The art-E-fact ontology: a possible contribution to CIDOC CRM

C. Lamsfus*, G. Karagiannis**, S. Sotiropoulou**, H. Eskudero*, G. Marcos*, M.T. Linaza* and S. Daniilia**

*Asociación Centro de Tecnologías de Interacción Visual y Comunicaciones, VICOMTech, Spain
**Ormylia Diagnosis Centre, Greece

Abstract

This paper describes a domain ontology which has been designed within the European art-E-fact IST project to represent the associated knowledge and to enable the description, exchange and sharing of multimedia added-value content for the creation of artistic expressions both within and between cultural institutions.
This paper proposes an approach that extends the documentation domain-specific aspects of CIDOC-CRM with a domain ontology for the generation of added-value content in the cultural heritage area. The result could be an extensible model which could lead to a common search interface and the open exchange, sharing and integration of heterogeneous multimedia resources distributed across cultural institutions.

1. Introduction

Creating art is the genesis of an original, impulse expression of feeling, thoughts, passions, behaviour, etc. Expression is an output of what creators obtain in their internal worlds, through their cultural background and environment, as well as through their technical skills. The huge amount of experiences and the stochastic way of assimilating and mixing them is the kernel of the final expresions that rise out.

So technically speaking, authors, artists or content generators should be aware of this rich internal world which is provided to them through the desciption- metalevel ontology. The technical skills of the author aided by tools, which will retrieve content from distributed databases using the metalevel ontology, will rise out in an optimal way following the memory of creating art.

Interoperability between these databases has to be provided on both a technical and an informational level. Problems that might arise owing to heterogeneity of the data are already well-known within the distributed database systems community: structural heterogeneity and semantic heterogeneity.

In order to achieve the latter, the meaning of the information that is interchanged has to be understood across the systems. Semantic conflicts occur whenever two contexts do not use the same interpretation of the information. The use of ontologies for the explication of implicit and hidden knowledge is a possible approach to overcome the problem of semantic heterogeneity.

For the semantic technologies to succeed, there is a need not only of modelling languages, but also methodologies for extracting and defining the knowledge to be represented. One of the major hurdles facing the cultural content creation is the lack of suitable ontologies.

Section 2 of the paper presents a brief description of the CIDOC CRM. Section 3 describes in more detail different aspects of the art-E-fact ontology, including a brief description of the project, the scope of the ontology, the description of the metadata and teh applied form. Section 4 presents a comparison between both domain ontologies and finally, Section 5 provides some conclusions.

2. The CIDOC Conceptual Reference Model (CRM)

The CIDOC CRM is a formal ontology intended to facilitate the integration, mediation and interchange of heterogeneous cultural heritage information.

The primary role of the CRM is to enable information exchange and integration between heterogeneous sources of cultural heritage information [Doe03]. It aims at providing the semantic definitions and clarifications needed to trans-

form disparate, localised information sources into a coherent global resource within a larger institution, in intranets or on the Internet.

More concretely, it defines and is restricted to the underlying sematics of database schema and document structures used in cultural heritage and museum documentation in terms of a formal ontology. It does not aim at proposing what cultural institutions should document. Rather it explains the logic of what they actually currently document, and thereby enables semantic interoperability.

Among the specific functionalities, the CRM aims at supporting, for example, the following ones:

- Serve as a common language for domain experts and IT developers to formulate requirements and to agree on system funcionalities with respect to the correct handling of cultural contents;
- Support the implementation of automatic data transformation algorithms from local to global data structures without loss of meaning. This is useful for data exchange or data information integration; as well as
- Support associative queries agains integrated resources by providing a global model of the basic classes and their associations to formulate such queries.

2.1. Scope of the CIDOC CRM

The overall scope of the CIDOC CRM can be summarised in simple terms as the curated knowledge of museums [CDG*03]. The Intended Scope of the CRM may be defined as all information required for the exchange and integration of heterogeneous scientific documentation of museum collections. This definition requires further elaboration.

- The term "scientific documentation" is intended to convey the requirement that the descriptive information handled by the CRM should be sufficient for serious academic research. The CRM is intended to provide the level of detail and precision expected and required by museum professionals and researchers in the field.
- The term "museum collections" includes collections, sites and monuments relating to fields such as social history, ethnography, archaeology, fine and applied arts, natural history, history of sciences and technology.
- The documentation of collections includes the detailed description of individual items within collections, group of items and collections as a whole.
- Information required solely for the administration and management of cultural institutions falls outside the Intended Scope of the CRM.

2.2. Applied Form

The CRM is a domain ontology in the sense used in computer science. It has been expressed as an object-oriented semantic model, so that it can be converted to machine-readable formats such as RDF Schema, OWL, DAML+OIL [DHL03]. It is composed of 81 classes and 132 unique properties. It does not attempt to articulate the inheritance of properties by subclasses throughout the class hierarchy.

3. The art-E-fact ontology

3.1. Description of the project

The aim of the art-E-fact project (IST-2001-37924) is to create a generic platform for Interactive Storytelling in Mixed Reality that allows artists to create artistic expressions in an original way within a cultural context between the virtual and the physical reality.

With the means of Interactive Storytelling and Mixed Reality of virtual autonomous characters, multimedia, physical props and devices, and multimodal human-oriented interactions for artistic expression are enabled [Iur04]. The target platform is both a new medium for the communication of informational content, and a new form of act. The main objectives of the art-E-fact project are:

- to develop a generic platform for Interactive Storytelling in Mixed Reality that allows artists to create artistic expressions in an original way, within a cultural context between the virtual ("new") and the physical ("traditional") reality,
- to use the platform to actually build a compelling Mixed Reality installation that facilitates the access to a knowledge base of inspirational material of art history- reflecting the way humans created art since at least 4000 years,
- to involve artists and the analysis of artistic methods, on from the beginning of the project through all project phases, as well as
- to create a showcase within an interdisciplinary team that can be used for the evaluation of artistic methods, as well as for the diffusion and exploitation of the results, leading to more accessible tools for artistic expression in the future.

Artists can create a Mixed Reality exhibit by using the generic system to shape a specific instance of expression (Figure 1). They make choices of specific interaction devices and physical props to be used for anthropomorphic interactions, as well as corresponding interaction metaphors; they define dialogues with a degree of autonomy and behaviour of virtual characters, and they create multimedia elements to be accessed during run-time.

art-E-fact aims at the provision of Mixed Reality technology addressing two directions: firstly, to provide a generic platform for artistic expression to enable interactive exploration of artworks, and secondly, to ease the task of artwork creation by providing a standard VRML compatible VR framework enhanced with interaction and sensor features.
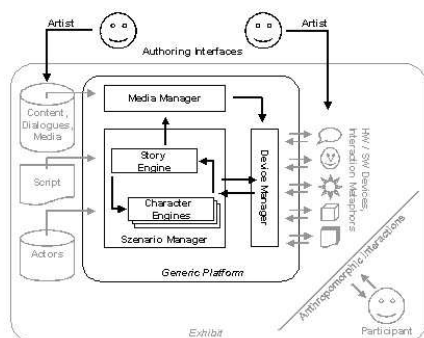
Figure 1: *Generic platform (black, including authoring interfaces) to be used by the artists who create an exhibit (grey), which is an interactive storytelling application in Mixed Reality.*

The Mixed Reality generic platform for Interactive Storytelling serves as an experimental platform allowing authors with artistic or humanistic backgrounds to make design decisions that go beyond the state-of-the-art creation systems of digital media. In summary, it is possible for artists to include anthropomorphic interactions such as gestures, body poses into their design of Mixed Realities, and to direct lifelike avatars in order to act.

3.2. Scope of the art-E-fact ontology

The art-E-fact ontology is part of the outcome of the art-E-fact project. Since the target of the art-E-fact project is to create stories about artworks and thus create art, we have to be taught by the experience that was gained the last 4000 years of civilization.

As noted above, this is not just a conception of the experts performing the scientific diagnosis, but it is also a tool for artists, authors and content generators. Artists using the ontology have to create stories or experiences concerning one or more selected artworks, including its main features, technical data, historical context, etc. All this information will be included within the Cultural Content concept. The domain metalevel ontology conception will lead them to assimilate the internal world of the creator of an artwork, and create and tell stories.

The following is an enumeration of the merits identified for the art-E-fact ontology.

- A common vocabulary. The description of the target world needs a vocabulary agreed among people involved.
- Data structure. In this sense, the art-E-fact ontology implements a conceptual structure appropriate for information description and exchange.
- Access to data. The ontology facilitates browsing, navigating and retrieving content using semantic concepts.

- Reusability. The explicit description about assumptions of the domain contributes to making the system understandable and transparent, and hence its reusability increases.
- Interoperability. The goal of the ontology is to build coherent access to artists, authors and content generators so that they can access distributed data managed by different organisations.
- Semantic web technology. The ontology will allow other users and systems to access the art-E-fact system using semantic web technology, providing the system that is accessing has permissions.

Scoping the ontology has been mainly based on two brainstorming sessions with the artists and the content providers. Having these brainstorming sessions allowed us to produce most of the potentially relevant terms and phrases. At this stage, the terms alone represented the concept, thus concealing significant ambiguities and differences of opinion.

A clear issue that arose during these sessions was the terminology differences among different art styles, between the Greek traditional iconography and the traditional European painting schools. The concepts listed during the brainstorming sessions were grouped in areas of work corresponding to naturally arising sub-groups. Most of the important concepts and many terms were identified. The main work of building the ontology was then to produce definitions. During scoping, most of the important concepts and many terms have been identified.

3.3. Description of the conception

The scientific diagnosis and documentation of artworks provide artists, authors or content generators with a rich knowledge background with plenty of multidimensional data and metadata. There is a special relation among the metadata, which reveals all the knowledge concerning the artwork obtained from the diagnosis procedure.

The artwork is related to five levels of knowledge, enriched with a set of metadata or descriptors of the data of the diagnosis. All these levels of knowledge or "thematic entities" in the ontology conception are supported by the scientific diagnosis results and the related documentation.

- The entity "Work identification" consists of general historical data, identifying aspects such as subject, title, category, type, dimensions, current location, context, ownership or creator of the artwork.
- The entity "Description" consists of information concerning the descriptive details of the theme and forms of representation, providing a better understanding of the context, such as representation, persons, background, decorative elements, inscriptions or sceneries.
- The entity "Aesthetic appearance" concerns mainly with plastic elements, which provide the appreciation of the style/aesthetic appearance of the artwork, such as the

style, manner, composition set-up, colour, drawing style or texture.

- The entity "Technical" includes technical information both revealing the techniques and the materials used in the creation of the artwork, such as support, preparatory layers, underdrawings, painting materials, varnishes or stratigraphy, and also concerning exams of the condition, such as diagnosis or conservation treatments history.
- The entity "Interpretation" is provided compared or associated with analogous or totally unlike artworks, such as thematic relationships, persons, symbols, styles or techniques.

These main entities and their metadata are supported, documented and provided by the scientific diagnosis, which has been applied to the artworks.

3.4. Applied form

Among the three possible alternatives to define the classes to build the art-E-fact ontology (top down, bottom up and a combination of both development process), a combined development process has been used. The most representative concepts have been defined first and then they have been specified appropriately in order to get a representation of the knowledge stored in the databases. The art-E-fact domain ontology is composed of 84 classes and 173 properties and has been implemented in a RDF Schema.

4. Comparison between the CIDOC CRM and the art-E-fact ontology

The CIDOC CRM and the art-E-fact ontologies reflect a commitment to the expression of common concepts underlying the data structures used by their users. The art-E-fact model, driven by artists and content generators requirements, was motivated by the need to describe added-value content for the creation of stories. The CIDOC CRM model, motivated by cultural artifacts, documentation experts and museum requirements, focuses on documentation processes among cultural institutions.

- The intended scope of the CIDOC CRM has been defined as all the information required for the scientific documentation of cultural heritage collections, with a view to enabling wide area information exchange and integration of heterogeneous sources. The main objective of the art-e-fact ontology is not devoted to documentation, but to content description and comprehension. At this point, there should be clarified the distinction between information for documentation and the generation of content for diffusion of cultural heritage.
- The term cultural heritage collections is intended to cover all types of material collected and displayed by museums and related institutions, as defined by ICOM. This includes collections, sites and monuments relating to natural history, ethnography, archaeology, historic monuments, as

well as collections of fine and applied arts. The presented ontology is also valid for interpretation centres and humanistic research institutions, which may have access to data and are not included among the ICOM concept.

- The scope of the CIDOC CRM is the curated knowledge of museums, while the scope of the art-E-fact project is the content generation by the artists.
- The CIDOC CRM is specifically intended to cover contextual information: the historical, geographical and theoretical background in which individual items are placed and which gives them much of their significance and value. Meanwhile, the art-E-fact ontology takes into account different levels of knowledge in order to provide rich content to build interactive amazing stories.

Therefore, the main difference between both ontologies is the application domain. There is no incompatibility between both models. Moreover, it should be possible to consider the art-E-fact ontology as an extension in the area of content description.

5. Conclusions

As it is encouraged by the CIDOC CRM, we believe the art-E-fact ontology could be an extension for the needs of the content generation community and its applications.

Compatibility of extensions with the CRM means that data structured according to an extension must also remain valid as a CRM instance. In practical terms, this implies query containment: any queries based on CRM concepts should retrieve a result set that is correct according to the CRM's semantics.

A sufficient condition for the compatibility of an extension with the CRM is that CRM classes subsume all classes of the extension, and all properties of the extension are either subsumed by CRM properties, or are part of a path for which a CRM property is a shorcut.

In order to evaluate this compatibility, and as it has been done before for other ontologies, we are planning to follow a formal, analytical, deterministic method to compare and converge ontologies based on the OntoClean approach. Furthermore, in addition to comparing the concept definitions of each model, we propose bilateral meeting with the CIDOC CRM expert group in order to investigate how well the concepts cope in order to improve both models.

6. Acknowledgements

References

[CDG*03] CROFTS N., DÖRR M., GILL T., STEAD S., STIFF M.: *Definition of the CIDOC Conceptual Reference Model*. ICOM/CIDOC CRM Special Interest Group, 2003. 2

[DHL03] DOERR M., HUNTER J., LAGOZE C.: Towards a core ontology for information integration. *Journal of Digital Information 4(1)* (2003). 2

[Doe03] DOERR M.: The cidoc conceptual reference module: An ontological approach to semantic interoperability of metadata. *AI Magazine 24(3)* (2003), 75–92. 1

[Iur04] IURGEL I.: From another point of view: art-e-fact. In *Proc. TIDSE '04* (2004), vol. 1, pp. 26–35. 2