INTRODUCTION

Over the past decade there has been a growing recognition of the need to address the fragility of the digital information that is collected in artistic activities. This need has been strengthened by the fact that digital information has been rapidly gaining ground within the performing arts community, due to significant breakthroughs in human-computer interaction technologies, such as human motion capture and analysis.

Preserving digital information aims at ensuring the intelligibility of this information at any given time in the near or distant future. Digital preservation has to address changes that inevitably occur in hardware or software, in the organisational or legal environment, as well as in the designated community, i.e. the people that will use the preserved information. In order to be preserved against these changes, digital information has to be enriched with metadata, usually referred to as Representation Information, which can be used for the interpretation of information. In addition, the user community for whom the information is being preserved must also be taken into account. Representation Information needs therefore to be connected to the Knowledge Base of the designated community. Ontologies offer the means for organizing and representing the semantics of this knowledge base.

The European IST project CASPAR (http://www.casparpreserves.eu/) aims to build a pioneering framework to support the end-to-end preservation lifecycle of scientific, artistic and cultural information. The focus of CASPAR is specifically on preserving knowledge for future archive intelligibility and information system/services interoperability. The CASPAR framework is based on the full use of the OAIS (Open Archival Information System) Reference Model [1], which provides a consistent set of concepts, terminology and framework for the development of archival information systems and related standards.

Interactive Multimedia Performance (IMP) preservation comprises part of the Contemporary Arts testbed of CASPAR, dealing with preservation of artistic contents. In particular, the IMPs produced by the MvM (Music via Motion) system [2] have been considered for preservation. For this purpose, we have adopted an ontology-driven approach [3, 4] that reuses and extends existing standards, such as the CIDOC Conceptual Reference Model (CIDOC-CRM) [5, 6].

This paper introduces a layering architecture for preservation ontologies, based on the Heraclitus II framework [7]. A set of extensions for the CIDOC-CRM ontology are also proposed, in order to describe concepts of the IMP domain. The remainder of this paper is organized as follows: The ontology layering architecture, as well as its construction and integration are discussed in sections 2 and 3. Section 4 presents the extended CIDOC-CRM ontology for the IMP domain. Finally, the paper is concluded in section 6.

2. ONTOLOGY LAYERING ARCHITECTURE

The Heraclitus II framework [7] considers ontologies as a semantically rich knowledge base for information management and proposes ways for the management and evolution of this knowledge base. The Heraclitus II ontology layering architecture comprises a collaborative scheme for ontology layering and integration so that responsibilities are distributed among ontology authors.

Figure 1 shows the ontology layering architecture of Heraclitus II that has been modified to accommodate the preservation ontologies of the MvM case study. The lower layers represent more generic and all-purpose ontologies, while the upper layers are customized for certain uses. When traversing the layers from bottom to top, each layer reuses and extends the previous ones. In addition, whenever a layer extends the ones below it (e.g.
with the insertion of new concepts), these extensions are propagated to the lower layers. Each layer is maintained by a different group of ontology authors, depending on the expertise that each layer requires. The integration of the ontology pyramid layers is achieved with the use of ontology mapping between ontologies belonging to the same layer (intra-layer), or different ones (inter-layer).

The Lexical Ontology layer contains domain-independent ontologies of a purely lexicographical nature. This layer handles lexicographical issues, such as multilingualism. An example of such an ontology is the widely adopted WordNet [8].

Modelling of a certain domain is the main characteristic of the Domain Ontology layer. In the context of the MvM case study, this layer describes the complex relationships that exist among different components of a performance. The extended CIDOC-CRM ontology presented in section 4 of this paper is used in this layer to describe the IMP domain.

The Data Ontology layer describes the data that we want to preserve, namely the fields in which it can be decomposed, their structure (simple or composite), their type (integer, real, enumerated, array, record, list), their coding (ASCII, binary), their location (rank, length), etc. Examples for this layer can be constructed with the use of the EAST language [9] for the syntactic description of data and DEDSL (Data Entity Dictionary Specification Language) [10] for their semantic description.

Finally, on top of the pyramid lies the Application Ontology layer, containing software development ontologies that represent the internal organization of the software applications that have produced the preserved data, e.g. Max/MSP (www.cycling74.com). The application ontologies describe the relations between components of a software tool, e.g. in the form of UML diagrams. This type of ontologies allows for the interconnection between software structures and ontological data. More specifically, if certain ontological data are mapped to programming data structures, this can facilitate the process of ontology-driven software development [11].

3. ONTOLOGY PYRAMID CONSTRUCTION & INTEGRATION

In order to apply the Heraclitus II framework in the MvM case study, we have taken the following steps. First, we constructed each ontology layer according to the requirements of the case study under consideration. Following ontology construction, the resulting layers have been mapped to each other in order to be integrated into the Heraclitus II ontology pyramid.

Ontology construction is semi-automatic, with the human author being assisted by clustering or classification algorithms [12]. For the construction of each ontology layer, the designated ontology author group needs to be involved. More specifically, these groups are the shown in Table 1.

<table>
<thead>
<tr>
<th>Ontology layer</th>
<th>Author group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical</td>
<td>Lexicographers</td>
</tr>
<tr>
<td>Domain</td>
<td>Domain experts</td>
</tr>
<tr>
<td>Data</td>
<td>Data curators</td>
</tr>
<tr>
<td>Application</td>
<td>Software developers</td>
</tr>
<tr>
<td>Inter-layer ontology mappings</td>
<td>Author groups of corresponding layers</td>
</tr>
<tr>
<td>Intra-layer ontology mappings</td>
<td>Author group of corresponding layer</td>
</tr>
</tbody>
</table>

Table 1. Assignment of ontology author groups to ontology layers

Starting from the bottom of the pyramid, lexicographers have the knowledge on language structures that is required in this level. For the Domain layer, a more diverse group is suitable. These are professionals on a certain domain, e.g. in our case.

![Figure 1](image-url). The Heraclitus II ontology pyramid for the MvM case study
performers, production managers, system managers etc. The designated ontology authors for this layer are the data curators, namely the people whose duty is to keep the archive usable over time. Finally, software developers are responsible for the Application layer, as they have to adjust the application ontologies according to their software modelling.

With regard to the ontology mappings, depending on their scope, their maintenance is a joint effort of one or two author groups. For example, an inter-layer ontology mapping between the lexical and the domain layer is maintained through the collaboration of the authors of these two layers, namely lexicographers and domain experts. Intra-layer ontology mappings are assigned to the author group responsible for the corresponding layer.

The integration of the ontology pyramid layers into one manageable scheme is achieved via ontology mapping. This is performed either manually by the ontology authors, or semi-automatically with the use of various ontology mapping techniques [13-18]. The ontology mappings of Heraclitus II are comprised of relations between ontology objects belonging to the same or different layers. These relations can be either structural, namely referring to the structure of the mapped ontologies, e.g. via is-a relations, or semantic when mapping two ontology objects via a semantic relation, such as an employer-employee relation. OWL Full [19] offers a variety of constructs for representing structural ontology mappings, including owl:subClassOf, owl:sameAs, owl:inverseOf, owl:equivalentClass, and owl:equivalentProperty.

4. THE EXTENDED CIDOC-CRM ONTOLOGY

The CIDOC Conceptual Reference Model (CRM) has been proposed as a standard ontology for enabling interoperability amongst digital libraries [5, 6, 20]. CIDOC-CRM was originally designed to describe cultural heritage collections in museum archives. The meta-schema of CIDOC-CRM is illustrated in Figure 2. CIDOC-CRM’s conceptualisation of the past is centred on Temporal Entities (e.g. events). People (Actors) and objects (Conceptual Objects and Physical Objects) involved, time (Time-Spans) and Places are documented via their relationships with the Temporal Entities.

In order to describe the IMP domain in the Domain Ontology layer of the Heraclitus II ontology pyramid, we propose a set of extensions to CIDOC-CRM. These extensions have the following objectives:

- To provide a domain specific vocabulary for describing objects related with IMPs.
- To provide a vocabulary for describing the interrelationships between digital objects and the operations performed on them in the digital preservation context.

More specifically, the following concepts have been introduced:

- "IMP5.Instrument": a specialisation of CIDOC-CRM "E22.Man-Made Object" for modelling musical instruments (e.g. cellos, violins, drums, etc.) used in a performance.
- "IMP6.Equipment": a specialisation of CIDOC-CRM "E22.Man-Made Object" for modelling equipment (e.g. a microphone, a sound mixer or a computer, etc.) used in a performance.
- "IMP8.Performance Procedure": a specialisation of CIDOC-CRM “E29.Design or Procedure” for describing the procedure in which a performance should be carried out.
5. CONCLUSION

The present paper investigated the area of digital preservation of interactive multimedia performances. More specifically, the ontology management approach that we have adopted in the context of the CASPAR project was presented. This approach considers ontologies as a semantic knowledge base of the designated community, containing the necessary metadata for the preservation of IMPS. A layering architecture for this knowledge base is proposed, which defines different types of preservation ontologies and assigns responsibilities among diverse groups of ontology authors. The ontologies of this architecture range from generic to more specialized ones and are integrated with the use of ontology mappings. In addition, a number of extensions to the CIDOC-CRM ontology are proposed. These extensions mainly regard the description of various concepts belonging to the IMP domain.

We are currently working on the deployment of the proposed ontology architecture within the CASPAR framework. We want to acquire further insight into preservation issues arising from the Contemporary Arts testbed of the project. We also plan to evaluate our architecture based on the outcomes of its deployment within CASPAR.

6. ACKNOWLEDGEMENT

Work partially supported by European Community under the Information Society Technologies (IST) programme of the 6th FP for RTD - project CASPAR contract IST-033572. The authors are solely responsible for the content of this paper. It does not represent the opinion of the European Community, and the European Community is not responsible for any use that might be made of data appearing therein.

7. REFERENCES


