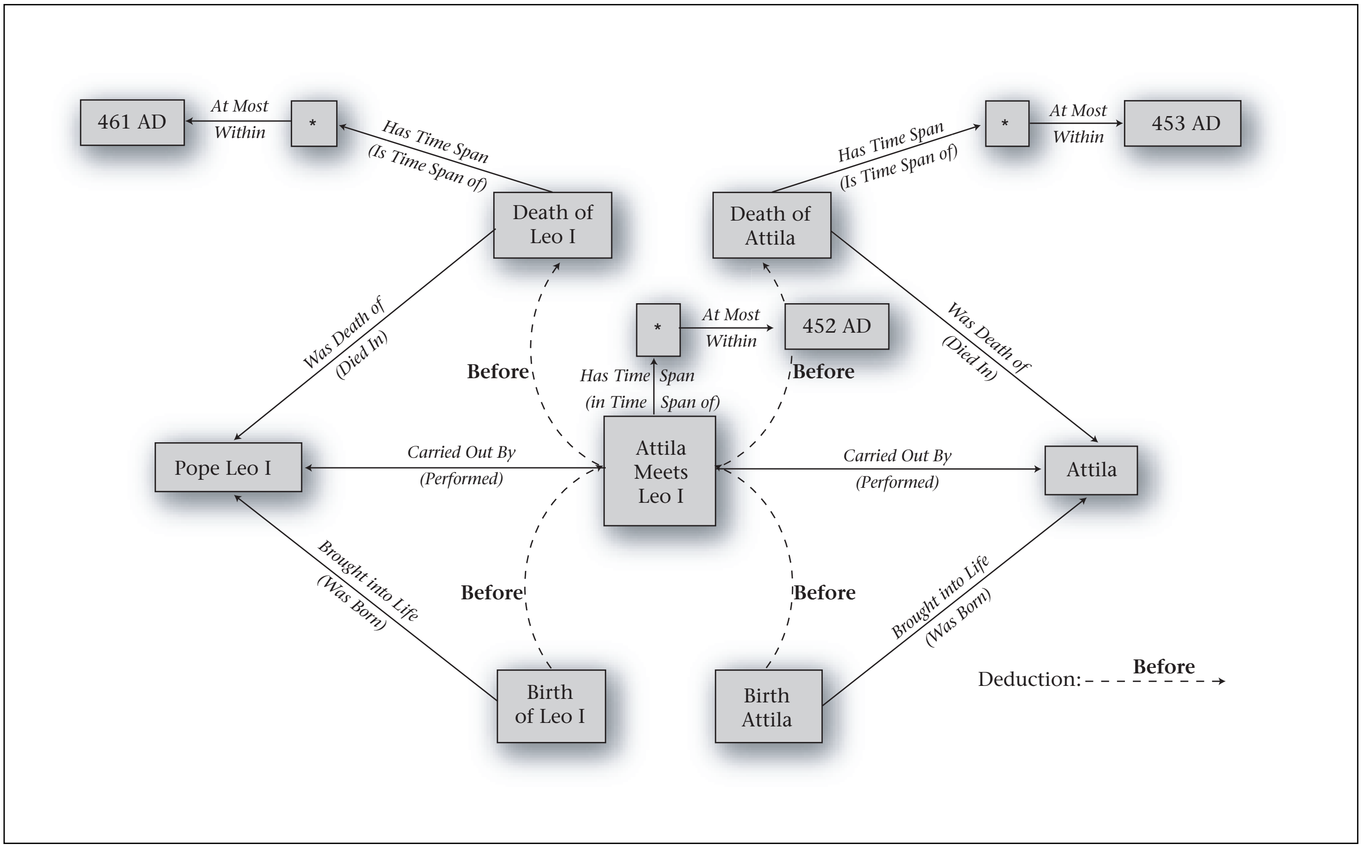

Figure 6. Pope Leo I Meeting Attila the Hun.

can be used in rules independent from further extension of the model.

The next notion relevant in this context is the properties brought into existence and taken out of existence, limiting the existence of things that have a persistent existence, that is, that 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 these events can be useful for temporal reasoning without explicit time: Using participation of other things in the same event, one can derive further termini. Because I perceive events as continuous processes with nonzero extent and infinitely divisible, I argue that each item participates partially in its creation. Therefore, the respective specializations such *has created* appear in both hierarchies.

The properties in tables 1 and 2 characterize the semantics of data structures in the cultural area. Figure 6 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 might be wrong, the four deductions are true if the meeting has happened. Each death date constrains the meeting and both birth dates, the meeting date constrains both death and birth dates, and so on. A maximum lifespan 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.

Note that any extension of the model with another property that implies participation, for example, *was injured in,* would not be captured by the previous reasoning in some implementations, unless it is an explicitly declared subproperty of P11. Because subproperties are not supported by the older Object Management Group (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 further exploration.

Properties of Locating

The question of where it is can be answered in natural language by relation with two different kinds of entities: (1) geometric areas and (2) ob-

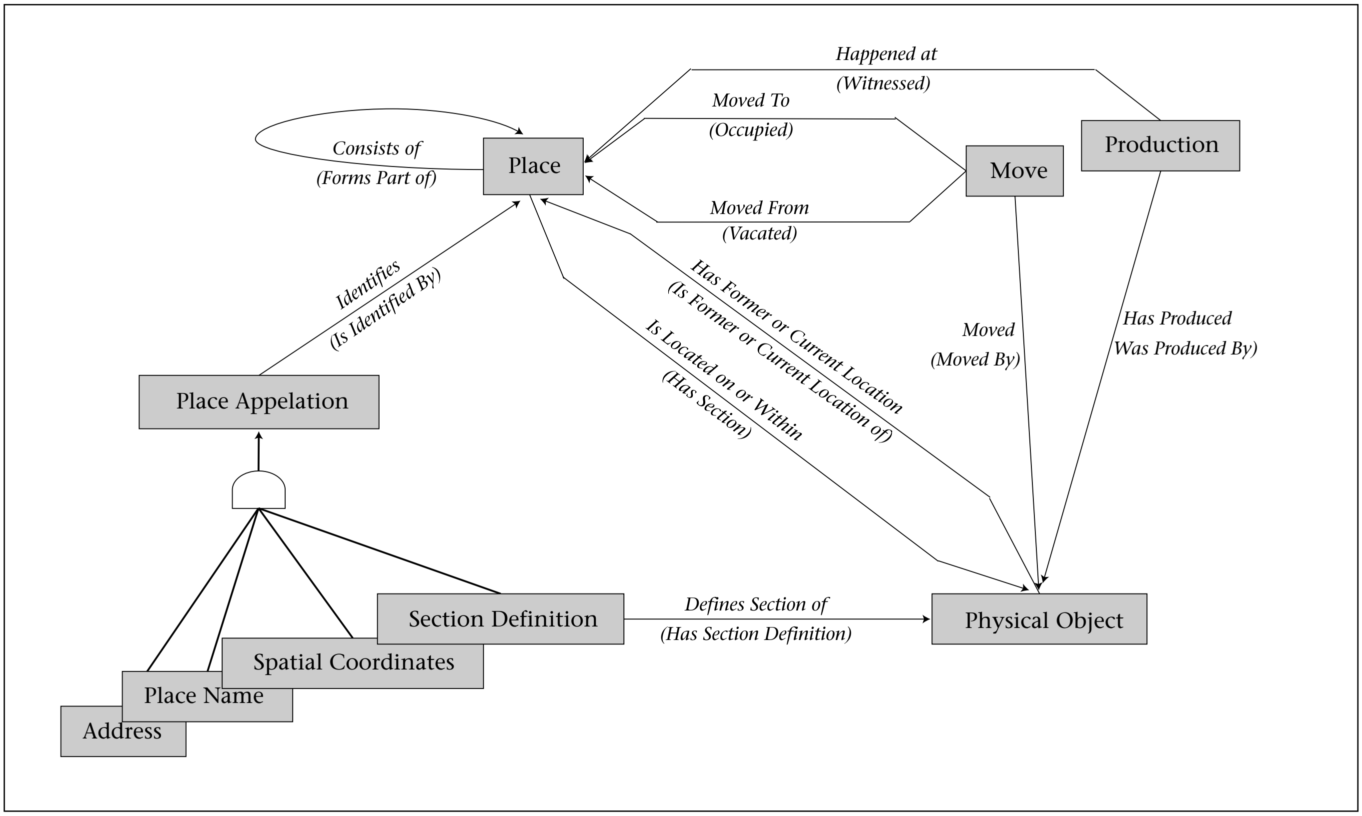

Figure 7. Properties of Locating Items.

jects. Examples of areas are *in France*, *in Athens,* and *39N 124E*. Points given by spatial coordinates are typically understood as the center of a wider, extended area. Objects can be in the proper sense ("bona fide objects" [Smith and Varzi 1997]), such as *on Queen Elizabeth (the ship), in my suitcase, at home*, or they can be in landscape and other features ("fiat objects" [Smith and Varzi 1997]), such 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 the earth. These objects, in turn, can 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. Because 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 the result of a move or a construction in situ, as with buildings. This view is the result of a longer discussion. The notion *place* is ambiguous in English and gives rise to endless confusion in database design. In particular, I take the position that there is no image of a place because it is not a material entity.

Places are identified by proper names or names referring to topological characteristics of object types, so-called *segments* (Gerstl and Pribbenow 1996), or E46 Section Definition in the CRM, such as bow, head, neck, and bottom.

Addresses in general need not be places; their function is often that of a contact point for some person or organization (P76 has contact points [provides access to]), be they physical letterboxes or post office boxes. Figure 7 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 or maximal area elements.

Notions of Influence

The knowledge of what influenced or motivated a human activity and, in turn the persistent things that have come upon us, is culturally most relevant. The working group has not yet developed a systematic understanding of the different forms of influence and their mutual relations. Some are more physical, such as using a mould or a tool. The influence of a mould on a produced object is strong and can often be

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
P62	depicts object (is depicted by) (mode of depiction : Type)	E24 Physical Manmade Stuff	E18 Physical Stuff
P63	depicts event (is depicted by) (mode of depiction : Type)	E24 Physical Manmade Stuff	E5 Event
P65	shows visual item (is shown by)	E24 Physical Manmade 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	E19 Physical Object
P20	had specific purpose (was purpose of)	E7 Activity	E7 Activity

Table 3. Properties of Influence.

verified on the object afterward. 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 has a weak influence; it is more an intellectual influence than a physical one. Further, activities are influenced by other activities, such as orders, or just by the emotions they raise. If a real influence exists, a temporal sequence can be deduced. In contrast to "hard facts," as described in Participation and Spatiotemporal Reasoning, the notions described here vary over a continuum of stronger and weaker influence, which can be verified more or less easily afterward. To date, the CRM contains the properties of influence that are given in table 3.

The properties P15, P33, and P16 describe plans, prototypes, and physical tools (moulds, hammers, and so on) 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, and so on.

The properties P62, P63, P65, P67, and P70 describe an influence that 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, and so on, play an extraordinary role in historical studies. All range values of these properties must have existed before the respective process that manifested them in the product.

The properties P17, P18, and P20 describe an influence that originates in the activity itself, such as orders, impressions, or emotions. P20, in particular, captures sequences of planned activities. For example, in the sentence, "George of Kyriaze orders a commemoration cross for donation to the Metropolitan Church of Ankara,"[14] there is a specific 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 article, the properties of influence have been revised and a final form was decided.[15,16]

Applications

It would go beyond this article to describe applications in detail. Several installations based on the CIDOC CRM have already been made (Crofts 1999). A recent test together with Consortium for Computer Interchange of Museum Information (CIMI) aimed at demonstrating that the semantics of heterogeneous museum records are preserved under the CIDOC CRM.[17] Two examples are interesting: The Clayton collection of the Natural History Museum in London and the Australian Museums Online (AMOL) initiative both use flat records in the ACCESS database. The Clayton collection describes a complex relation between plant specimens, initial and current classification events, and classification documents. These records can automatically be 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: (1) data structures (such as the Clayton data) need not implement the complexity of an ontology for information integration to be interpretable, and (2) an ontology can help to create interpretable data structures. More such tests will be carried out in the near future.

Conclusion

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

The CIDOC CRM has achieved a relatively high degree of maturity and completeness in capturing the conceptualizations 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 can 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 the experience in training nonexperts in basic knowledge representation principles has been encouraging. The complexity of the domain is intriguing. Philosophical considerations and long discussions were necessary to clarify the role of the modeled 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, because it is not needed for either work in isolation. It was interesting to learn that not all domain concepts are equally suited as a basis for information integration.

The methodology presented here appears to be contrary to well-known object-oriented methodologies for designing the controlling software of information systems. I want to point out that there can 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 article tried to illustrate ontology from a point of view seldom taken: the relationships between entities as a driving force for the logical structure rather than the nature of involved individuals. This approach seems to be appropriate to analyze data structures in contrast to terminological systems. A coherent analysis of (nonunary) 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 my 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 conceptualization 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. I 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 seem to be interesting topics for further investigation.

The CIDOC CRM is envisaged to become an ISO standard in 2004. In parallel to the standardization work, I 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 helpful (as discussed in the subsection entitled Participation and Spatiotemporal Reasoning, subproperties play a crucial role for that). From the point of contents, the CIDOC CRM still touches only fundamental concepts, and many extensions will be useful to allow for more reasoning, such as temporality of properties, phases of objects, a coherent model of influence, and modeling performing arts. I see also a need to clarify philosophical questions of foundational character about the nature of the knowledge we describe.

Acknowledgment

I want to express my particular gratitude to Christopher Welty for the time he invested in editing this article, substantially contributing to its current concise form and readability.

Notes

1. The CIDOC CRM Home Page. 2001. cidoc.ics.forth.gr/what_is_crm.html.

2. Foundations of Data Warehouse Quality (DWQ), European ESPRIT IV Long Term Research (LTR) Project 22469, 1996–1999. www.dbnet.ece.ntua.gr/~dwq/.

3. The Art Museum Image Consortium (AMICO). amico.org. AMICO data dictionary version 1.3, amico.org/AMICOlibrary/dataDictionary.html.

4. www.fordham.edu/halsall/mod/1945YALTA.html.

5. www.getty.edu/research/tools/vocabulary/tgn/.

6. CIDOC. 1995. International Guidelines for Museum Object Information: The CIDOC Information Categories. CIDOC.

7. N. Crofts, C. Dallas, I. Dionissiadou, M. Doerr. 1997. Notes on the Data Modeling Meeting in Crete, July 1997. cidoc.ics.forth.gr/docs/notes_data_-modelling_1997_crete.doc.

8. N. Crofts, C. Dallas, I. Dionissiadou, M. Doerr. 1997. Notes on the Data Modeling Meeting in Crete, July 1997. cidoc.ics.forth.gr/docs/notes_data_ modeling_1997_crete.doc.

9. The Semantic Index System—SIS." ICS FORTH, Information Systems Laboratory. 2003. Heraklion, Crete, Greece. www.ics.forth.gr/isl/r-d-activities/sis.html.

10. Crofts, N., et al. 2001. CRM Scope Definition: Proposal of the Steering Committee of the CIDOC CRM SIG, July 7, 2001.

11. Crofts, N., et al. 2001. CRM Scope Definition: Proposal of the Steering Committee of the CIDOC CRM SIG, July 7, 2001.

12. Doerr, M., ed. 2000. Agios Pavlos Extensions—Add-ons for the Completion of the CIDOC CRM. cidoc.ics.forth.gr/docs/agios_pavlos_extensions.rtf.

13. N. Crofts, I. Dionissiadou, M. Doerr, P. Reed, eds. 1998. CIDOC Conceptual Reference Model—Information Groups. ICOM/CIDOC Documentation Standards Group. cidoc.ics.forth.gr/docs/info_groups.rtf.

14. Doerr, M., and Dionissiadou, I. 1998. Data Example of the CIDOC Reference Model—Epitaphios GE34604. cidoc.ics.forth.gr/docs/crm_example_1.doc

15. The CIDOC CRM Home Page. 2001. cidoc.ics.forth.gr/what_is_crm.html.

16. N. Crofts, M. Doerr, T. Gill, S. Stead, M. 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.

17. J. Perkins. 2000. ABC/Harmony CIMI Collaboration Project. September 30. www.cimi.org/public_docs/Harmony_long_desc.html.

References

Analyti, A.; Spyratos, N.; and Constantopoulos, P. 1998. On the Semantics of a Semantic Network. *Fundamenta Informaticae* 36(2–3): 109–144.

Baker, T. 2000. A Grammar of Dublin Core. *D-Lib Magazine* 6(10): 3.

Bekiari, C.; Constantopoulos, P.; and Bitzou, T. 1992. DELTOS: A Documentation System for the Antiquities and Preserved Buildings of Crete, Requirements Analysis. Technical Report FORTH-ICS/TR-60, Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece.

Bekiari, C.; Gritzapi, C.; and Kalomoirakis, D. 1998. POLEMON: A Federated Database Management System for the Documentation, Management, and Promotion of Cultural Heritage. Paper presented at the Twenty-Sixth Conference on Computer Applications in Archaeology, 24–28 March, Barcelona.

Bergamaschi, S.; Castano, S.; De Capitani De Vimercati, S.; Montanari, S.; and Vincini, M. 1998. An Intelligent Approach to Information Integration. Paper presented at the International Conference on Formal Ontology in Information Systems (FOIS98), 6–8 June, Trento, Italy.

Bower, J. M., and Baca, M. 1994. *Union List of Artist Names*. Getty Art History Information Program. New York: G. K. Hall.

Calvanese, D.; De Giacomo, F.; Lenzerini, M.; Nardi, D.; and Rosati, R.; 1998. Description Logic Framework for Information Integration. Paper presented at the Sixth International Conference on the Principles of Knowledge Representation and Reasoning (KR'98), 2–5 June, Trento, Italy.

Christoforaki, M.; Constantopoulos, P.; and Doerr, M. 1995. Modeling Occurrences in Cultural Documentation. Paper presented at the Third Convegno Internazionale di Archeologia e Informatica, 22–25 November, Rome, Italy.

Constantopoulos, P. 1994. Cultural Documentation: The CLIO System. Technical Report, FORTH-ICS/TR-115, Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece.

Crofts, N. 1999. Implementing the CIDOC CRM with a Relational Database. *MCN Spectra* 24(1): 1–6.

Crofts, N.; Dionissiadou, I.; Doerr, M.; and Stiff, M. 2001. Definition of the CIDOC Object-Oriented Conceptual Reference Model, Version 3.2.1, Working Document ISO/TC46/SC4/WG9/N2, International Organization for Standardization.

Dionissiadou, I., and Doerr, M. 1994. 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, 31–38. Silver Spring, Md.: Museum Computer Network.

Doerr, M. 2001a. CIDOC Conceptual Reference Model, Correlation Test Project, Results, Foundation for Research and Technology—Hellas, Crete, Greece.

Doerr, M. 2001b. Mapping of the AMICO Data Dictionary to the CIDOC CRM, Technical Report FORTH-ICS/TR-288, Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece.

Doerr, M. 2000. Mapping of the DUBLIN CORE Metadata Element Set to the CIDOC CRM, Technical Report FORTH-ICS/TR-274, Foundation for Research and Technology—Hellas, Heraklion, Crete, Greece.

Doerr, M., and Crofts, N. 1999. Electronic Esperanto: The Role of the Object-Oriented CIDOC Reference Model. Paper presented at the ICHIM'99, 22–26 September, Washington, D.C.

Gerstl, P., and Pribbenow, S. 1996. A Conceptual Theory of Part—Whole Relations and Its Applications." In *Data and Knowledge Engineering 20,* 305–322. Amsterdam, The Netherlands: North Holland-Elsevier.

Guarino N. 1998. Formal Ontology in Information Systems. In *Formal Ontology in Information Systems,* ed. N. Guarino, 3–15. Amsterdam, The Netherlands: IOS.

Guarino, N., and Giaretta, P. 1995. Ontologies and Knowledge Bases, Toward a Terminological Clarification. In *Toward Very Large Knowledge Bases,* ed. N. J. I. Mars, 25–32. Amsterdam, The Netherlands: IOS.

Gruber, T. R. 1993. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. Technical Report, KSL93-04, Knowledge Systems Laboratory, Stanford University.

Karvounarakis, G.; Christophides, V.; Plexousakis, D.; and Alexaki, S. 2001. Querying Querying RDF Descriptions for Community Web Portals. Paper presented at the Seventeenth French National Conferences on Databases BDA, 29 October–2 November 2001, Agadir, Morocco.

Lagoze, C., and Hunter, J. 2001. The ABC Ontology and Model. DC-2001, International Conference on Dublin Core and Metadata, 22–26 October, Tokyo.

MDA. 1998. SPECTRUM: The UK Museum Documentation Standard. 2d ed. Cambridge, United Kingdom: Museum Documentation Association.

Mylopoulos, J.; Borgida, A.; Jarke, M.; Koubarakis, M. 1990. TELOS: Representing Knowledge about Information Systems. *ACM Transactions on Information Systems* 8(4): 325–362.

Quatrani, T . 1998. *Visual Modeling with Rational Rose and UML*. Reading, Mass.: Addison-Wesley.

Ranganathan, S. R. 1965. A Descriptive Account of Colon Classification. Bangalore, India: Sarada Ranganathan Endowment for Library Science.

Smith, B., and Varzi, A. 1997. Fiat and Bona Fide Boundaries: Toward on Ontology of Spatially Extended Objects. In *Proceedings of the International Conference COSIT'97, October 15–18,* 103-119. Lecture Notes in Computer Science 1329. New York: Springer.

Theodoridou, M., and Doerr, M. 2001. Mapping of the Encoded Archival Description DTD Element Set to the CIDOC CRM, Technical Report FORTH-ICS/TR-289. FORTH, Crete, Greece.

Wiederhold, G. 1992. Mediators in the Architecture of Future Information Systems. *IEEE Computer Magazine* 25(3): 38–49.

Martin Doerr holds a Ph.D. in physics from the University of Karlsruhe, Germany, and is a senior researcher at the Foundation for Research and Technology—Hellas (FORTH). Since 1992, he has participated in a series of projects on cultural information systems and taught courses in cultural informatics. His e-mail address is martin@ ics.forth.gr.