Ontologies for Cultural Heritage

Martin Doerr

ICS-FORTH, Greece, martin@ics.forth.gr

1 Introduction

In the cultural heritage domain information systems are increasingly deployed, digital representations of physical objects are produced in immense numbers and there is a strong political pressure on memory institutions to make their holdings accessible to the public in digital form. The sector splits into a set of disciplines with highly specialized fields. Due to the resulting diversity, one can hardly speak about a "domain" in the sense of "domain ontologies" [33]. On the other side, study and research of the past is highly interdisciplinary. Characteristically, archaeology employs a series of "auxiliary" disciplines, such as *archaeometry, archaeomedicine, archaeobotany, archaeometallurgy, archaeoastronomy*, etc., but also historical sources and social theories.

Interoperability between various highly specialized systems, integrated information access and information integration increasingly becomes a demand to support research, professional heritage administration, preservation, public curiosity and education. Therefore the sector is characterized by a complex schema integration problem of associating complementary information from various dedicated systems, which can be efficiently addressed by formal ontologies [14, 32, 33].

There is a proliferation of specialized terminology, but terminology is less used as a means of agreement between experts than as an intellectual tool for hypothesis building based on discriminating phenomena. Automated classification is a long established discipline of archaeology, but few terminological systems are widely accepted. The sector is, however, *more focused* on establishing knowledge about facts and context in the past than about classes of things and the laws of their behavior. Respectively, the concatenation of related facts by *co-reference* [56] to particulars, such as things, people, places, periods is a major open issue. *Knowledge Organisation Systems* (KOS, [62]) describing people and places are employed to a certain degree, and pose similar technical problems as ontologies, but the required scale is very large. In this chapter, we describe how ontologies are and could be employed to improve information management in the cultural heritage sector.

S. Staab and R. Studer (eds.), *Handbook on Ontologies,* International Handbooks on Information Systems, DOI 10.1007/978-3-540-92673-3,

2 The Cultural Heritage Domain

Layman may think of cultural heritage primarily as fine arts collections and regard the description and indexing of these objects as relatively straightforward and reasoning more as a matter of scholarly reflection about their ideal values than a matter of logic. In reality, cultural heritage is more than as a domain. It comprises a broad spectrum of functions about the study and preservation of physical evidence of the past of all sorts of human activities [19].

2.1 What is Cultural Heritage?

In a narrower sense, we may regard the cultural heritage as the things preserved by the *memory institutions*, i.e. museums, sites and monuments records ("SMR"), archives and libraries. Their international umbrella organizations are: the International Council of Museums (ICOM,[1]) the International Federation of Library Associations (IFLA,[2]) and the International Council of Archives (ICA.[3]) They maintain their specific documentation policies and standards.

Following ICOM, "A museum is a non-profit making, permanent institution in the service of society and of its development, and open to the public, which acquires, conserves, researches, communicates and exhibits, for purposes of study, education and enjoyment, material evidence of people and their environment" [60] and "Museums" hold primary evidence for establishing and furthering knowledge" [61]. SMRs are typically departments of a Ministry of Culture, pursuing similar goals as museums, but for immobile sites. Archives keep very large amounts of original material – mostly written and images – in their historical order, such as administrational records, letters from VIPs, photographic collections and others.

To a certain degree, libraries may also preserve cultural heritage when they keep unique books, however their focus is on mediating access to non-unique information sources. In contrast, most cultural heritage objects are a rather mute evidence of past events that acquire relevance from understanding the context of their origin and history. The object may appear less as an information source in its own right than as an "illustration" of the past. This distinction is important to understand the difference between library and cultural heritage information, and the immense complexity of the latter.

One can appreciate the diversity of cultural heritage from the following list of major kinds of collections:

- History of arts and modern arts (graphics, painting, photography, sculpture, architecture, manuscripts, religious objects),

[1] http://www.icom.org
[2] http://www.ifla.org
[3] http://www.ica.org

- Historical heirloom (treaties, letters, manuscripts, drawings, photos, films, personal objects, weapons),
- Archaeology (sherds, sculptures, tools, weapons, household items, human remains),
- Design (furniture, tableware, cars, etc.),
- Science and technology (machinery, tools, weapons, vehicles, famous experiments, discoveries),
- Ethnology (costumes, tools, weapons, household items, religious objects, etc.)
- Immobile sites (architecture, sculpture, rock art, caves)
- To a certain degree, natural history collections, such as paleontology, biodiversity, mineralogy are also evidence of human activities (i.e. research) and hence culture.

Handling information about all those kinds of things implies the use of very rich terminology, multilingual and often specific to particular communities or even to particular scholars. Agreement on common terminology is difficult and equivalent terms in other languages are often missing. It is an obvious challenge for employment of formal ontologies that poses not only technical problems, but also intellectual challenges in the approximation of intuitive or traditional concepts by logical definitions, such as the possible narrower and wider meanings of the same term, objective declaration of discriminating features or fuzzy transitions of instances from one class to another.

2.2 Functions of Cultural Heritage Information

One can distinguish kinds of cultural heritage information systems by their major functions. Those are:

- Collection management (acquisition, registration, "deaccession", inventory, loans, exhibitions, insurance, rights, protection zones) [29, 30]
- Conservation (diagnosis of deterioration, preventive measures, interventions, treatments and chemical agents)[78]
- Research (investigation, description, interpretation)
- Presentation (portals, teaching, publication)

In each of these four areas quite distinct and highly specialized information systems exist, created and maintained by different players. On the other side, information in all those systems overlaps and should be mutually accessible in order to do the job. One of the major challenges of cultural heritage information management is the interoperability of those system and integration of information across function and discipline.

Collection management systems are offered by several commercial vendors. They are mostly built on Relational or hierarchical database systems. Many customized systems are built on demand by IT experts. They support the technical management and administration of collections or sites and monuments. A comprehensive, internationally accepted definition of their functions can be found in [29]. *Curators* provide basic descriptions of the objects that serve their identification and handling, but do also research and justify their relevance, i.e. why the object is kept in a museum. Archival collections typically consist of millions of leaves. It is unusual to describe each item. Rather, curators document the historical context under which coherent sets of documents were created or brought together as finding aids for researchers, so-called collection-level metadata. Only recently, more fine-grained documentation is occasionally considered [23].

Conservation information may be part of the collection management or separate from it. It deals with the scientific, material analysis of the objects, preventive measures and interventions. Loan management and historical research may need those data. Art and monument conservation is an underestimated sector of financial importance. *Art conservators* are scientists who need, similar to doctors, to accumulate and exchange immense knowledge about diagnosis methods, treatments and side effects [78]. There are a few dedicated websites and systems for information exchange between experts [3,58] and learning [20], but there is still a wide market for such systems. Since they deal with *categorical* (general) knowledge, such systems should better be ontology-based.

Research information systems are highly specialized and mostly built on demand for specific projects. There are reference systems that list consolidated, uniform descriptions of all known items of a certain kind, such as Roman Inscriptions [15] or the Union List of Artist Names [12]. There are many systems[4] that integrate information from thousands of sources for statistical or other kinds of analysis. Particularly important became GIS-based reasoning systems, such as for archaeological site prediction, and systems for running automatic classification (see for instance, [24,38]). Unfortunately, idiosyncratic design and insufficient management of source references frequently make the reuse of the integrated information impossible after the project ends.[5] More effective means of data transformation and migration are still to be developed. Ontologies could play an important role in that.

Presentation systems give access to cultural heritage information to the general public or a community of subscribers, in particular teachers and academics (see Chapter "Ontology-Based Recommender Systems"). We estimate that more than 95% of museum objects are not in any exhibition, and archives are mostly closed to the public. Therefore there is a strong political pressure to make at least object descriptions from the collection management systems publicly accessible. Museum portals (see Chapter "Ontology-Based Recommender Systems") may present parts of collections. The scale is immense:

[4] For instance, those published by the conference series Computer Applications & Quantitative Methods in Archaeology `http://caa.leidenuniv.nl/proceedings/`

[5] Round Table discussion at the 8th EAA ANNUAL MEETING, 24–29 September 2002, Thessaloniki–Hellas, `http://www.symvoli.com.gr/eaa8/mple.htm#P5`

larger museums hold millions of objects. The Smithsonian Institutions hold over 100 million objects. Other presentation systems may take the form of an electronic exhibition, complementary information to a physical exhibition, or the form of an electronic publication that elaborates a particular subject matter. Ontologies play a major role to provide structured access points and to structure the subject matter itself in these systems.

Recent efforts deal with the capturing and preservation of *performing arts* and *oral tradition* [11, 34, 49]. Since there is no object to be described, traditional models of documentation are not appropriate, and new models are discussed.

3 The Schema Integration Problem

Most of the professional systems referred to above are based on fairly complex database schemata. For instance, CIDOC proposed until 1995 a standard Relational Schema for museums with more than 400 tables. As described above, cultural heritage information is distributed in many different systems which complement each other. One source may relate Roman names to Roman inscriptions, another Roman inscriptions to stones, another stones to place of finding, and another places to coordinates [25]. But still most efforts to integrate heritage information concentrate on finding minimal common description elements for objects as *finding aids* rather than documentation. This is motivated by practice from the library communities, in particular the Dublin Core Consortium.[6]

3.1 Metadata and Application Profiles

Since libraries and Digital Libraries hold objects that contain data, they use to call the descriptions of their objects “metadata”, i.e. data about data. This term has also been adopted by museums for their object descriptions, even though their objects are not data. There is a plethora of attempts to structure metadata as flat lists of properties, which may be aggregated in so-called “application profiles” [9, 37], and the mapping and data transformation between different metadata formats may be called a “metadata crosswalk” [62]. The labels of metadata properties, such as “creator”, “date”, etc., remind concepts. Therefore several authors recently regard metadata schemata as “vocabularies” or a kind of ontologies. We regard this as a confusion of information models with ontologies, as elaborated by [71, 72]. It is to remind that formal ontologies were introduced to computer science to describe common conceptualization behind multiple schemata [32, 33], and not to become a synonym of schemata. Further, the reduction of complex object histories to a flat set of

[6] `http://dublincore.org/`

properties can only be achieved by semantic overloading of these properties, which conflicts with the definition of an ontology, as shown in [18, 43].

Nevertheless, numerous digitization projects of cultural objects create digital libraries with Dublin Core metadata elements as minimal standard. Also wide-spread is the use of MPEG7 ([8, 40] and Chapter "Ontologies for Cultural Heritage"), the metadata standard for multimedia objects, for obvious reasons, which is a far richer representation of the structure, history and subject of the object. There is serious research on automatic matching of metadata elements in order to support schema mapping and merging which is based on comparison of metadata elements with ontologies. The idea is to detect similarities between schema elements and the underlying concepts by the similarity of naming and properties. The underlying concepts are found in a formal ontology, such as WordNet [26].

3.2 ISO21127

Information integration based on finding aids for the objects actually fails to integrate the information about the wider historical contexts these objects illustrate and from which they get their relevance. If a serious integration of the relevant contents of cultural heritage information is intended, richer models must be employed. For instance, the Research Libraries Group in California successfully integrated in their Cultural Materials Initiative data from about a thousand cultural institutions encoded in about a hundred different schemata into a far richer schema, virtually without loss of information. This schema was derived from the CIDOC CRM ontology, now ISO21127, which is currently the most elaborated ontology for the integration of cultural heritage information.

The CIDOC CRM is a formal ontology [16] intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information. It was developed by interdisciplinary teams of experts, coming from fields such as computer science, archaeology, museum documentation, history of arts, natural history, library science, physics and philosophy, under the aegis of the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Its development started *bottom up*, by reengineering and integrating the semantic contents of more and more database schemata and documentation structures from all kinds of museum disciplines, archives and recently libraries.

The development team applied strict principles to admit only concepts that serve the functionality of global information integration, and other, more philosophical restrictions about the kind of discourse to be supported (for more details see [19]). The application of these principles was successful in two ways. On the one side, the model became very compact without compromising *adequacy* [71]. The very first schema analyzed in 1996, the CIDOC Relational Data Model with more than 400 tables (described by [66]), could be reduced to a model of about 50 classes and 60 properties, with far wider applicability than the original schema. On the other side, the more schemata were analyzed, the fewer changes were needed in the model (see version history.[7]) The present model contains 80 classes and 132 properties, representing the semantics of may be hundreds of schemata. As a result of the successful reformulation of the original Relational model, CIDOC started the standardization process in collaboration with ISO in 2000. The model was accepted as ISO21127:2006 in September 2006.

Deliberately, the CIDOC CRM ontology is presented in a textual form to demonstrate independence from particular knowledge representation formats. There exists however a formal definition in TELOS [59]. The CRM distinguishes individual classes from properties (binary relationships). Properties are directed and bidirectional, with distinct labels for each direction. It employs strict multiple inheritance (without exceptions) for both classes and properties. It foresees multiple instantiation, i.e. one particular item can accidentally be instance of more than one class. Domain and range of properties are associated with the quantifiers zero, one or many. There exist valid equivalents in KIF, RDFS and OWL, to the degree the respective constructs are supported. Four ideas are central to the CRM (see Fig. 1):

1. The possible ambiguity of the relationship between entities and the identifiers ("Appellations") that are used to refer to the entities are a part of the historical reality to be described by the ontology rather than a problem to be resolved in advance. Therefore, the CRM distinguishes the nodes representing a real item from the nodes representing only the names of an item.
2. "Types" and classification systems are not only a means to structure information about reality from an external point of view, but also part of the historical reality in their nature as *human inventions*. As such they

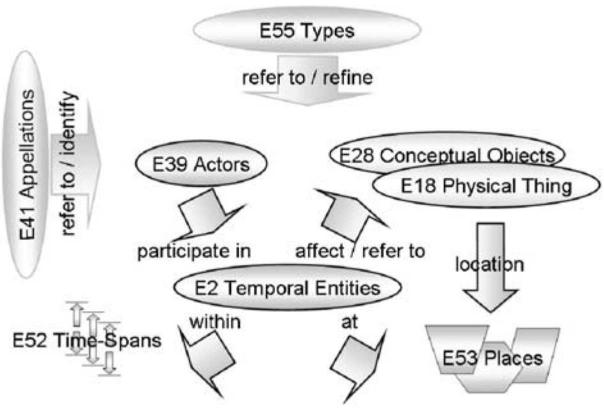

Fig. 1. Top-level entities of the CIDOC CRM

[7] http://cidoc.ics.forth.gr

fall under “Conceptual Objects”, inheriting properties of creation, use, etc. Similarly, all documentation is seen as part of the reality, and may be described together with the documented content itself. This reification problem is not appropriately dealt with in current ontology languages. The CRM is forced to use some workarounds we do not analyze here further. Terminology, i.e. classes that are not contributing as *domain* or *range* to the relationships expressed in data structures, are not part of the core ontology itself but regarded as instances of “Type” for practical reasons.
3. The normal human way to analyze the past is to split up the evolution of matters into discrete events in space and time. Thus the documented past can be formulated as series of events involving “Persistent Items” (also called endurants, see [19]) like Physical Things and Persons. The involvement can be of different nature, but it implies at least the presence of the respective items. The linking of items, places and time through events creates a notion of “world-lines” of things meeting in space and time (see Fig. 2). Events, seen as processes of arbitrary scale, are generalized as “Periods” and further as “Temporal Entities” (also called perdurants [19]). Only the latter two classes are directly connected to space and time in the ontology. The Temporal Entities have fuzzy spatiotemporal boundaries which can be approximated by outer and inner bounds.
4. Immaterial objects (“Conceptual Objects”) are items that can be created but can reside on more than one physical carrier at the same time, including human brains. Immaterial items can be present in events through the respective physical information carriers (see Fig. 3). Immaterial items cannot be destroyed, but they disappear when the last carrier is lost.

As a standard, the use of CRM concepts is not prescriptive, but provides a controlled language to describe common high-level semantics that allow for information integration at the schema level. It is intended to serve

1. As an intellectual guide to good practice of conceptual modeling in the sector.

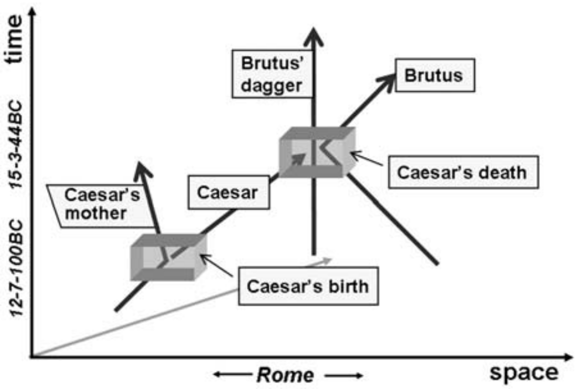

Fig. 2. Historical events as meetings

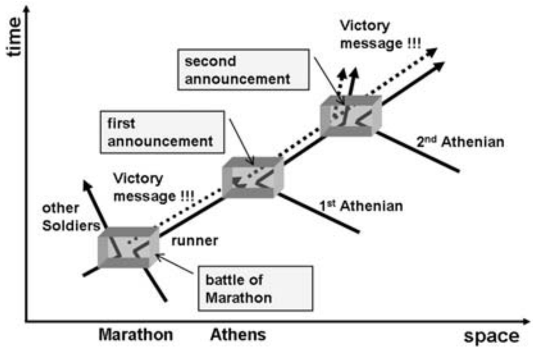

Fig. 3. Information exchange as meetings

2. As global model for information integration in a "Local as View" (LAV, [13]) or data warehouse manner.
3. As an intermediate model for data migration.

The coverage of the CRM for cultural heritage data has been validated by mappings from numerous data structures of the sector to the CRM. Even the common library format MARC ('Machine Readable Cataloguing') can be adequately mapped to it [49]. Such a mapping can be seen as an interpretation of the data structure elements in terms of the ontology. If the ontology is implemented as a schema (such as in RDFS), the mapping can also be seen as a specification for Local as View (LAV) schema integration. The examples of mappings from Dublin Core or EAD to the CRM [18, 43] show how well-defined common semantics can be associated with typical metadata formats. In particular they allow for describing explicitly the cases of semantic overloading (such as the use of DC.date for various events). Even MPEG7 has been aligned with the CRM [40]. The CRM is increasingly used in real integrated information environments for cultural heritage systems. A list of references can be found on [1, 2]. Due to the characteristic focus of the empirical base of the CRM, i.e. data structures used for collection descriptions, it is relatively poor in describing family relations, rights, and intellectual processes. The latter has been recently complemented by the FRBRoo model [7, 48].

3.3 FRBRoo and Performing Arts

The FRBR model ('Functional Requirements for Bibliographic Records') was designed as an entity-relationship model by a study group appointed by the International Federation of Library Associations and Institutions (IFLA) during the period 1991–1997 [68]. It was published in 1998. Its innovation is to cluster publications and other items around the notion of a common conceptual origin – the 'Work', in order to support information retrieval and to initiate a new bibliographic practice. It distinguishes four levels of abstraction from conception to the book in my hands: The Work, Expression, Manifestation, Item. Its focus is domain-independent and can be regarded as the most advanced formulation of library conceptualization [48].

Initial contacts in 2000 between the two communities eventually led to the formation in 2003 of the International Working Group on FRBR/CIDOC CRM Harmonisation. The common goals were to express the IFLA FRBR model with the concepts, ontological methodology and notation conventions provided by the CIDOC CRM, and to merge the two object-oriented models thus obtained. Although both communities have to deal with collections pertaining to cultural heritage, those collections are very different in nature: Most of library holdings are non-unique exemplars of publications, i.e. products of industrial processes. FRBR focuses therefore on the "abstract" characteristics that all copies of a single publication should typically display in order to be recognised as a copy of that publication. The history of individual copies and of the immaterial content is not regarded as particularly relevant in library catalogues and therefore widely ignored by FRBR. Of course, libraries do also hold unique items, such as manuscripts; but there are no internationally agreed standards how to deal with such materials, and FRBR mentions them but does not account for them in a very detailed way.

Museums, on the other hand, are mainly concerned with unique items – the uniqueness of which is counterpoised by a focus on the cultural circumstances under which they were produced and through which they are interrelated. CIDOC CRM highlights therefore the physical description of singular items, the context in which they were produced, and the multiple ways in which they can be related to other singular items, categories of items, or even just ideological systems or cultural trends. Of course, museums may also have to deal with exemplars of industrially produced series of artefacts, but CIDOC CRM covers that notion just with the multi-purpose E55 Type class. Museum objects may be referred to in literature kept in libraries. Museum objects may illustrate subjects described in literature. Literature and objects may be created by the same persons, in common events.

The Working Group has submitted the final draft of FRBRoo, i.e. the object-oriented version of FRBR, harmonized with CIDOC CRM, for public review by IFLA in February 2008. This formal ontology is intended to capture and represent the underlying semantics of bibliographic information and to facilitate the integration, mediation and interchange of bibliographic and museum information.

The major innovation of FRBRoo is a realistic, explicit model of the intellectual creation process (see Fig. 1), which should still be developed further in the future for the benefit of librarians and scholars from the various museum disciplines. FRBRoo makes a fundamental distinction between internal representations of our mind (Work), sets of signs or symbols human can interpret (Expression), and physical information carriers.

The idea is that products of our mind, as long as they stay in one person's mind only, are relatively volatile and not evident. In an event of first externalization, the "Expression Creation", concepts of a Work are made manifest by creating an Expression on a first physical carrier. This may be just another person's memory, as in the case of oral tradition, a paper manuscript or a computer disc. In its current draft version, FRBRoo includes a model of performing arts, connecting the interpretation of theatre plays with the recording and documentation of performances. It distinguishes and relates the three intellectual contributions (works) of the creation of the play, of the interpretation and the recording with the associated symbolic forms and physical carriers. This part of the model has been developed and tested in first examples in collaboration with the European funded project CASPAR on Digital Preservation. Even though there is a rising interest in documenting and preserving non-material culture, there are few other models about performing arts [11]. Jane Hunter has done interesting research on representing indigenous knowledge and its oral traditions [41].

3.4 Other Core Ontologies

Independent from the CRM, the European funded project IndeCs, a consortium of multimedia experts, developed around 1997 a core model to trace the provenance of contributions and associated intellectual property rights in multimedia products and implemented a respective information system. This model was taken up by the ABC ontology. The latter is an outcome of the Harmony Project, which was funded cooperatively by the Distributed Systems Technology Cooperative Research Centre (DSTC) (Australia), the Joint Information Systems Committee (JISC) (UK), and the National Science Foundation Digital Libraries Initiative (NSF DLI2) (US). The original goal and continuing motivation of the ABC work arose from the need to integrate information from multiple genres of multimedia information within digital libraries. The researchers working on the Harmony Project have each been involved in a number of metadata initiatives including Dublin Core and MPEG-7.

Complete details of the ABC ontology are described in [46]. It is far smaller than the CRM, just 13 classes and 14 properties. As the CIDOC CRM, ABC describes temporality in a first-class manner. Modeling change over time is critical to the description of digital content due to its inherent fluidity and the linkages of provenance to integrity or trust [45]. ABC includes both the notions of "events" and "situations", which respectively model transitions (i.e. verbs) and existential properties. The inspiration for these concepts lies in process models such as Petri Nets [63] and temporal extensions for first-order logic such as Situational Calculus [55]. Due to these temporal concepts, ABC is able to definitively model time periods during which certain properties of an object are static. It is also able to model events or transitions marking property modification, for example during the change of a version of a digital object. Finally, ABC builds on the concepts developed in the FRBR model [68]. These concepts – *works, expressions, manifestations,* and *items* – give ABC the ability to link entities that have common intellectual property origins. Work in the library community has proven the utility of these concepts [47].

Similarities between ABC and CRM aims and solutions were so striking, that both teams collaborated between 2001 and 2003 on a harmonization project, in which both ontologies adopted concepts from each other and rearranged properties and IsA hierarchies, until a merged representation was possible [21]. The CRM did not adopt the concept of a *situation*: In the end, the representation of object history in ABC as a chain of states (*situation*) and state transitions (*events*) turned out to be redundant, making knowledge revision complex, and causing problems to integrate interconnected histories of multiple objects and agents. ABC has been mainly used in research. Another interesting core ontology is DOLCE [54]. It is product of careful reengineering of the core concepts of WordNet, a linguistic resource derived from dictionaries, enriched with theory-based foundational relationships such as participation, part-whole, constitution, etc. It is rigorously formulated in logic, making it rather difficult for domain experts to comprehend and use it. In contrast to the CRM, space and time are regarded as dependent properties of things, and not as things existing in a potentially empty space-time – the only, but deep incompatibility between both ontologies. Otherwise, many concepts exist in both ontologies. Some concepts in DOLCE are characteristic for other kinds of discourse than that found in data structures for heritage documentation. Interesting enough, museums are not much interested to analyze iconographic representations by discrete schema elements. With the aim of digital archive interoperability in mind, D'Andrea et al. [17] took the CIDOC CRM common reference model for cultural heritage and enriched it with the DOLCE D&S foundational ontology to better describe and subsequently analyze iconographic representations, from, in this particular work, scenes and reliefs from the meroitic time in Egypt.

3.5 Characteristics of Ontologies for Cultural Heritage

Ontologies that deal with semantics equivalent to those of data structures, as the ones presented above, contain few classes and are rich in relationships [19,51], in contrast to terminological ontologies for classifying individual items. Data structures can be seen as equivalents of propositions about a domain ("possible states of affairs", [33]). Therefore their semantics reveal characteristic parts of the discourse of a domain or sector. So what characterizes the discourse in cultural heritage as reflected in data structures and ontologies?

Cultural heritage can be seen as the material evidence of human activities of social relevance in the past. Therefore

- Information is mesoscopic, i.e. at a human scale, neither astronomic nor microscopic, except for microscopic analysis of traces and materials. Information is discrete. Processes are reported or become evident as discrete events involving discrete things, in contrast to geological or meteorological phenomena.

- Information is event centric. Things, people and ideas connect and relate via events.
- Its description is retrospective, in contrast to information to plan the future, such as for manufacturing.

Information is naturally incomplete at some scale. It can be complemented but not be completed. Its description serves a kind of detective work to reconstruct *possible pasts*. The distinction between evidence and conclusion is vital. Therefore information cannot be normalized and integrated on the basis of the assumed past, such as on absolute dates, geographic coordinates, cause-and-effect, states-and-state-transitions. The documentation of the process of observation is necessary to interpret correctly the observed evidence. Even the fact that some scholar classifies an object with a certain term is documented as a historical, intellectual process (this holds equally for biodiversity). Information is about material facts [35]. Observed individual facts are the basis to induce categorical behavior, such as "all Pharaohs were mummified".

The above characteristics hold equally well for other descriptive sciences, such as geography, biodiversity, paleontology, clinical observation and epidemic studies, but also for the documentation of experiments and observations in natural sciences. Whereas the latter formulate their conclusions about their observations in categorical theories ("F=m*a", or "any non-supported material object in the atmosphere of Earth will fall"), scholars interpreting cultural heritage would generally hesitate to formulate their categorical conclusions or hypotheses in a formal representation (see also [31]). Rather, interpretation is normally presented as text rendering implicitly a wealth of associated belief values. Therefore the presented ontologies are surprisingly domain and discipline independent. It is the retrospective discourse that determines their characteristic form. Ontologies describing the formal structure of iconographic subjects can be seen as an exception to this (see [17] and Sect. 4).

Also surprising is the fact that scholars hesitate to formulate in objective terms causes and causation [50]. Whereas in the domain of jurisdiction characteristic ontologies are being elaborated that detail contributions of individuals in activities, modern scholars prefer a more distant stance of multiple views and possible truths [39]. Noteworthy are the promising attempts of [27,67] to formally structure archaeological argumentation, which could lead to an ontology or better epistemology of cultural heritage argumentation, even though vehement arguments against this approach are not missing [39].

Particularly in ethnology and archaeology (as in biodiversity), some information is documented in a partially categorical form, such as: "The boomerang is a hunting weapon of the Australian Aborigines". I.e. a particular community is associated with characteristic kinds of things and kinds of activities. The described object is seen as example of the category and an illustration of the activity. There is neither currently a formal ontology nor a suitable ontology language which would give a realistic account of the relationship between such partially categorical statements and the individual facts as perceived by the domain expert, and there is no dedicated "metaontology" which could be instantiated with such partially categorical statements.

The CIDOC CRM makes a practical distinction between core classes and classes appearing as terminology motivated by the fact that they appear typically as data in data structures, in order to make fine distinctions between the kinds of the referred items. Even though knowledge representation does not distinguish between the two, it is an empirical fact that the sector uses to organize terminology differently, in vocabularies and thesauri, which may more and more be developed into formal ontologies in the proper sense. Consequently, the RDF schema SKOS [5], a W3C First Public Working Draft, suggests the encoding of terms from vocabularies and thesauri as particulars, and not as RDF classes. We follow this distinction here to structure this chapter. Cultural heritage terminology pertains mostly to classes detailing kinds of material things, which is quite similar and or even overlapping with product classification [69].

Other terminologies of the sector characteristically pertain to:

- Materials, conservation agents
- Information objects
- Processes, deterioration, activities
- Social roles
- Literary and iconographical subjects

In the following section we describe the role of terminology and the most important ontologies in the sector.

4 Terminology in Cultural Heritage

In many collaborations and discussions with museum curators and archaeologists we encountered a negative position towards the use of controlled vocabularies or even formal ontologies. Experts tend not to agree with the terminology used by colleagues [31]. This is in strong contrast to the library sector, which cannot exist without standardized terminology. We attribute this to the fact that in the cultural heritage sector terminology is less used as a means of agreement between experts than as an intellectual tool for hypothesis building based on discriminating phenomena. Consequently, automated classification is a long established discipline of archaeology, but few terminological systems are widely accepted. They are built ad-hoc for specific research questions.

The renowned archaeologist Franco Niccolucci posed the question, if archaeologists are "fuzzy" [38]. He discussed the notion of a neolithic tool, a scraper. From a given set of similar stone tools, several archaeologists did each classify a different subset as *scrapers*. The background of this disagreement is that the concept is used to deduce hypothetic function in the past from observed morphology. It represents already a debatable hypothesis. Many archaeologists develop their own *typologies*. There is a continuous demand for specialized reasoning systems (for instance, [24]).

4.1 Information Access by Terminology

The diversity and number of small ontologies, in the order of a hundred to a thousand terms each, puts interesting challenges to ontology matching and alignment. Only automated tools have a chance to exploit this expert terminology for retrieval and reasoning across local systems.

The task of librarians is not hypothesis building, but providing access to information. Quite naturally, they have a long tradition to agree on common terminology as access points. It is not easy for cultural heritage experts to appreciate the need for shared search terms (see for instance [31]), and there is still enough conviction work to be done. In contrast to [31], we assume that cultural heritage terminology could be separated into an upper, stable level suitable for search, and a lower volatile level supporting hypothesis building. This is motivated from our experience building ISO21127 and various information systems. The largest and stable thesaurus in the sector, the Art & Architecture Thesaurus (AAT, [64]), with more than 30,000 concepts, comes actually from a library background (see below).

A problem with classification of material objects are the different aspects (facets), under which the classification may be done. Dominant aspects are the function of the object, its shape or appearance, elements or principles of construction [22]. These three aspects are partially related. For instance, a typical hammer may have a classical shape and construction, but a motorized hammer may only share function, but not the other aspects. Other aspects are forms specific to historical periods or nations. The effect may confuse thesaurus and ontology editors when building IsA hierarchies, and may mislead users when they apply classification terms.

The so-called facet analysis tries to resolve this problem (e.g. [53, 74]) by systematic separation of the concepts for each facet. "Faceted classification", which goes back to the Indian Librarian Ranganathan (1965) [65], employs the systematic combination of classification terms for each relevant facet. For example, the AAT has removed the term "mills" because it can be constructed from "grinding & factories". The method greatly reduces complexity and depth of term hierarchies, and improves the quality of the ontology. On the other side, faceted classification can be seen as a precursor of employing Description Logic (DL) – simply the roles between the combined terms are implicit. It is assumed that the user has an intuitive understanding of the meaning, and that it is unambiguous. The use of DL in cultural heritage is still in the beginning.

It is standard for museum portals and other cultural information systems that provide information about material objects to offer faceted access by type of object, person, place, date. MuseumFinland [42] employs a faceted Finnish ontology, may be the most advanced system in terms of formal representation of terminological concepts. The UK national project FACET [75] with the UK Science Museum's collections database on Thesaurus-Based Query Expansion employs a combination of novel techniques with a faceted theasurus (the Getty Art and Architecture Thesaurus). Aroyo et al. [28] employ the VRA metadata scheme and encode terminology from the Getty Art and Architecture Thesaurus, Union List of Artist Names and ICONCLASS in SKOS [5].

All terminological systems contain very general terms as root elements of their hierarchies. These may vary considerably and cause unnecessary inconsistency between the ontologies, because the purpose of these ontologies is not to solve the philosophical questions these general terms are associated with. For instance, the AAT subsumes under *visual works* material and immaterial things, such as *paintings* and *electronic images*. In the CIDOC CRM, material and immaterial things are disjoint concepts, because reasoning differs considerably for the two. In integrated information systems depending on rich data structures, this incompatibility can interfere with schema integration. The use of a shared core ontologies to enable interoperability between different domain ontologies has been proposed a decade ago by [33]. The ongoing British STAR project [6] is now investigating cross search over different archaeological datasets and grey literature with the CIDOC CRM core ontology as an integrating framework for the datasets and domain thesauri.

4.2 Major Terminological Systems

The AAT is the most widespread ontology in cultural heritage. It has the form of a thesaurus, compatible with ISO2788. Its topic is art and architecture, but covers a wide range of archaeological and ethnological materials as well as any kinds of object that may be subject of art in some way, such as weapons. It was originally developed by merging culture-relevant subject keywords from several large library systems. It is built for faceted classification. Its major facets are: Activities, Agents, Materials, Objects, Physical Attributes, Styles & Periods, Associated Concepts.

The broader term and narrower term relationship are used in the sense of IsA. Its originally monohierarchical ("tree") generalization structure has been extended to polyhierarchical (directed acyclic graph). The AAT introduces so-called guide terms, (node labels in ISO2788) to group terms under minor facets, such as function or form, but there is no rigorous logic applied to this organization principle. The AAT has been translated into Spanish and Dutch.

English Heritage (EH) maintains also a very large thesaurus of terms for mobile and immobile objects for the United Kingdom, as well as the French MERIMEE thesaurus.

The multilingual thesaurus attached to the European HEREIN project [4] intends to offer a terminological standard for national policies dealing with architectural and archaeological heritage, integrating concepts from the above resources. Beyond just correlating concepts from different languages, the project decided to create for each language a new generalization-specialization hierarchy and to harmonize concepts manually. However, they did not preserve the concepts as found in other sources or link to them. We regard this as problematic, as an opportunity for interoperability seems to have been thrown away unnecessarily.

Remarkable is the successful use of SHIC [70], a classification system of human activities, for the description of museum objects by several British museums. Rather than characterizing the object, only the function or utility of an object for a human activity is regarded. This focus on one uniform aspect ("facet") avoids the ambiguity in the application of other terminological systems. In 1950, the Netherlands Institute for Art History (Rijksbureau voor Kunsthistorische Documentatie or RKD) began its collaboration with Henri Van der Waal on the development of ICONCLASS,[8] with the publication of mounted and annotated photographs of Dutch works of art, the so-called DIAL (Decimal Index of the Art of the Low Countries). From 1950 until 1982, 28 sets of 500 cards were produced, making for a total of 14,000 items. In the RKD images database, which can be consulted via the RKD website, a large number of Dutch works of art is made accessible with the help of ICONCLASS notations. In September 2006, the RKD acquired the rights for the ICONCLASS software from the Royal Academy of Arts and Sciences (KNAW) in Amsterdam. The ICONCLASS System is the only more widespread system for iconographic classification. It is a kind of faceted classification system with a hierarchy of concepts defined by decimal codes. It comprises general, Christian and Greco-Roman subjects. Concepts can be modified by keys to express, for instance, 'head of X'. So far, no formulation as a formal ontology has been undertaken. Van Gendt [76] could only partially represent ICONCLASS in SKOS. Even though it is a genuine aspect of cultural heritage, iconographic classification is not regarded as part of standard museum documentation.

CAMEO is a searchable information center developed by the Museum of Fine Arts, Boston [58]. The MATERIALS database contains chemical, physical, visual, and analytical information on over 10,000 historic and contemporary materials used in the production and conservation of artistic, architectural, archaeological, and anthropological materials. It offers only search by keywords and alphabetic order. The European funded Project CRISATEL developed a system and an ontology employing multiple generalization for art conservation comprising materials, techniques and methods of investigation and intervention for paintings, but the system has not been taken up by the community yet [20].

4.3 KOS of Particulars and Information Extraction

Understanding cultural heritage lives from contextual knowledge and concatenation of facts. Therefore it is most important to be precise about particular

[8] `http://www.iconclass.nl/`

persons, places, historical periods and objects, which appear as the major constituents that connect multiple facts. Relevant resources about particulars are organized as Knowledge Organisation Systems, sometimes also called "ontologies", even though the term does not apply to lists of particulars.

For instance, large reference lists of Persons are maintained by national libraries [10]. The Getty Research Institute maintains ULAN, the Union List of Artist Names [64], and TGN, the Thesaurus of Geographic Names [36], which lists a million of historical and current placenames organized in a hierarchy of geographic spatial inclusion. Also the Alexandria Gazetteer is an important resource of current placenames. In these resources, only the schema and the typologies can be regarded as a kind of ontology. The Alexandria Gazetteer contains an interesting list of feature types to classify places.

A large part of cultural heritage documentation, primary and secondary literature is in textual form. Even though databases have become standard tools for collection descriptions, curators prefer to express the relevant historical facts in free text. Therefore, automated information extraction (IE) becomes more and more important. Extracted information could be used to produce structured metadata and to instantiate ontology-based knowledge representation systems. Full text retrieval systems and text mining systems use to recognize concept names from ontologies. Ontologies should be tailored for this purpose, for instance be enriched with frequent synonyms. To our knowledge, there has been no such attempt for the more popular cultural heritage vocabularies.

So-called Named Entity Recognition of names of persons, things, places [57, 73], or recognition of date and time expressions [52] works reasonably well. Most systems use KOS or "gazetteers" to guess if a name is likely to be a person or a place name. Some languages, like Latin, have more distinct grammatical forms for location expressions, which makes the job to distinguish these categories easier [73]. Note that detecting a name does not mean that the individual has also been identified.

As the core ontologies presented in this chapter show, event information is particularly important for cultural heritage. Automatic event recognition could bring a break-through in the access to relevant historical knowledge. Automatic event recognition is the next step after recognizing named entities and dates. An event can normally be described by the kind of action, the participating things and people, date and place. Event recognition should be combined with NER. So far, there has been not too much work in this direction (for instance, [44,52,77]). An obstacle is the lack of formal ontologies relating characteristic action verbs, such as "printed", "discovered", "broke", "shot" with typical events, such as activities of creation, finding and destroying things, meetings, birth, death and killing.

5 Conclusions

Current ontologies for cultural heritage exhibit a focus on the material and physical aspects of the past. This is quite natural, since "heritage" in the narrower sense implies material evidence of the past. Information about events in the physical world is central to the understanding of heritage information and explicit formal representation of events a key element to integrate heritage information. Interesting is the convergence of core ontologies to very similar forms, which can be integrated, and their independence from a particular "cultural" view. The work of historians is more a detective work than that of a judge. This determines widely the character and focus of cultural heritage ontologies. Information is incomplete. More important than the conclusions is the careful collection of all evidence that could support the one or the other view about the past. In contrast to that, natural sciences would get rid of experimental data after a theory has been sufficiently supported by experiments.

Conclusions and judgment about the past are rather published in scholarly texts than encoded in data structures. This focus may be due to the characteristics of the reasoning in the sector, or just be enforced by the fact that IT methods have penetrated the sector from core documentation and management of physical collections. In the latter case, one may expect that cultural ontologies may in the future extend to other applications in the sector as well. May be formal ontologies dealing with the intellectual structuring of the sector, such as iconography, social interaction, and causation will find more attention in the future. Generally, we expect a greater diversity of conceptualization in the intellectual structure than in the description of material aspects, as represented by the CIDOC CRM.

Since many scholars question the utility of standardized terminology, the formalization of the major terminological systems in the sector is still poor, but this may be overcome by a gradual transfer of know-how and better appreciation of the specifics of cultural conceptualization by ontology engineers. The sector shows enough interest in using ontologies to solve the interoperability of data structures and engages in real implementations. Ontology languages seem to be sufficiently expressive for terminological problems. In the area of data structures semantics, reification problems (i.e. simultaneous use of ontologies and documentation of the discourse about them and documenting facts together with their observation), as well as partially categorical statements cannot sufficiently be described with current ontology languages.

In general, in the years of our collaboration with memory institutions and scholars we found that a major obstacle to introducing advanced computer science methods in the sector is a general underestimation of the complexity of cultural heritage conceptualization by the IT experts, which is equaled by the inability of domain experts to describe their conceptualizations in conscious, objective terms. Whoever wants to deal with the subject effectively must be prepared for a long knowledge engineering phase.

References

1. The CIDOC Conceptual Reference Model, Applications, http://cidoc.ics.forth.gr/uses_applications.html.
2. The CIDOC Conceptual Reference Model, References, http://cidoc.ics.forth.gr/references.html.
3. Conservation OnLine, Resources for Conservation Professionals http://palimpsest.stanford.edu/.
4. HEREIN Thesaurus: European Heritage Network, http://www.european-heritage.net/sdx/herein/thesaurus/introduction.xsp.
5. Simple Knowledge Organisation Systems (SKOS) - home page, http://www.w3.org/2004/02/skos.
6. STAR Project - Semantic Technologies for Archaeological Resources, http://hypermedia.research.glam.ac.uk/kos/star.
7. Successive drafts of the model and minutes of the Harmonisation Groups meetings are available from the CIDOC CRM Web site at http://cidoc.ics.forth.gr/frbr_inro.html and the FRBR Review Group Web site at http://www.ifla.org/VII/s13/wgfrbr/FRBR-CRMdialogue_wg.htm.
8. International Organization for Standardisation, MPEG-7 Overview ISO/IEC JTC1/SC29/WG11 N6828 http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm, 2004.
9. T. Baker, M. Dekkers, R. Heery, M. Patel, and G. Salokhe. What Terms Does Your Metadata Use? Application Profiles as Machine-Understandable Narratives. *Journal of Digital Information 2(2)*, 2(No. 65), 2001.
10. R. Bennett, Chr. Hengel-Dittrich, E. O'neill, and B. Tillett. VIAF (Virtual International Authority File): Linking Die Deutsche Bibliothek and Library of Congress Name Authority Files. In *WLIC2006*, Seoul Korea, 2006.
11. P. Bonora, Ch. Ossicini, and G. Raffa. From Relational Metadata Standards to CRM Ontology: a Case Study in Performing Arts. In *Proc. of CIDOC2006*, Gothenburg, 2006.
12. J.M. Bower and M. et al Baca. *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.
13. A. Cali. Reasoning in Data Integration Systems: Why LAV and GAV Are Siblings. In *Proc. of ISMIS 2003*, Lecture Notes in Computer Science 2871, pages 562–571. Springer, 2003.
14. D. Calvanese, G. Giacomo, M. Lenzerini, D. Nardi, and R. Rosati. Description Logic Framework for Information Integration. In *Proc. of the 6th International Conference on the Principles of Knowledge Representation and Reasoning (KR'98)*, pages 2–13., Trento, Italy, 1998.
15. M. Clauss. Epigraphic Data Bank http://www.manfredclauss.de/gb/index.html. Clauss/Slaby, 2003.
16. N. Crofts, M. Doerr, T. Gill, S. Stead, and M. Stiff. Definition of the CIDOC Conceptual Reference Model. Version 4.2 http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.doc, 2005.
17. A. D'Andrea and G. Ferrandino. Shared Iconographical Representations with Ontological Models. In *Proc. of Computer Applications and Quantitative Methods in Archaeology Conference, CAA2007*, Berlin, Germany, 2-6 April 2007. CAA.

18. M. Doerr. Mapping of the Dublin Core Metadata Element Set to the CIDOC CRM. Technical Report TR-274, FORTH-ICS, July 2000.
19. M. Doerr. The CIDOC CRM - An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine*, 24(3), 2003.
20. M. Doerr. Modelling Learning Subjects as Relationships. *Intuitive Human Interface*, LNAI 3359:200–214, 2004.
21. M. Doerr, J. Hunter, and C. Lagoze. Towards a Core Ontology for Information Integration. *Journal of Digital information*, 4(1)(169), 2003.
22. M. Doerr and D. Kalomoirakis. A Metastructure for Archeological Terminology. In *Proc. of the Computer Applications and Quantitative Methods in Archaeology Conference (CAA 2000)*, Ljubljana, Slovenia, April 18-21 2000.
23. M. Doerr, G. Markakis, M. Theodoridou, and M. Tsikritzis. DIATHESIS: OCR based semantic annotation of newspapers. In *Proc. of the third SEEDI International Conference: Digitization of cultural and scientific heritage*, Cetinje, Montenegro, September 13-15 2007.
24. M. Doerr and A. Sarris, editors. *CAA2002, The Digital Heritage of Archaeology, Proceedings of the 30th Conference*, Heraklion, Crete, April 2003. Hellenic Ministry of Culture.
25. M. Doerr, K. Schaller, and M. Theodoridou. Integration of Complementary Archaeological Sources. In *Computer Applications and Quantitative Methods in Archaeology Conference*, Prato, Italy, April 2004.
26. Ch. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. MIT Press, 1998.
27. J.-Cl. Gardin. The structure of archaeological theories. *Studies in Modern Archaeology Vol. 3: Mathematics and Information Science in Archaeology: A Flexible Framework*, 3:7–25, 1990.
28. P. Gorgels, L. Aroyo, Y. Wang, R. Brussee, l. Rutledge, and N. Stash. Personalized Museum Experience: The Rijksmuseum Use Case. In *Proc. Museums and the Web 2007 Conference*, Toronto, Canada, 2007. Archives & Museums Informatics.
29. A. Grant, editor. *SPECTRUM - The Museum Documentation Standard*. The Museum Documentation Association, 1994.
30. A. Grant, J. Nieuwenhuis, and T. Petersen, editors. *International guidelines for museum object information: the CIDOC information categories*. 1995.
31. M. Greengrass. E-science Challenges in the World of Historical Studies, 2006.
32. T.-R. Gruber. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal of Human-Computer Studies*, 43((5-6)):907–928, 1995. Special issue on Formal Ontology in Conceptual Analysis and Knowledge Representation.
33. N. Guarino. Formal Ontology in Information Systems. In *FOIS'98*, pages 3–15, 1998.
34. M. Guercio, J. Barthelemy, and A. Bonardi. Authenticity Issue in Performing Arts using Live Electronics. In *Proceedings of the 4th Sound and Music Computing Conference (SMC07)*, pages 226–229, Lefkada, Greece, 2007.
35. G. Guizzardi, H. Herre, and G. Wagner. On the General Ontological Foundations of Conceptual Modeling. In *Proc. Conceptual Modeling - ER 2002: 21st International Conference on Conceptual Modeling*, volume 15, pages 65–78, Tampere, Finland, 2002. Springer-Verlag Berlin Heidelberg.
36. P. Harpring. Proper Words in Proper Places: The Thesaurus of Geographic Names. *MDA Information, 2(3), 5-12.*, 2(3):5 12, 1997.

37. R. Heery and M. Patel. Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne*, (Issue 25), 2000.
38. S. Hermon and F. Niccolucci. A Fuzzy Logic Approach to Typology Archaeological Research. In *CAA 2002. The Digital Heritage in Archaeology. Computer Applications and Quantitative Methods in Archaeology*, pages 307–310, Heraklion, Crete, April, 2002 2002. Proceedings of the 30th Conference, CAA.
39. I. Hodder. *The Archaeological Process: An Introduction*. Blackwell Publishers, 1999.
40. J. Hunter. Combining the CIDOC/CRM and MPEG-7 to Describe Multimedia in Museums. *Museums and the Web*, 2002.
41. J. Hunter. *The Role of Information Technologies in Indigenous Knowledge Management*, chapter Chapter 9. Indigenous knowledge and libraries. Australian Academic & Researchers Libraries., 2005.
42. E. Hyvonen, E. Makela, M. Salminen, A. Valo, K. Viljanen, S. Saarela, M. Junnila, and S. Kettula. Museumfinland Finnish Museums on the Semantic Web. *Journal of Web Semantics*, 3((2)):224–241, 2005.
43. C. Kakali, I. Lourdi, Th. Stasinopoulou, L. Bountouri, C. Papatheodorou, M. Doerr, and M. Gergatsoulis. Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. In *Proceeding of 10th International Conference on Asian Digital Libraries*, Hanoi, Vietnam, 10-13 December 2007. To appear.
44. M. Keynes. Ontology-driven Event Recognition on Stories. Technical report, KMI-TR-135, Knowledge Media Institute, 2003.
45. C. Lagoze. Business Unusual: How "event awareness" may breathe life into the catalog. In *Bicentennial Conference on Bibliographic Control in the New Millennium: Confronting the Challenges of Networked Resources and the Web*, Washington, DC., 2000.
46. C. Lagoze and J. Hunter. The ABC Ontology and Model. *Journal of Digital Information 2(2)*, (Article No. 77), 2001.
47. P. LeBoeuf. Frbr and Further. *Cataloging & Classification Quarterly*, 32((4)): 15–22, 2001.
48. P. LeBoeuf. *Functional Requirements for Bibliographic Records (Frbr): Hype or Cure-All*. Haworth Press, Inc, 2005.
49. P. LeBoeuf, Ch. Lahanier, G. Aitken, P. Sinclair, P. Lewis, and K. Martinez. Integrating Museum & Bibliographic Information: The SCULPTEUR Project. In *Proc. ICHIM 2005 Conference*, Paris, 2005.
50. J. Lehmann, S. Borgo, C. Masolo, and A. Gangemi. Causality and Causation in DOLCE. In Vieu L. (eds) Varzi, A.C., editor, *Formal Ontology in Information Systems, Proceedings of the International Conference FOIS 2004*, pages 273–284, Torino, November 4-6 2004. IOS Press Amsterdam.
51. A. Magkanaraki, S. Alexaki, V. Christophides, and D. Plexousakis. Benchmarking RDF schemas for the Semantic Web. In *Proc. First International Semantic Web Conference on the Semantic Web*, number 12, pages 132–146, Sardinia, Italy, 2002. Springer Berlin/Heidelberg.
52. J. Makkonen and H. Ahonen-Myka. Utilizing Temporal Information in Topic Detection and Tracking. In *Proc. of the 7th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2003)*, pages 393–404, Trondheim, Norway, 2003. Springer.
53. A. Maple. Faceted Access: A Review of the Literature, 1995.

54. C. Masolo, S. Borgo, A. Gangemi, N. Guarino, and A. Oltramari. Ontology Library (final). IST Project 2001-33052 WonderWeb: Ontology Infrastructure for the Semantic Web. Deliverable D18., 2003.
55. J. McCarthy. *Programs with Common Sense*, pages 403–418. Semantic Information Processing. MIT Press, Minsky, M. (ed). edition, 1969.
56. C. Meghini, M. Doerr, and N. Spyratos. Sharing Co-reference Knowledge for Data Integration. *DELOS Network of Excellence*, 2007. In Second DELOS Conference - Working Notes. Pisa, Italy.
57. A. Mikheev, M. Moens, and C. Grover. Named Entity Recognition without Gazetteers. In *EACL*, pages 1–8, 1999.
58. Boston Museum of Fine Arts. CAMEO: Conservation & Art Material Encyclopedia `http://cameo.mfa.org/`.
59. J. Mylopoulos, A. Borgida, M. Jarke, and M. Koubarakis. Telos: Representing Knowledge about Information Systems. *ACM Transactions on Information Systems*, 8((4)):325–362, 1990.
60. International Council of Museums ICOM. ICOM Definition of a Museum `http://icom.museum/statutes.html#2`, 2001.
61. International Council of Museums ICOM. ICOM Code of Ethics for Museums `http://icom.museum/ethics.html`, 2006.
62. M. Patel, T. Koch, M. Doerr, and Chr. Tsinaraki. D5.3.1: Semantic Interoperability in Digital Library Systems. Deliverable, Project no.507618, DELOS, A Network of Excellence on Digital Libraries, June 2005.
63. JL. Peterson. *Petri net theory and the modeling of systems.* Prentice-Hall, 1981.
64. Oxford University Press, editor. *Art & Architecture Thesaurus.* Getty Trust Publications, 1994.
65. SR. Ranganathan. *A descriptive account of Colon Classification.* Bangalore: Sarada Ranganathan Endowment for Library Science., 1965.
66. PA. Reed. CIDOC Relational Data Model. A Guide. ICOM/CIDOC, 1995.
67. V. Roux and P. Blasco. Logicisme et format scd: dune epistemologie pratique a de nouvelles pratiques editoriales hermes. 2004. CNRS - editions.
68. K.G Saur, editor. *IFLA Study Group on the functional requirements for bibliographic records. Functional requirements for bibliographic records: final report,* `http://www.ifla.org/VII/s13/frbr/frbr.htm`, volume vol. 19 of *New Series.* UBCIM Publications, Munich, 1998.
69. E. Schulten, H. Akkermans, G. Botquin, M. Doerr, N. Guarino, N. Lopes, and N. Sadeh. Call for Participants: The E-CommerceProduct Classification Challenge. *IEEE Intelligent System*, 16((4)):86–c3., 2001.
70. The SHIC Working Party. *Social, Historical and Industrial Classification (2nd edition).* The Museum Documentation Association, Cambridge, UK, 1993.
71. B. Smith. Ontology. *The Blackwell Guide to the Philosophy of Computing and Information*, pages 155–166, 2003. Floridi, L. (ed). Oxford: Blackwell.
72. B. Smith. Against Idiosyncrasy in Ontology Development. In B. Bennett and Fellbaum C., editors, *Formal Ontology in Information Systems (FOIS 2006)*, pages 15–26, Amsterdam, 2006. IOS Press.
73. DA. Smith and G. Crane. Disambiguating Geographic Names in a Historical Digital Library. In *Proc. of the 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001)*, pages 127–136, Darmstadt, Germany, 2001. Springer.

74. D. Tudhope and C. Binding. A Case Study of a Faceted Approach to Knowledge Organisation and Retrieval. *DigiCULT, Resource Discovery Technologies for the Heritage Sector*, (Thematic Issue 6):2833, 2004.
75. D. Tudhope, C. Binding, D. Blocks, and D. Cunliffe. Query Expansion via Conceptual Distance in Thesaurus Indexed Collections. *Journal of Documentation*, 62((4)):509–533, 2006. Emerald Group Publishing Limited.
76. M. vanGendt, A. Isaac, L. vanderMeij, and S. Schloback. Semantic Web Techniques for Multiple Views on Heterogeneous Collections: A Case Study. In *Proc. of the 10th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2006)*, pages 426–437, Alicante, Spain, 2006. Springer.
77. M. Vargas-Vera and D. Celjuska. Event Recognition on News Stories and Semi-Automatic Population of an Ontology. *Web Intelligence*, pages 615–618, 2004.
78. SM. Vinas. *Contemporary Theory of Conservation*. Elsevier Oxford, 2005.