Enhancing the Semantic Interoperability of Multimedia through a Core Ontology

Jane Hunter
DSTC Pty Ltd
University of Qld, Australia
jane@dstc.edu.au

Abstract

A core ontology is one of the key building blocks necessary to enable the integration of information from diverse multimedia sources. A complete and extensible ontology that expresses the basic concepts that are common across a variety of domains and media types and that can provide the basis for specialization into domain-specific concepts and vocabularies, is essential for well-defined mappings between domain-specific knowledge representations (i.e., metadata vocabularies) and the subsequent building of a variety of services such as cross-domain searching, tracking, browsing, data mining and knowledge acquisition. As more and more communities develop metadata *application profiles* which combine terms from multiple vocabularies (e.g., Dublin Core, MPEG-7, MPEG-21, CIDOC/CRM, FGDC, IMS) such a core ontology will provide a common understanding of the basic entities and relationships which is essential for semantic interoperability and the development of additional services based on deductive inferencing.

In this paper we first propose such a core ontology (the ABC model) which was developed in response to a need to integrate information from multiple genres of multimedia content within digital libraries and archives. Although the MPEG-21 RDD was influenced by the ABC model and is based on a model extremely similar to ABC, we believe that it is important to define a separate and domain-independent top-level extensible ontology for scenarios in which either MPEG-21 is irrelevant or to enable the attachment of ontologies from communities external to MPEG, for example, the museum domain (CIDOC/CRM) or the bio-medical domain (ON9.3).

We then evaluate the ABC model's ability to mediate and integrate between multimedia metadata vocabularies by illustrating how it can provide the foundation to facilitate semantic interoperability between MPEG-7, MPEG-21 and other domain-specific metadata vocabularies. By expressing the semantics of both MPEG-7 and MPEG-21 metadata terms in RDF Schema/DAML+OIL (and eventually the Web Ontology Language (OWL)) and attaching the MPEG-7 and MPEG-21 class and property hierarchies to the appropriate top-level classes and properties of the ABC model, we have defined a single distributed machine-understandable ontology. The resulting distributed machine-processable ontology provides semantic knowledge which is non-existent within declarative XML schemas or XML-encoded metadata descriptions. Finally, in order to illustrate how such an ontology will contribute to the interoperability of data and services across the entire multimedia content delivery chain, we describe a number of valuable services which have been developed or could potentially be developed using the resulting merged ontologies.

Keywords: Ontology, Multimedia, MPEG-7, MPEG-21, ABC, Semantic Interoperability

1. Introduction

Audiovisual resources in the form of still pictures, graphics, 3D models, audio, speech, video will play an increasingly pervasive role in our lives, and there will be a growing need to enable computational interpretation and processing of such resources and the automated generation or extraction of semantic knowledge from them. The Moving Pictures Expert Group (MPEG) has developed the Multimedia Content Description Interface (MPEG-7) [1], which aims to define a rich set of standardised tools to enable machines to generate and understand audiovisual descriptions for retrieval, categorisation and filtering purposes. Significant progress has been made on automatic segmentation, scene-change detection, and the recognition and detection of low-level features for multimedia content - however little progress has been made on machine-generation of semantic descriptions of audiovisual information.

The objectives of MPEG's latest initiative, MPEG-21 [2] (ISO/IEC 18034-1), are to:

- Provide a vision for a multimedia framework to enable transparent and augmented use of multimedia resources across a wide range of networks and devices to meet the needs of all users;
- Facilitate the integration of components and standards in order to harmonise technologies for the creation, management, manipulation, transport, distribution and consumption of content;

- Provide a strategy for achieving a multimedia framework by the development of specifications and standards based on well-defined functional requirement through collaboration with other bodies.

Although MPEG-21 aims to eventually focus attention on semantic description schemes and knowledge integration, it's current effort is concentrated on the declaration and identification of digital items and the definitions of a Rights Expression Language and a Rights Data Dictionary.

Both MPEG-7 and MPEG-21 have also concentrated on defining XML representations of their description schemes using the XML Schema language [3]. Although XML Schemas provide support for explicit structural, cardinality and datatyping constraints, they provide little support for the semantic knowledge necessary to enable efficient and flexible mapping, integration, and knowledge acquisition.

Central to the development of semantic analysis and knowledge mining tools for multimedia is the need for ontologies which express the key entities and relationships used to describe multimedia in a formal machine-processable representation (e.g, RDF Schema [4], DAML+OIL [5] or the recently published Web Ontology Language (OWL) working draft [6]). The knowledge representation provided by such ontologies can be used to develop sophisticated services and tools which perform knowledge-based reasoning, knowledge adaptation, knowledge integration and sharing and knowledge acquisition, specifically for semantically-rich audiovisual content.

Consequently we have developed the following three ontologies which we will describe in detail:

- A top-level core ABC ontology [7];
- An ontology for MPEG-7;
- An ontology for MPEG-21;

We believe that the definition of a seperate domain-independent top-level ontology has a number of advantages. Currently there are overlaps, redundancies and incompatibilities between the semantics of terms used in both MPEG-7 and MPEG-21. We believe that the provision and employment of a single common model with one set of semantic definitions, by both MPEG-7 and MPEG-21, would greatly facilitate the efficiency and interoperability of multimedia delivery systems based on these two standards.

A separate top-level core ontology also enables other domain-specific ontologies to be incorporated as they are required. Different communities are developing application profiles which combine metadata elements from a number of existing standardized metadata vocabularies to satisfy their own unique requirements. For example, there are educational communities who want to combine IMS [8] with Dublin Core [9] and MPEG-7 to describe multimedia learning objects. Similarly the geo-spatial and cultural heritage communities are combining FGDC [10] and the CIDOC/CRM [11] (respectively) with MPEG-7 to describe multimedia within their domains. A core ontology provides the glue to enable these domain-specific ontologies to be harmonized with the MPEG-7 and MPEG-21 ontologies.

The representation of the semantics of MPEG-7 and MPEG-21 terms as machine-processable ontologies will also enable the construction of efficient knowledge-based multimedia systems which are capable of automatically extracting and aggregating semantic information (objects, events, properties, relations) about the audiovisual data. The extracted semantic metadata can be used for classification, summarisation, indexing, searching and efficient retrieval of multimedia content. For example, given the appropriate inferencing rules, a complete distributed and well-structured ontology could conceivably enable the "subject" of an image or the "genre" of a video to be automatically deduced from a combination of low-level MPEG-7 visual or audiovisual descriptor values.

Hence in the remainder of this paper we describe a simple event-aware model (the Harmony Project's [12] ABC model [7]) which can be used to describe, record and yet clearly differentiate between the overlapping, interacting and recursive events in the life cycle of a multimedia asset. We then analyse both the MPEG-7 and MPEG-21 metadata models to determine those domain-specific aspects which are not covered by ABC. We represent these aspects in RDF Schema[4]/DAML+OIL[5] (but will move to OWL [6] when it is more stable) and determine the most appropriate attachment points for adding the semantics of MPEG-7 and MPEG-21 to the ABC ontology. The final outcome is in essence, a single, machine-understandable, extensible ontology (conceptually illustrated in Figure 1) which is distributed across the ABC, MPEG-7 and MPEG-21 namespaces. Finally, in order to illustrate how such an ontology will contribute to the interoperability of data and services across the entire multimedia content delivery chain, we describe a number of scenarios and implementations which have been developed or are currently under development using the resulting merged ontologies.
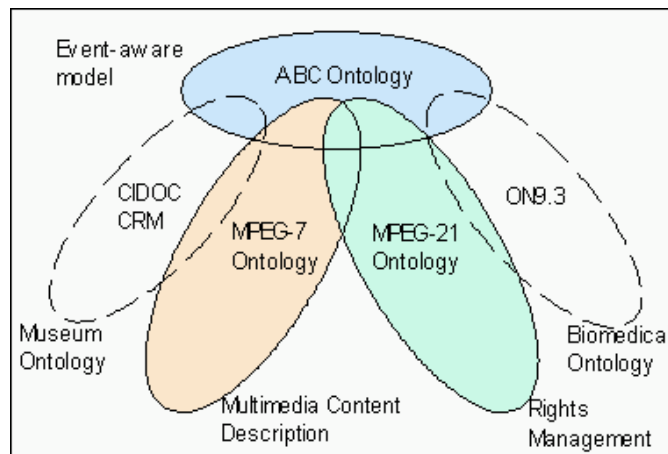
Figure 1 - Proposed Approach

2. The ABC Ontology

There are a number of metadata models which recognise the importance of events or actions in unambiguously describing resources and their lifecycles and in facilitating semantic interoperability between metadata domains.

The Harmony project's ABC model [7] was motivated by the recognition that many existing metadata approaches are based on a resource-centric traditional cataloguing approach which assumes that the objects being described, and therefore their attributes, are more or less stable. Such an approach is inadequate for modeling the creation, evolution, transition, usage and rights of objects over time or for supporting advanced queries such as *who was responsible for what, when and where?*

The <indecs> 2rdd model [14] (which was submitted to the MPEG-21 CfP and adopted as the underlying model for the Rights Data Dictionary) uses *Actions* as the starting point for their model so as to enable interoperability between interdependent rights metadata and descriptive metadata.

Both the <indecs> and ABC models have informed each other at different stages since their inceptions. However the <indecs> model is more concerned with the description of rights and permissions whilst the ABC model has been deliberately designed as a primitive ontology so that individual communities are able to build on top of it. ABC provides a set of domain-independent base classes which act as either attachment points for domain-specific properties or super classes which can be sub-classed to create domain-specific classes. System builders might use the ABC principles as the basis for implementing tools that permit mapping across descriptions in multiple metadata formats. Hence we have chosen to use the ABC model as the starting point or foundation to which we will attach the semantics of the MPEG-7 terms and the MPEG-21 rights data dictionary and thus enable semantic interoperability between these two initatives.

The ABC model has been designed to model physical, digital and analogue objects held in libraries, archives, and museums and on the Internet. This includes objects of all media types, text, image, video, audio, web pages, and multimedia. It can also be used to model abstract concepts such as intellectual content and temporal entities such as performances or lifecycle events that happen to an object. In addition the model can be used to describe other fundamental entities that occur across many domains such as: agents (people, organizations, instruments), places and times.

A detailed specification of all ABC classes and Properties and an RDF schema representation are available at [15]. The top-level class hierarchy is illustrated in Figure 2 below.
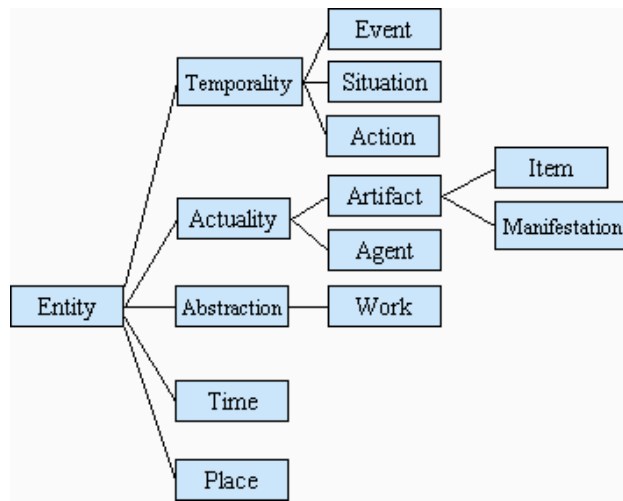
Figure 2 - ABC Class Hierarchy

The primitive category at the core of the ABC ontology is the *Entity*. In sections 2.1-2.3 below, we describe the three main classes (other than *Time* and *Place*) which lie at the second level of the ontology: *Temporality*, *Actuality*, and *Abstraction*.

2.1 Temporality Category

A distinguished aspect of the ABC model is the manner in which it explicitly models time and the way in which properties of objects are transformed over time. Other descriptive models such as AACR2 [16] and Dublin Core [9] imply some time semantics. For example, the DC date element and its qualifiers [17] created and modified express events in the lifecycle of a resource. However by expressing these events in this second-class manner (i.e., not making the temporal entities ontological entities) it becomes difficult to associate agent responsibility with those events and connect them with changes in state of the resource. In contrast, the ABC model makes it possible to unambiguously express situations in which object properties exist, the transitions that demark those situations, and the actions and agency that participate in those transitions. In brief summary, ABC models time as follows:

- A *Situation* provides the context for framing time-dependent properties of (possibly multiple) entities. Entities, such as a person or a document, may have properties that exist only in the context of a situation and other properties that are constant across the context of a description. For example, in a description of the "my car", the property "has make Honda" is constant across the entire description, but the property "has color red" applies before I paint it, and the property "has color green" applies after I paint it. Concurrently, the green paint can has the property "is full" before I paint the car and the property "is empty" after I paint it, but always has the property "produced by Dulux". ABC models this through the use of situations to which are bound existential facets of entities, which provide the attachment points for situation-specific properties of entities (the color of the car and the fullness of the paint can). These existential facets can co-exist with a single universal facet of each entity, to which the time-independent properties are bound (e.g., the model of the car or the producer of the paint can). From the perspective of first-order logic, the existential facet corresponds to "there exists a situation in which an instance of the entity exists with a property set", and the universal facet corresponds to "for all situations in the description the entity exists with certain property set".
- An *Event* marks a transition from one situation to another. Events always have time properties. The effect is that a situation implicitly has time duration as defined by its bounding events (associated via precedes and follows properties). As an example, the model could express the loan of the Mona Lisa to the Metropolitan Museum for a fixed period (e.g., May 1, 2000 through May 30, 2001) as follows: an existential facet of the Mona Lisa with a property "located at the Metropolitan" could be associated with a situation that is related via precedes and follows properties with two events, one of which gives the time of the loan, the other the time of the return. The use of the hasPresence property with an Event models the fuzzy concept of the participation of an Agent in the Even. More precise notions of participation require the Action concept as described below.
- An *Action* provides the mechanism for modeling increased knowledge about the involvement and responsibility of agents in events. Specifically, it denotes a verb in the context of an event. The hasAction property connects an Action to an Event. Actions provide the ontological framework for expressing these "verbs" and associating the specific agency with them. In addition, the *involves* property (and its sub-properties) makes it possible to explicitly associate actions with effects on existential facets of entities. Finally, the hasParticipant property, and its possible domain-specific sub-properties, makes it possible to precisely specify the association of an Agent with an Action. The combination of these makes it possible to clearly state entity derivations (e.g., translations, reformatting, etc.) and modifications and who or what is responsible for them.

2.2 Actuality Category

The Actuality ontology category encompasses entities that are sensible - they can be heard, seen, smelled, or touched. This

contrasts with the Abstraction category, which encompasses concepts. As described in Section 2.1 entities that are Actualities, can have one universal or time- independent facet and many existential or time- dependent facets. ABC expresses this notion through the inContext property that associates an Actuality with a Situation. For example, an ABC description of Bill Clinton might have an existential Actuality with property "President of the United States" that is related via the phaseOf property to one universal Actuality with the property "born in Arkansas". The existential facet would be related via the inContext property to a Situation that follows an Event representing Clinton's election in 1992. The result is a statement that expresses the "sameness" of the two entities (they are both "Bill Clinton"), but the fact that one is an existential facet and one is a universal facet. The ABC model also incorporates intellectual creation semantics influenced by the IFLA FRBR [18]. A sub-category of Actuality, Artifact, expresses sensible entities that are tangible realizations of concepts, and that can be manifested in multiple ways; e.g., as Manifestations and Items as expressed in the FRBR.

2.3 Abstraction Category

The Abstraction category makes it possible to express concepts or ideas. Entities in this category have two notable characteristics:

1. They are never in the context of a situation. While it can be argued that an idea is "born" at some time, ABC treats the "birth of an idea" when it is manifested in some sensible way; e.g., when it is told, demonstrated, or shown in some manner.
2. Correspondingly, ideas cannot exist in isolation in the model. They must be bound to some Actuality through the hasRealization property.

The main use of the Abstraction category is to express the notion of Work in the FRBR [12] sense; that is, as a means of binding together several Manifestations of an intellectual Expression. For example, an ABC description of the Hamlet might instantiate a Work that binds the folio manifestation, a Stratford performance, and a Penguin edition.

2.4 An Illustrative Example

Consider the following example narrative:

The book, "Charlie and the Chocolate Factory" was written by Roald Dahl in 1964. The first edition (a hardcover, illustrated by Joseph Shindleman) was published in 1985 by Knopf. A second edition was published in 1998 by Puffin. It was a paperback illustrated by Quentin Blake. In 1999, a 3 hour audiocassette recording of the book was produced by Caedmon. It was narrated by Robert Powell.

The graphical representation corresponding to the ABC model for this example is shown in Figure 3 below. The narrative is modelled as a sequence of Events and States. Events have Actions and Agents associated with them. States provide the context for Manifestations. Manifestations which are based on a common "idea" or intellectual property, are linked by a common Work.
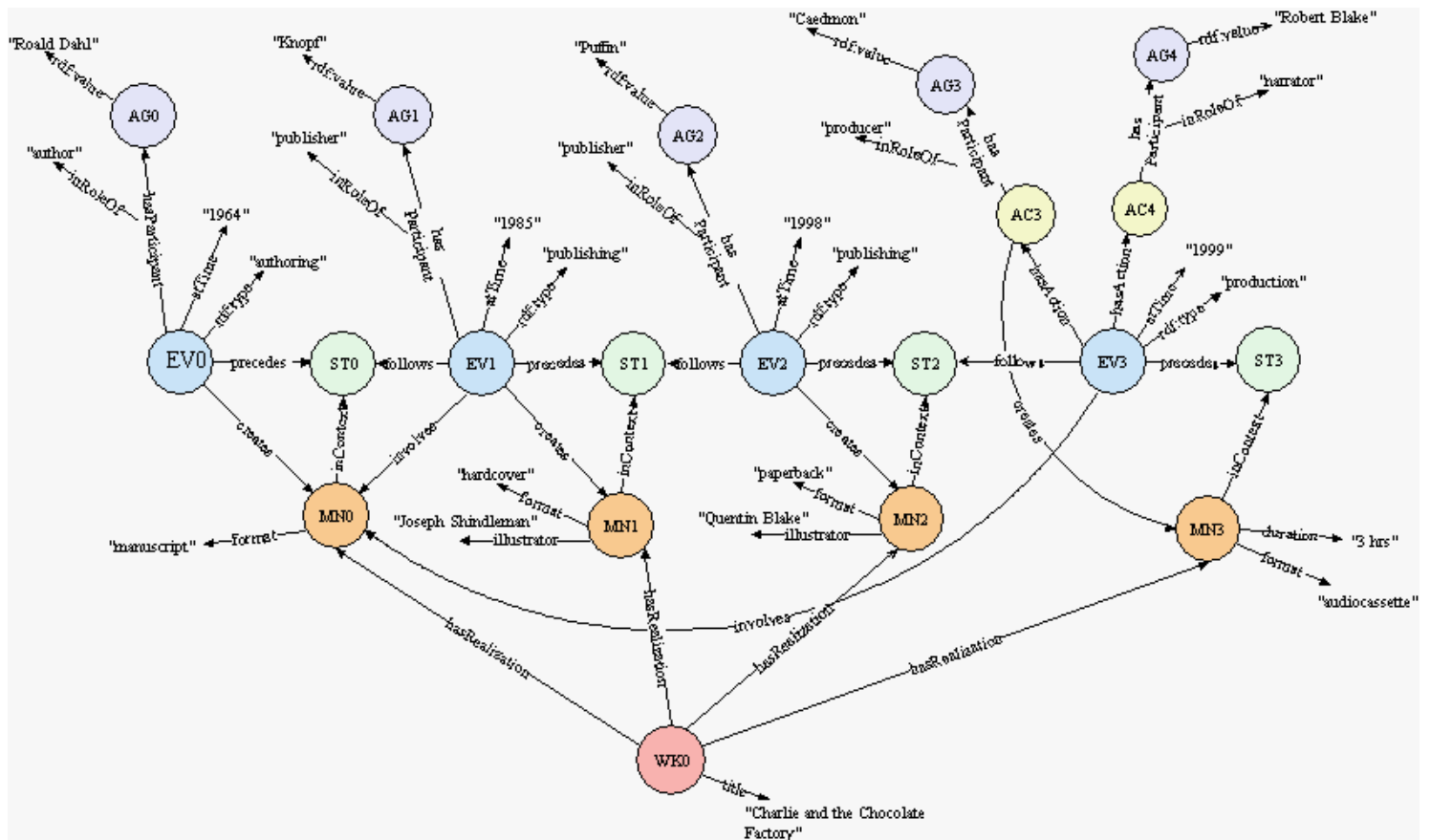
Figure 3 - ABC Model of a Typical Example

3. Adding Multimedia Metadata Semantics through the MPEG-7 Ontology

In March 2002, the Moving Pictures Expert Group (MPEG), a working group of ISO/IEC, published the final standard for MPEG-7 [1], the "Multimedia Content Description Interface", a standard for describing multimedia content. The goal of this standard is to provide a rich set of standardized tools to enable both humans and machines to generate and understand audiovisual descriptions which can be used to enable fast efficient retrieval from digital archives (pull applications) as well as filtering of streamed audiovisual broadcasts on the Internet (push applications). MPEG-7 can describe audiovisual information regardless of storage, coding, display, transmission, medium, or technology. It addresses a wide variety of media types including: still pictures, graphics, 3D models, audio, speech, video, and combinations of these (e.g., multimedia presentations). Initially MPEG-7 definitions (description schemes and descriptors) were expressed solely in XML Schema [3,19,20]. XML Schema has been ideal for expressing the syntax, structural, cardinality and datatyping constraints required by MPEG-7.

However semantic interoperability is necessary to enable systems to exchange data (e.g., metadata descriptions), to understand the precise meaning of that data and to translate or integrate data across systems or from different metadata vocabularies. At the 57th MPEG meeting in Sydney, it was recognized that there was a need to formally define the semantics of MPEG-7 terms; and to express these definitions in a machine understandable, interoperable language. RDF Schema [4] was the obvious choice due to its ability to express semantics and semantic relationships through class and property hierarchies and its endorsement by the W3C's Semantic Web Activity [21]. Consequently the Adhoc Group for MPEG-7 Semantic Interoperability was established and these requirements formed the mandate. An MPEG-7 ontology was developed and expressed in RDF Schema and DAML+OIL extensions [22,23]. The extensions provided by DAML+OIL were necessary to satisfy certain requirements such as the need for multiple ranges and sub-class specific constraints. The basic class hierarchy of MPEG-7 content and segments is shown in Figure 4 below e.g., the MPEG-7 class *VideoSegment* is a subclass of both *Video* and *Segment*.
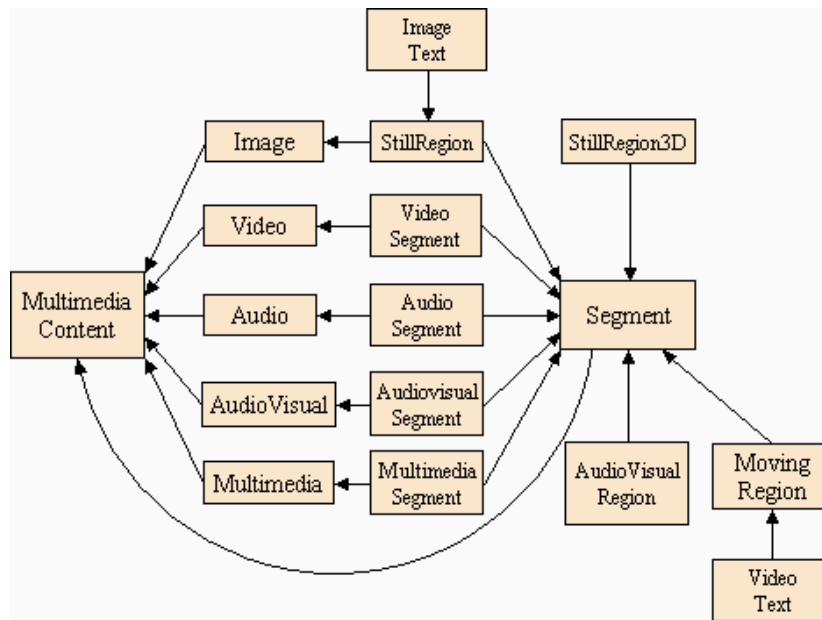
Figure 4 - MPEG-7 Multimedia and Segement Class Hierarchy

Associated with each of the subclasses in Figure 4 are various properties which define permitted relationships between the segment classes corresponding to specific structural or organisational description schemes and the permitted audio, visual and audiovisual attributes associated with different types of multimedia segments. We will not describe these properties here but detailed descriptions are available from [23].

ABC is designed to provide a top level set of classes and properties which can act as attachment points for domain-specific metadata ontologies. In the case of the MPEG-7 ontology, the obvious attachment point for the MPEG-7 class hierarchy is the ABC Manifestation class, as shown in Figure 5 below.
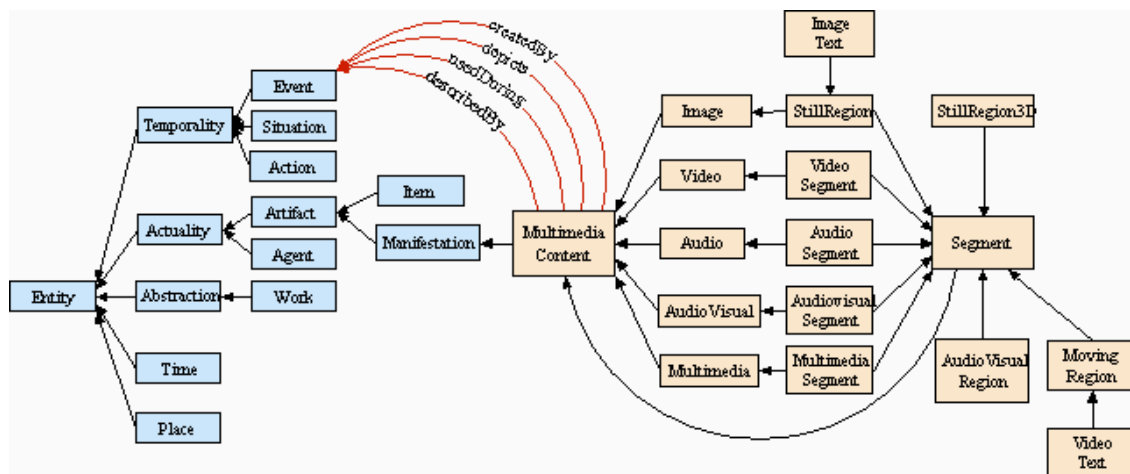

Figure 5 - Extending ABC with MPEG-7

By basing the MPEG-7 ontology on top of the ABC event-aware model, it then becomes possible to re-use the underlying event model within different contexts within MPEG-7. The red property arcs between the *MultimediaContent* class and the *Event* class in Figure 5 illustrate how the Event model can be used to describe multimedia creation, usage and metadata attribution events, as well as the semantic events occurring within the content (as defined in the MPEG-7 SemanticDS [1]).

The modified version of the MPEG-7 ontology, based on ABC, and represented in RDF Schema, is available at: [26].

We believe that this more modular and object-oriented approach could usefully be employed by the MPEG-7 Multimedia Description Schemes [1] to resolve the overlaps, redundancies and incompatibilities which currently exist between the definitions of certain concepts which occur across multiple DSs. For example the *event* concept could be better represented by adding an EventDS to the MPEG-7 Basic Description Tools and modifying the MPEG-7 Creation and Production, DescriptionMetadata, Usage and Usage History and Semantics description tools so that they re-use this common EventDS. Such an approach would also contribute significantly to the ease with which MPEG-7 and MPEG-21 can be harmonized.

4. Adding Rights Metadata Semantics through the <indecs> ontology

The <indecs>2rdd model [14] was submitted to ISO IEC JTC1/SC29/WG11 (MPEG) by a consortium of organisations (consisting of Accenture, ContentGuard, EDItEUR, Enpia Systems, International DOI Foundation, MMG [a subsidiary of Dentsu], the Motion Picture Association and the Recording Industry Association of America with the International Federation of the Phonographic Industry), in response to the MPEG-21 Call for Proposals for a Rights Data Dictionary and a Rights Expression Language [24]. The <indecs>2rdd model aims to provide a comprehensive framework and initial vocabulary for the expression of any form of right or agreement that supports commerce in Digital Items. A decision was made at the 58th MPEG meeting in Pattaya, Thailand to base Part 6 of MPEG-21 (ISO/IEC 21000-6), the MPEG-21 Rights Data Dictionary, on the <indecs> model.[25]

The <indecs>/MPEG-21 RDD data model is similar to the ABC model in that it is based on *Actions* and *Contexts* (Events and Situations). The basic Context model which defines the fundamental relationships between entities in <indecs> is shown in Figure 6 [14]. Agents act through Actions which occur in a Context - either an Event or Situation. Contexts have Inputs and Outputs which are Resources of various different types.
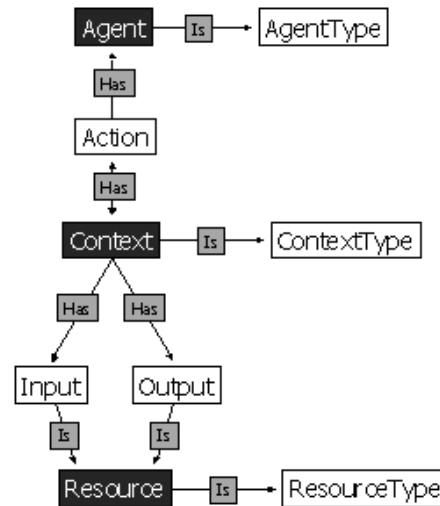

Figure 6 - The MPEG-21 RDD Context Model [7]

Figure 7 below illustrates most of the class hierarchies for the MPEG-21 RDD model and also serves to demonstrate how the ABC and MPEG-21 RDD models differ. (Due to space limitations, Figure 7 does not show the details of the Agent or Event subclass hierarchies, however these correspond closely to the Action class hierarchy which is shown here.)
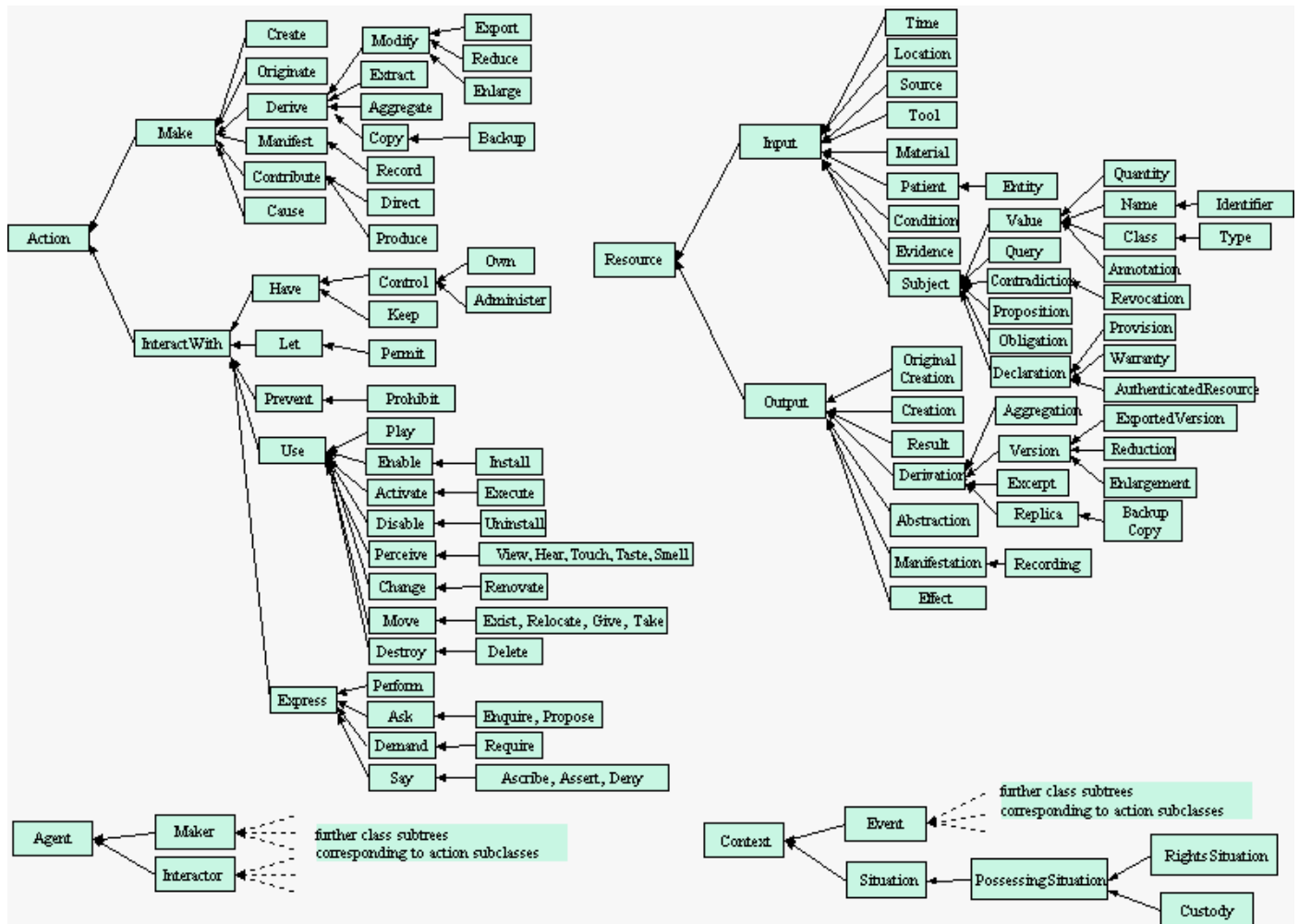
Figure 7 - The MPEG-21 RDD Class Hierarchy

In order to be able to attach the MPEG-21 semantic definitions to the appropriate places in the ABC ontology, whilst still maintaining correct meaning and not introducing inconsistencies or redundancies, we must first understand the differences, intersection points and overlaps between the two models.

The MPEG-21 RDD model has four top-level or superclasses: Action, Resource, Agent and Context. The Action, Agent, Event and Situation classes are identical to these same classes in the ABC model. In addition, we ascertained the following differences between the ABC and <indecs> models:

- Within the MPEG-21 RDD model, *Resources* are directly input to or output from Events. Within ABC on the other hand, *Situations* precede and follow Events. ABC *Situations* provide the context for either Manifestations or phases of manifestations which are being acted on within Events and Actions - this approach enables us to model properties which change over time.
- The ABC model uses the *precedes* and *follows* properties to relate Events to Situations, rather than *hasInput* and *hasOutput*. This is because we believe that the terms "input" and "output" imply causality, which may not always be the case. For example one may simply want to model a series of situations and events which occur in chronological order but are not actually generated or caused by one another.
- The MPEG-21 RDD model explicitly defines *Input* and *Output* class hierarchies to specify the precise kinds of interactions and relationships which occur between Resources, Events and Actions. The ABC model uses properties and sub-properties such as: *involves, hasPatient, usesTool, hasResult, creates* to model these interactions. It would be possible for ABC to extend these property hierarchies to support those MPEG-21 RDD *Input* and *Output* subclasses which are not currently supported.
- ABC defines *Time* and *Place* as top-level classes whereas the MPEG-21 RDD model defines *Time* and *Location* as subclasses of *Input*.
- The ABC *Artifact* class is essentially the same as the MPEG-21 RDD *Output* class. However whereas ABC only defines the Manifestation and Item (Copy) classes, MPEG-21 defines a more complex Derivation subclass hierarchy under Output.
- the MPEG-21 RDD model defines highly detailed subclass hierarchies for Action, Event, Agent and Situation which can simply be attached to the ABC *Action*, *Event* and *Agent* and *Situation* classes.

The above analysis reveals that it is possible to reconcile the MPEG-21 RDD model with the ABC model and to add support for rights metadata modeling, through judicious selection of appropriate attachment points. Using a colour scheme, Figure 8 below

illustrates, how we believe the MPEG-21 RDD ontology can best be attached to the ABC ontology to add specific support for rights metadata semantics. ABC classes are blue, MPEG-7 classes are orange and <indecs> classes are green. The MPEG-21 *Event, Situation, Action* and *Agent* sub-class hierarchies can be directly attached to the equivalent ABC classes. Due to space limitations we have not explicitly shown these subclass hierarchies in Figure 8. The MPEG-21 *Derivation* subclass tree can be attached to the ABC *Artifact* class. The ABC *Item* class is equivalent or synonymous to the MPEG-21 *Replica* class.
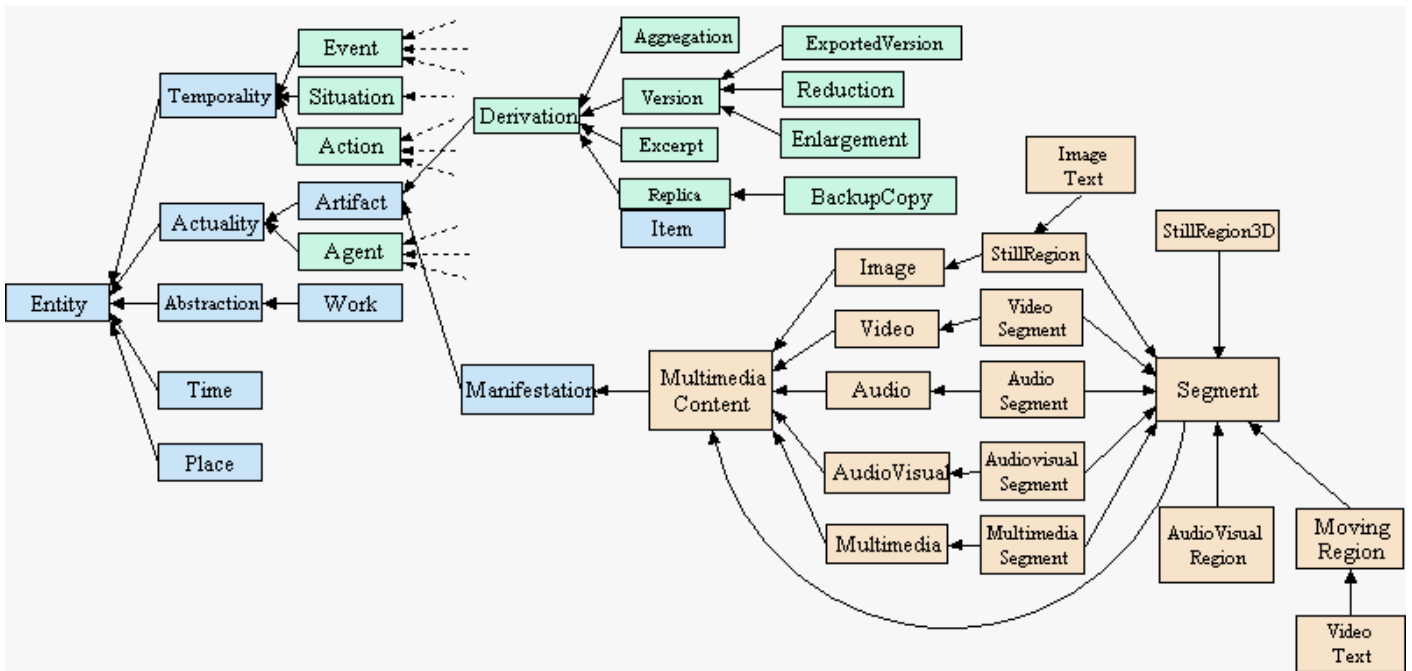
Figure 8 - The Class Hierarchy for the Aggregate Ontology

The remaining unsupported MPEG-21 *Input* subclasses can be accomodated through a sub-property hierarchy attached to the ABC *involves* property. Similarly any unsupported *Output* subclasses can be expressed as sub-properties of the ABC *hasResult* property. These additional sub-properties are shown in green in the property hierarchy diagrams in Figure 9.
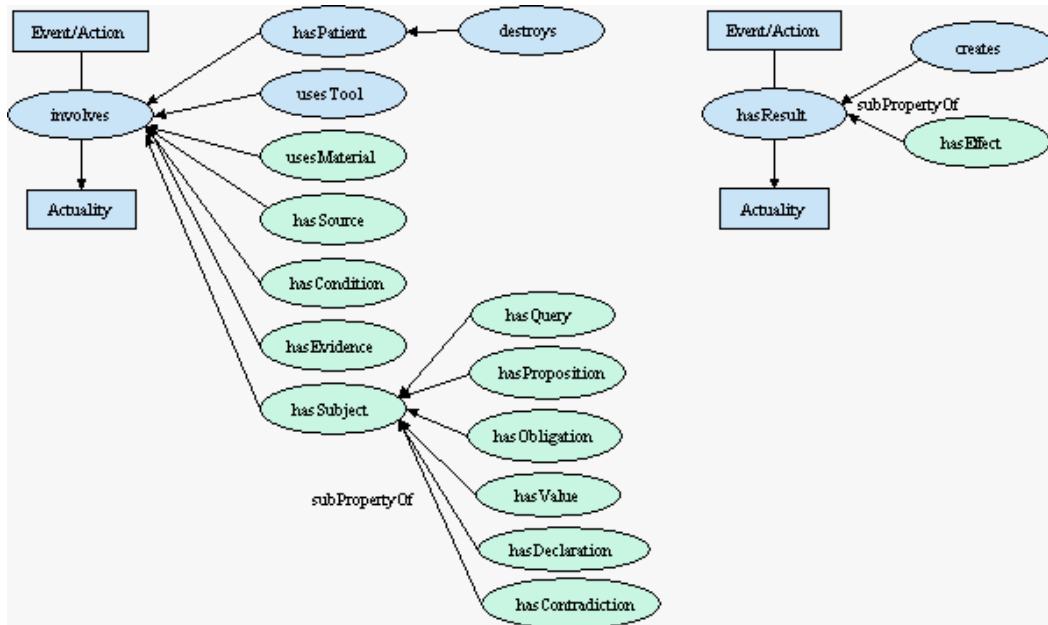

Figure 9 - The Extended Property Hierarchy for the Aggregate Ontology

Using the class and property hierarchies illustrated in Figures 8 and 9 we have been able to generate an RDF Schema/DAML+OIL representation of the MPEG-21 RDD ontology which is available at: [27].

DAML+OIL provides certain extensions to RDF Schema such as multiple range constraints, boolean combinations of classes and the ability to specify synonyms as equivalent classes and properties, which are required for our application. The ability to express equivalent classes and properties also enables mappings between metadata vocabularies to be defined within the ontology (e.g., mappings between XrML and <indecs>). The Web Ontology Working group [28] of the Semantic Web Activity

[21] is working on the development of a machine-readable markup language (OWL) that will allow adopters to define their own ontological content. A working draft of OWL [6] was recently published but future extensions and refinements are expected. We envisage that future versions of the ABC, MPEG-7 and MPEG-21 ontologies will be represented in OWL, when it becomes more stable.

---

5. Some Example Applications

There are a number of important scenarios which occur within information systems for which the semantic interoperability provided by a core ontology is essential:

- to enable a single search interface across heterogeneous metadata descriptions and content in distributed archives;
- to enable mapping between metadata vocabularies;
- to enable the integration or merging of descriptions which are based on complementary but possibly overlapping metadata schemas or standards;
- to enable different views of the one underlying and complete metadata description, depending on the users' particular interest, perpective or requirements.

In addition, the combination of the visual, audio and dynamic aspects of multimedia content together with the knowledge reasoning capabilities provided by semantic web technologies (and in particular, the ABC, MPEG-7 and MPEG-21 ontologies described above) generates a number of new possible scenarios:

- the ability to infer semantic descriptions or to detect conceptual entities (events, objects, places, actors) from combinations of low-level features - using pre-defined inferencing rules;
- the ability to infer semantic relationships between multimedia resources from the existing metadata and to map these to spatio-temporal relationships in order to generate coherent intelligent multimedia presentations.

In the following sub-sections we describe actual implementations of these scenarios which are possible only because of the semantic reasoning capabilities provided by multimedia ontologies, described above.

5.1 Query Mediation

One of the key objectives of a core ontology is to enable query mediation across heterogenous objects of different media types and described using different metadata vocabularies. In a collaboration between the Harmony project [12] and the CIMI Consortium [29], four CIMI members (Australian Museums Online (AMOL), Natural History Museum of London, Research Libraries Group/Library of Congress and theNational Museum of Denmark) contributed approximately 100 museum metadata records and the associated multimedia digital objects to enable an evaluation of the ABC model and its usability as means of mapping between disparate metadata ontologies. A detailed description of the images and data provided by these organisations is available at [30].

In order to evaluate the ABC model's ability to act as a query mediator, the metadata records from each organisation were mapped to the ABC model, represented as XML records and a single search interface was built using the Kweelt XQuery implementation. An online demonstration, available at [31], demonstrates the richer queries possible which can retrieve details of the events, agents or actions associated with resources. This exercise revealed that the mapping exercise was manual and tedious the first time for each organisation. But after it had been done once, an XSLT program could be written which automated the mapping process for the remainder of the records. This exercise also revealed that the ABC model was capable of supporting the metadata requirements of each of the organisations and of acting as a query mediator across their collections. Figure 10 illustrates the high level architecture of the search interface.
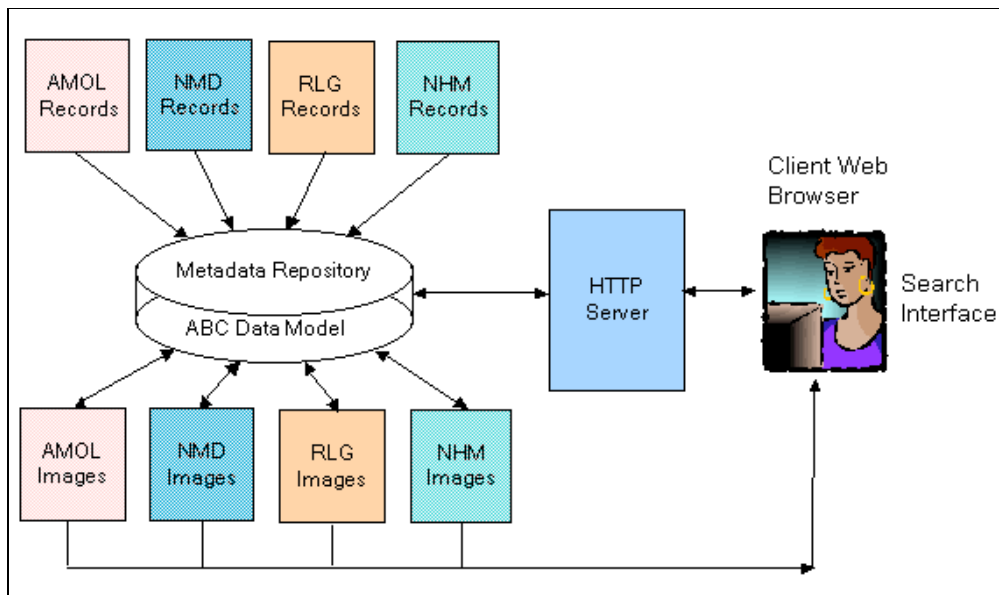
Figure 10 - ABC Query Mediation across CIMI Collections

5.2 Ontology Harmonization

A core ontology should support extensibility through the easy attachment of additional domain-specific ontologies. A typical example of a domain-specific ontology is the CIDOC/CRM - an ontology designed for information exchange in the cultural heritage and museum community. It's scope is the curatorial knowledge of museums and as such it is intended to cover contextual information such as the historical, geographical, cultural or theoretical background associated with artefacts held in museum collections, and which justifies their value and significance.

Many cultural organizations such as the Smithsonian are interested in developing metadata application profiles which combine the CIDOC/CRM with MPEG-7 (for content description) and MPEG-21 (for rights management) to describe and manage their multimedia museum collections. Hence the ability to harmonize the CIDOC/CRM with the ABC, MPEG-7 and MPEG-21 ontologies described above would be both a measure of the integrity of the ABC model and a very useful tool for many cultural institutions.

Through a series of three workshops which were held in 2001 and 2002, under the sponsorship of the DELOS Network of Excellence in Digital Libraries [32] and the Harmony project [12], the knowledge perspectives of the ABC and CIDOC/CRM ontologies were harmonized. A detailed description of this process is described in [33]. The result of this process was a harmonized ontology in which the top-level entities and properties of the CIDOC/CRM model were reconciled with the top-level entities and properties of the ABC model. Figure 11 illustrates the convergence of these two ontologies.
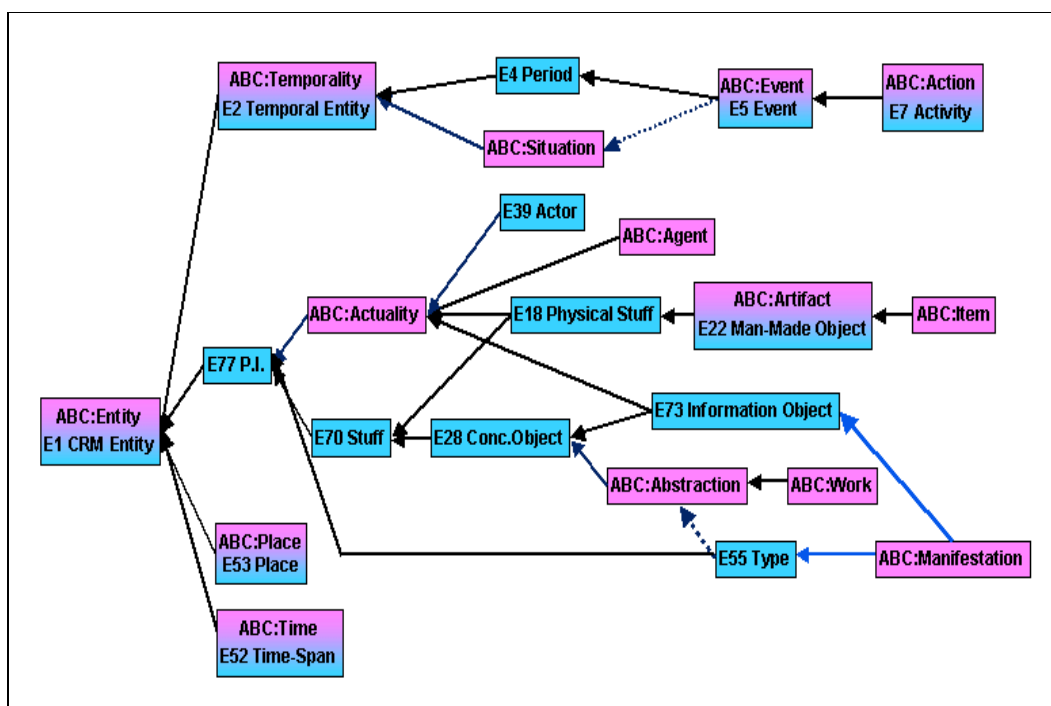
Figure 11 - The merged ABC and CIDOC/CRM class hierarchies

Once the top level classes have been harmonized it then becomes a trivial process to attach the museum-specific sub-class and sub-property hierarchies of the CIDOC/CRM to the top-level ABC classes.

5.3 Multimedia Aggregation

The majority of current search interfaces are still at the `hunter-gatherer' stage. The result of a typical search is a sequential list of URL's, referring to the HTML pages which match the metadata search field, displayed according to rank. The fact that there are semantic relations between the retrieved objects or that many more semantically related information objects exist, is ignored in the final presentation of results. But given machine-understandable rich semantic descriptions of multimedia resources, it should be possible to infer more complex semantic relationships between retrieved resources.

In a collaborative research project between DSTC and CWI [34], a system was developed which searches across repositories provided by the Open Archives Initiative (OAI) [35] data providers. By applying an iterative sequence of searches across the metadata, provided by the data providers, semantic relationships can be inferred between the mixed-media objects which are retrieved. Using predefined mapping rules, these semantic relationships are then mapped to spatial and temporal relationships between the objects. Using the Cuypers Multimedia Transformation Engine [36] developed by CWI, the spatial and temporal relationships are expressed within SMIL files which can be replayed as multimedia presentations. By using automated computer processing of metadata to organize and combine semantically-related objects within multimedia presentations, the system may be able to generate new knowledge by exposing previously unrecognized connections. In addition, the use of multilayered information-rich multimedia to present the results, enables faster and easier information browsing, analysis, interpretation and deduction by the end-user.

Figure 12 shows the high-level architecture of the system which was developed. Details are provided in a paper published at ECDL 2002 [37] and an online demonstration is available at [38] http://aries.ins.cwi.nl:8580/cocoon/cuypers/index.html
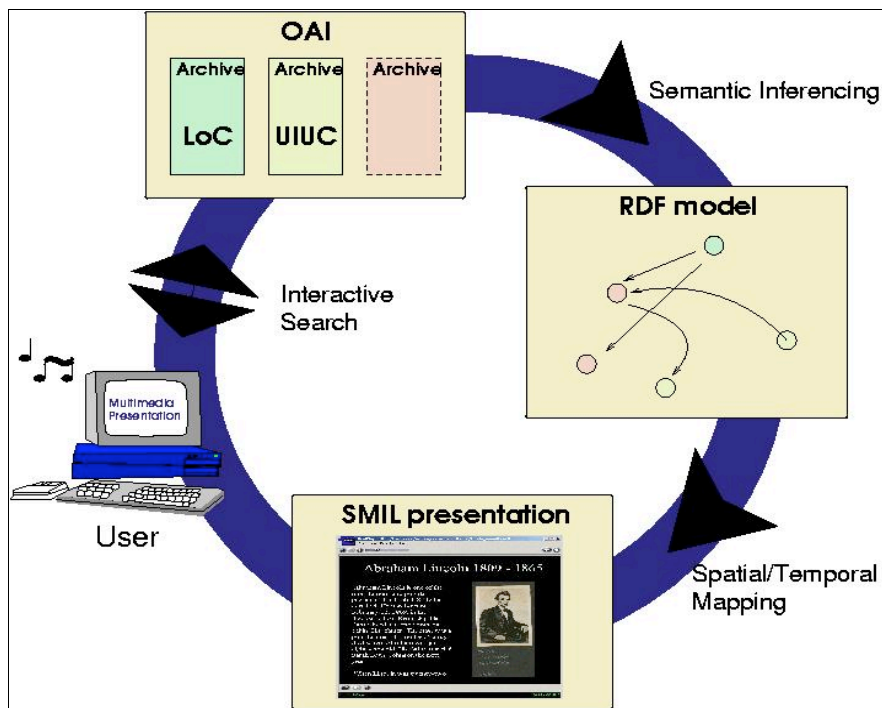

Figure 12 - Architecture of the Automatic Multimedia Presentation Generator

5.4 Conceptual Entity Recognition

Although significant progress has been made in recent years on automatic segmentation, scene-change detection, and the recognition and detection of low-level features for multimedia content, comparatively little progress has been made on machine-generation of semantic descriptions of audiovisual information. The representation of the semantics of MPEG-7 and MPEG-21 terms within machine-processable ontologies is expected to facilitate the future construction of efficient knowledge-based multimedia systems which are capable of automatically extracting and aggregating semantic information (objects, events, properties, relations) about the audiovisual data. For example, given the appropriate RDF inferencing rules, a complete distributed and well-structured ontology could conceivably enable the "subject" of an image or the "genre" of a video to be automatically deduced from a combination of low-level MPEG-7 visual or audiovisual descriptor values. Such extracted semantic metadata could then be used to automate the classification, summarisation, indexing, searching and efficient retrieval of multimedia content.

6. Conclusions

In this paper we have shown how the definition of a core, top-level but extensible, ontology which is separate and independent of any domain-specific metadata vocabulary is essential for: well-defined mappings between domain-specific metadata vocabularies; the integration of information from different domains; semantic interoperability between data services over networks; and the development of sophisticated knowledge mining services based on semantic inferencing.

By using the ABC event-aware model as the underlying framework and representing the ABC, MPEG-7 and MPEG-21 vocabularies as RDF Schema class and property hierarchies, we have been able to easily ascertain the intersections, differences and domain-specific aspects of each ontology. This has enabled us to determine the most appropriate attachment points on the ABC framework for MPEG-7 and MPEG-21 subclass and subproperty trees. In addition, the framework will continue to provide a basis for the attachment of further components of MPEG-21 as they are developed, or for incorporating non-MPEG, domain-specific semantic definitions from other communities e.g., the museum or medical communities.

Finally, by expressing the ABC, MPEG-7 and MPEG-21 ontologies in a machine-processable language (currently RDF Schema/DAML+OIL but moving to OWL), we have been able to generate an unambiguous machine-understandable formal representation of the semantics associated with multimedia description, management and delivery - paving the way for more effective resource discovery, transparent data exchange and automated integration, re-use, delivery and rights management of multimedia for content creators, distributors and consumers alike.

Acknowledgements

References

[1] ISO/IEC 15938-5 FDIS Information Technology - Multimedia Content Description Interface - Part 5: Multimedia Description Schemes, July 2001, Sydney

[2] ISO/IEC TR 21000-1:2001(E) Part 1: Vision, Technologies and Strategy, MPEG, Document: ISO/IEC JTC1/SC29/WG11 N3939 <http://www.cselt.it/mpeg/public/mpeg-21_pdtr.zip>

[3] XML Schema Part 0: Primer, W3C Recommendation, 2 May 2001, <http://www.w3.org/TR/xmlschema-0/>

[4] D. Brickley and R. V. Guha, "Resource Description Framework (RDF) Schema Specification," World Wide Web Consortium, W3C Candidate Recommendation CR-rdf-schema-20000327, March 27 2000. <http://www.w3.org/TR/rdf-schema/>

[5] DAML+OIL, March 2001. <http://www.daml.org/2001/03/daml+oil-index>

[6] Web Ontology Language (OWL) Guide Version 1.0, W3C Working Draft, 4 November 2002 <http://www.w3.org/TR/owl-guide/>

[7] C.Lagoze, J.Hunter, "The ABC Ontology and Model", Journal of Digital Information, Volume 2, Issue 2, November 2001, <http://jodi.ecs.soton.ac.uk/Articles/v02/i02/Lagoze/>

[8] IMS Learning Resource Metadata Specification Version 1.2.2

[9] Dublin Core Metadata Initiative, <http://dublincore.org>

[10] The Federal Geographic Data Committee Geospatial Metadata Standard

[11] CIDOC Conceptual Reference Model, http://cidoc.ics.forth.gr/

[12] The Harmony Project, http://metadata.net/harmony/

[13] ON9.3 Biomedical Core Ontology, http://saussure.irmkant.rm.cnr.it/onto/ON9.3-OL-HTML/index.html

[14]<indecs>2rdd, MPEG Document: ISO/IEC JTC1/SC29/WG11 W7610, 58th MPEG Meeting, Pattaya, December 2001;

[15] RDF Schema Representation of the ABC Ontology, November 2001, <http://metadata.net/harmony/ABC/ABC.rdfs>

[16] M. Gorman, The concise AACR2, 1988 revision. Chicago: American Library Association, 1989.

[17] Dublin Core Qualifiers, <http://purl.org/DC/documents/rec/dcmes-qualifiers-20000711.htm>

[18]"Functional Requirements for Bibliographic Records," International Federation of Library Associations and Institutions March 1998. <http://www.ifla.org/VII/s13/frbr/frbr.pdf>

[19] XML Schema Part 1: Structures, W3C Recommendation, 2 May 2001, <http://www.w3.org/TR/xmlschema-1/>

[20] XML Schema Part 2: Datatypes, W3C Recommendation, 2 May 2001, <http://www.w3.org/TR/xmlschema-2/>

[21] W3C Semantic Web Activity, <http://www.w3.org/2001/sw/>

[22] J. Hunter, "Adding Multimedia to the Semantic Web - Building an MPEG-7 Ontology", <http://archive.dstc.edu.au/RDU/staff/jane-hunter/swws.pdf>

[23] J. Hunter, "An RDF Schema/DAML+OIL Representation of MPEG-7 Semantics", MPEG, Document: ISO/IEC JTC1/SC29/WG11 W7807, December 2001, Pattaya

[24] MPEG-21 Requirements, "Call for Proposals for a Rights Data Dictionary and Rights Expression Language", MPEG Document: ISO/IEC JTC1/SC29/WG11 N4335, July 2001, Sydney

[25] B. Wragg et.al., "Report from Evaluation Meeting 'RDD/REL CfP'", MPEG Document: ISO/IEC JTC1/SC29/WG11 M7709, December 2001, Pattaya

[26] An RDF Schema/DAML+OIL representation of the MPEG-7 Ontology based on ABC, <http://metadata.net/harmony/MPEG7/MPEG7.rdfs>

[27] An RDF Schema/DAML+OIL representation of the MPEG-21/<indecs> Ontology based on ABC, <http://metadata.net/harmony/MPEG21/MPEG21.rdfs>

[28] W3C Web Ontology Working Group, <http://www.w3.org/2001/sw/WebOnt/>;

[29] CIMI Consortium for Interchange of Museum Information, <http://www.cimi.org/>;

[30] Images and Metadata Records Provided by CIMI Members, <http://archive.dstc.edu.au/CIMI/index.html>;

[31] XML Query Interface to CIMI/ABC Testbed, <http://sunspot.dstc.edu.au:9000/cocoon/xmlQry/index.html>;

[32] DELOS Network of Excellence for Digital Libraries, <http://delos-noe.iei.pi.cnr.it/>;

[33] M. Doerr, J. Hunter, C. Lagoze, "Towards a Core Ontology for Information Integration", submitted to JoDI, October 2002;

[34] CWI, Centrum voor Wiskunde en Informatica, <http://www.cwi.nl/>;

[35] The Open Archive Initiative, <http://www.openarchives.org/>;

[36] Lloyd Rutledge, Jim Davis, Jacco van Ossenbruggen, and Lynda Hardman. Inter-dimensional Hypermedia Communicative Devices for Rhetorical Structure. In Proceedings of the International Conference on Multimedia Modeling 2000 (MMM00), pages 89-105, Nagano, Japan, November 13-15, 2000.

[37]S.Little, J. Guerts, J. Hunter, "The Dynamic Generation of Intelligent Multimedia Presentations through Semantic Inferencing", ECDL 2002, Rome, September 2002 <http://archive.dstc.edu.au/maenad/ecdl2002/ecdl2002.html>;

[38]The Cuypers Multimedia Transformation Engine <http://aries.ins.cwi.nl:8580/cocoon/cuypers/index.html>;

Vitae

Jane Hunter is a Senior Research Fellow at the Distributed Systems Technology Centre, at the University of Queensland. Her research interests are multimedia metadata modelling and interoperability between metadata standards across domains and media types. She was chair of the MPEG-7 Semantic Interoperability Adhoc Group, editor of the MPEG-7 Description Definition Language ISO/IEC 15838-2 and is the liaison between MPEG and the W3C.