Towards Semantic Based Information Exchange and Integration Standards: the art-E-fact ontology as an extension to the CIDOC CRM (ISO/CD 21127) Standard

Carlos Lamsfus
VICOMTech
San Sebastián, Spain, E-20009
clamsfus@vicomtech.es

María Teresa Linaza
VICOMTech
San Sebastián, Spain, E-20009
mtlinaza@vicomtech.es

Tim Smithers
VICOMTech
San Sebastián, Spain, E-20009
tsmithers@vicomtech.es

ABSTRACT

Ontologies have been established as effective and efficient means of knowledge sharing and are being widely used to conceptually model domains of knowledge. With the growing use of ontologies in various domains of interest, the problem of overlapping knowledge in a common domain becomes critical. In this context, much work has already been done developing semi-automated applications that enable the merging, mapping or alignment of ontologies.

On the other hand, a big effort is being currently developed by many communities (e.g. eLearning, telemedicine, cultural heritage) in order to standardize their contents and data models facilitating the integration and exchange of content coming from heterogeneous data sources.

This paper presents the work that is currently being carried out in two different directions: first, to align two existing ontologies (the art-E-fact project domain ontology and the CRM ontology); second, to pursue a culture of re-using existing ontologies and content, generalizing the alignment method so that it becomes standard and applicable in other domains so that existing knowledge can be re-used and shared easily.

Keywords
Ontology, alignment, cultural heritage, standards, re-use.

INTRODUCTION
A unified representation for Web data and resources is needed in today's large scale Internet data management systems. This unification of standards will allow machines to meaningfully process the available information and to exchange and integrate data coming from distributed databases and information management systems. This has been occurring, e.g. in the context of eLearning with the development of the SCORM (http://www.adl.net) and AICC (http://www.aicc.org) standards, or in the context of telemedicine applications with the development of standard data transport protocols such as HL7 and ISO/IEEE/CEN 11073, among others.

In the area of cultural heritage, there have also been some initiatives to enable distrubuted data exchange and integration. Interoperability between databases has to be provided on both technical and informational (semantic) levels. Problems that may arise due to heterogeneity of the data are already well-known within the distributed database systems community: structural heterogeneity and semantic heterogeneity.

Problems related to structural heterogeneity of distributed databases have already been solved by the database management systems community. Furthermore, in order to achieve semantic interoperability, i.e. achieve communication between two agents that work in overlapping domains, the meaning of the information that is being interchanged has to be understood across both systems. The use of ontologies for the description of implicit hidden knowledge is a possible approach to overcome the problem of semantic heterogeneity.

This paper presents the work that is currently being carried out in two important lines of research:

- First, research on information integration and exchange standards in the context of cultural heritage: suitability of alignment of the art-E-fact domain ontology into the CIDOC CRM ontology;
- Second, research on ontology merging and alignment, in order to align the art-E-fact ontology into the CIDOC CRM ontology.

This paper is organized as follows. Section 2 presents previous work upon which this article is based briefly describing the CIDOC Conceptual Reference Model and the art-E-fact project and domain ontology. Section 3 is a comparative analysis of both the art-E-fact and CRM ontologies and summarizes the differences between them. Section 4 describes the working method that is currently under development. Finally, Section 5 gives some conclusions.

PREVIOUS WORK

The CIDOC Conceptual Reference Model (CRM)

The CIDOC CRM, a core ontology explaining the extended meaning of data structures from humanities and cultural heritage, including history of science, is the outcome of a long-term disciplined knowledge engineering activity, which excels in its ontological commitment, i.e. acceptance of its constructs by domain experts.

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 transform disparate, localised information sources into a coherent global resource within a larger institution, in intranets or in the Internet. More concretely, it defines and it is restricted to the underlying semantics of database schema and document structures used in cultural heritage and museum documentation in terms of a formal ontology.

The following are some of the most important functionalities of the CRM:

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

The success of the CIDOC CRM lies in the fact that the explanation of common meaning can be done by a very small set of primitive concepts and relationships, in contrast to the data structures that suggest to the user what to say about an object. The relations in data structures that connect items directly by highly specific, diverse kinds of relationship can frequently be expressed by data paths composed of a few fundamental relationships defined in the core ontology.

The CIDOC CRM has become the most promising core element for realizing semantic interoperability in Archives, Libraries and Museums, by its capability to link the intellectual structure of highly diverse sources and products of scientific and scholarly discourse with the elements formally handled by information systems. The CIDOC CRM is currently being elaborated by the International Standards Organization as Committee Draft ISO/DIS21127 and the CIDOC CRM Special Interest Group (SIG) to become an ISO standard.

The overall scope of the CIDOC CRM can be summarised in simple terms as the curated knowledge of museums [CDG03]. 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.

The art-E-fact project

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

In other words, art-E-fact's general purpose was to make art accessible in a different way from traditional methods. For example, a piece of art could be introduced, explained and even discussed by two virtual characters in the form of a story. They set up a conversation first with each other and then with the visitor or participant, who will feel involved in the virtual world generated by the platform.

Furthermore, intuitive and easy to use Mixed Reality based interaction techniques enable the visitor to explore the art work in depth by using physical devices that will be recognized by a gesture recognition system. With art-E-fact, the visitors of the museum will deepen their understanding of the complex issues surrounding the history, techniques, and social circumstances behind the individual artworks.

So, if we want to make art accessible in the form of stories to visitors of museum exhibitions we have to provide artists, users and content generators in general with a tool (the art-E-fact Authoring Tool) that allows them to create this kind of art with the features we are describing.

Creation of art is the genesis of an original expression of feelings, thoughts, passions 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 expressions that arise.

Therefore, if art-E-fact's target is "to tell" stories about (existing) artworks the author or content generator (using the art-E-fact Authoring-Tool) should be aware of this rich internal world which is provided to him/her through the art-E-fact description-metalevel ontology. The author's technical skills aided by the Authoring Tool, which retrieves the required information from the ArtWorks Database (AWDB) using the metalevel ontology, arise in an optimal way following the memory of creating art.

Description of the conception of the art-E-fact ontology

The art-E-fact domain ontology is composed of 84 classes and 173 properties and has been implemented in RDF Schema. It represents the artworks and its relational data stored in the AWDB and it is referred 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, people, background, decorative elements, inscriptions or sceneries;
- The entity "Aesthetic appearance" is concerned 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 condi-

tion, 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 that has been applied to the artworks.

COMPARISON OF THE CRM AND art-E-fact ONTOLOGIES

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

- 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.
- In the context of the CRM, 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 art-E-fact 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. On the other hand, 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, we believe it should be possible to consider the art-E-fact ontology as an extension in the area of content description and generation of the CRM.

ALIGNMENT OF THE ART-E-FACT AND CRM ONTOLOGIES

As proved before the art-E-fact domain ontology is (conceptually) complementary to the CRM and could therefore be considered as an extension of the standard model. We are proposing to incorporate the art-E-fact domain ontology into the CRM as part of the standard to study and research different ontology alignment methods, developing a new one if necessary.

We have previously presented the different levels of knowledge or thematic entities the art-E-fact domain ontology covers as well as the scope of the CIDOC CRM. The CRM covers ("only") information about cultural heritage collections which would be equivalent to the first thematic entity (Work Identification) defined in the art-E-fact ontology. Therefore, we are going to link (unite) common concepts of both ontologies, allowing this way the CRM ontology to access to the rest of knowledge levels covered by art-E-fact but not contemplated in it. Figure 1 (conceptually) shows the work that we are currently carrying out.
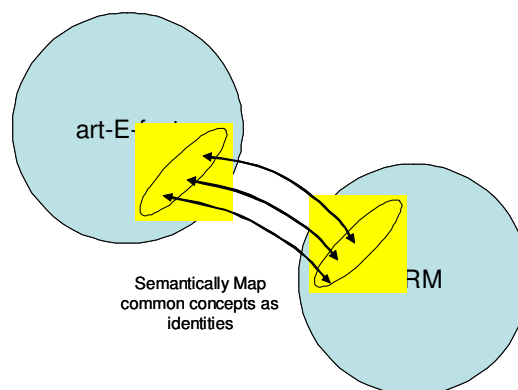

Figure 1: alignment of the art-E-fact and CRM ontologies

For the alignment of both ontologies we are going to use a rule-based methodology by the means of the emerging Semantic Web rule languages. Reasoning languages for the Web are an emerging technology that does not exist today. This technology will soon represent an essential breakthrough for Web systems and applications. One possible rule-based ontology language that we can use in this process is the Web Rule Language (WRL) for the Semantic Web. This language is located in the Semantic Web stack next to the Description Logic based Ontology Language (OWL).

The ontology vocabulary can be specified using WRL or OWL, or using their common semantic subset, denoted by the WRL-Core subset of WRL and the OWL-DP subset of OWL [Grosof *et al.*, 2003]. With common semantic subset we mean in this context that every WRL-Core has a corresponding OWL-DP ontology and vice versa, where both ontologies entail exactly the same set of ground facts.

So, the alignment work of both ontologies can roughly be summarized in the following tasks:

- Since the alignment of both ontologies is going to be carried out using the WRL language, we have to use the OWL DL version of the CRM and art-E-fact ontologies;
- The art-E-fact ontology was built using the RDF(s) language, therefore we have to represent this ontology using OWL DL;
- We have to standardize the art-E-fact ontology. Here, the word standardize stands for giving to the art-E-fact domain ontology's classes, slots etc. the same name they would have in the CRM ontology;
- Then, we have to identify the common concepts (classes) in both the art-E-fact and CRM domain ontologies and link them. To do so, we are going to follow a semantic rule-based process with WRL as rule language. This is the reason because we have to use the OWL (DL) version of both ontologies;
- In this linking process we have to be specially careful with the possible subclasses and slots each of the linked classes has, so that no information at all is lost in the alignment process;

We have now presented a very specific example of ontology alignment. In order for this process to be generic enough, we should also work on the standardization of the linking process so that it can also be used in the alignment of other ontologies.

Much previous work related to ontology mapping and merging has been done. To generate the bridging axioms, we must first find out the correspondence between the concepts of the two ontologies, which is the target of ontology mapping. Lots of systems have been implemented to map ontologies, e.g. CUPID, GLUE, Chimaera, PROMT and many others.

SUMMARY AND CONCLUSIONS

Technology, in the wider sense of its meaning, is tending to standards in order to enable and ease information exchange across different and usually distributed information management systems, should they be mobile devices or desktop computers. Thus, standards are not just a need but a must.

In this paper we have presented the CIDOC CRM and the art-E-fact domain ontologies. Then, we have compared them and, finally we have justified why we built a new ontology from scratch after analyzing their differences. Now, we are trying to align the art-E-fact ontology with the CRM standard in our commitment with standards and the re-use of previously developed work. In order to achieve this goal, we will study different methodologies, tools and ways of doing it and we will apply the most suitable one.

This is, however, a very concrete example of aligning ontologies. Therefore, another work that is under development is the standardization of the mapping process so that this alignment process can be carried out in a semi-automated way for other kinds of ontologies.

Ontology alignment, merging or mapping processes are very tedious and time consuming. There are various (semi-automated) tools that have already been used in some previous initiatives, however at some point of the alignment process there has to be some kind of manual intervention. We shall very carefully study their functionality and see whether they are finally suitable or if we should develop a new aligning tool.

Standards are becoming a very important issue within the new Information and Communication Technologies (ICT) context. With this research project, we would like to contribute to the extension of the most important information and data exchange format there is in the area of cultural heritage, i.e. the CIDOC CRM, giving the possibility to general cultural heritage institutions to exchange and integrate added value data related to artworks. Moreover, some research is going to be done in the alignment of ontologies and we expect to contribute with methodologies, e.g. standard mapping methodologies.

ACKNOWLEDGEMENTS

REFERENCES

[1] Advanced Distributed Learning, http//www.adlnet.org

[2] Aviation Industrie CBTI Committee, http://www.aicc.org

[3] Crofts, N.; Dörr M.; Gill T.; Stead S.; Stiff M.: *"Definition of the CIDOC Conceptual Reference Model"*. ICOM/CIDOC CRM Special Interest Group, 2003.

[4] Dörr M.; Hunter J.; Lagoze C.: "Towards a core ontology for information integration. Journal of Digital Information 4(1)" (2003).

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

[6] Dou, D; McDermott; Qi, P, "Ontology translation by Ontology Merging and Automated Reasoning"

[7] Elvimaria Terzi, Athna Vakali, Mohand-Saïd Hacid, *"Knowledge Representaioin, ontologies and the Semantic Web"*

[8] Grosof, B. N.; Horrocks, I.; Volz, R.;, and Decker, S. *Description logic programs: Combining logic programs with description logic.* In Proc. of the Twelfth International World Wide Web Conference (WWW 2003), pages 48-57. ACM, 2003.

[9] Guarino, N; Welty, C. *"Evaluation Ontological decisions with OntoClean"*. Communications of the ACM, Vol 45, Nr. 2, pp. 61-66. February, 2002.

[10] Iurgel I.: *"From another point of view: art-E-fact"*. In Proc. TIDSE '04 (2004), vol. 1, pp. 26–35.

[11] Noy N.F.; Musen M.A., *"SMART: Automated Support to Ontology Merging and Alignment"*, Twelth Workshop on Knowledge Acquisition, Modeling, and Management, 1999. Banff, Canada.

[12] Noy N.F.; Klein M; *"Ontology evolution: not the same as Schema Evolution"*, July 2004

[13] REWERSE, Reasoning on the Web with Rules and Semantics, http://rewerse.net/

[14] The CIDOC Conceptual Reference Model (CRM), http://cidoc.ics.forth.gr

[15] The Web Rule Language, http://www.wsmo.org/wsml/wrl/wrl.html#wsml