Abstract

This paper discusses a novel metadata element set for describing cultural heritage objects and their digital representations, which is particularly useful for building virtual exhibitions. The element set is based on several international standards and current best practice in museums. Additional elements are used for describing new types of data used in virtual exhibitions and taking into account new ways of using that data. The implementation of the element set and a number of tools used to manipulate the metadata descriptions in the ARCO project are described.

Keywords: Metadata, Cultural Heritage, Application Profiles, XML, Interoperability

1. Introduction

With the onset of the digital age, there is a great deal of interest in improving access to various resources using the Web as the means of access and dissemination. This is of particular importance for museums which have collections of artefacts which they currently cannot exhibit in a low cost and easy way. This is due to the fact that either museums do not have enough space to exhibit all their collections, or due to their fragile nature, such objects cannot be placed on display.

In the ARCO project – Augmented Representation of Cultural Objects [1], we are developing a system providing museums with digitisation, management and presentation tools that will allow museum staff to make virtual representations of their artefacts in an inexpensive, efficient and simple manner. An important part of the ARCO project is development of a metadata element set for describing digitised museum artefacts and allowing museums to build virtual exhibitions available over the Internet.

A number of problems must be solved while developing metadata solutions for a system like ARCO, where the element set partially overlaps with existing standards, system specific extensions must be introduced in technical areas that change quickly, such as virtual reality, and the system should be suitable for a wide range of cultural heritage institutions with users having different levels of computer experience.

One of the key issues to be considered is interoperability with existing standards. Interoperability permits transfer of data from and to existing museum databases. Another important requirement is extensibility, which allows incorporation of new metadata elements as they come into use. Flexibility to customise the system to meet the requirements of a specific institution is also of great importance. The tools that operate on the metadata should be independent of the particular element set used and should provide as much support for non-experienced users as possible.

In this paper we present the ARCO Metadata Schema (AMS) and metadata management tools developed for the ARCO project that address the above mentioned problems. The remainder of this paper is organized as follows. In Section 2, the existing metadata standards in the cultural heritage sector and related projects are described. In Section 3 the ARCO system is presented in outline. Section 4 provides an overview of the AMS element set. In Section 5, the implementation of AMS in XML schemas is described. In Section 6 an overview of ARCO metadata-related tools is presented. Finally, section 7 concludes the paper and indicates future work.

2. Related Standards and Projects

2.1. Metadata in the Cultural Heritage Domain

A number of organisations and initiatives have attempted to address the wide-ranging metadata requirements of the cultural heritage sector [2]. Amongst these, some of the most notable are: the Consortium for the Computer Interchange of Museum Information (CIMI); the mda (formerly the Museum Documentation Association); the Art Museum Image Consortium (AMICO); the International Committee for Documentation (CIDOC), the
The European Museum’s Information Institute (EMII) and the Research Libraries Group (RLG).

One of the most important standards for the ARCO project is SPECTRUM [3]. This standard is co-ordinated by the mda [4]; it is a museum standard for documentation of UK museum collections. SPECTRUM has resulted from a collaboration of over one hundred practitioners working in the area of documentation in museums. It comprises procedures for documenting objects and the processes that they undergo. It also identifies and describes the information that needs to be maintained to support those procedures. The intention is that the standard should contain all those functions that are common to most museums. A particular institution would then choose and use those procedures that are most relevant to its own requirements. The advantage of adhering to SPECTRUM is that data exchange between organisations becomes much easier.

CIMI [5] is committed to bringing museum information to a wide audience, encouraging an open standards-based approach to the management and delivery of digital museum information, focused on interoperability and usage of common tools in a museum context. CIMI has also developed a metadata test bed based on SPECTRUM, the XML binding for which became publicly available in March 2003 [6].

The Art Museum Image Consortium (AMICO) [7] is building 'The AMICO Library', a resource that makes digitised versions of artworks available to educational institutions and other museums by subscription. The metadata schema associated with each item in the AMICO Data Dictionary [8] can record detailed information about the type of object that is being described.

CIDOC [9] has an international focus on the documentation interests of museums and similar organisations. The CIDOC Conceptual Reference Model (CRM) [10] is intended to cover all concepts relevant to museum documentation, but most particularly those needed for wide area data exchange. Due to the diversity of museum subjects, that goal can ultimately be achieved only by extensions to the CRM.

In 1997, the Research Libraries Group (RLG) launched the REACH project [11]. The aim of the project was to explore how existing information in museum collection management systems could be extracted and re-purposed to provide online access to museum object descriptive information. The REACH Element Set was to be used for exporting data from disparate museum collection management systems. The set has many commonalities with other cultural heritage data standards; it allows details of provenance, dimensions, materials and production techniques to be recorded alongside information such as title, subject, place of origin, and date of creation.

European Museums’ Information Institute (EMII) [12] aims to establish a working model for the provision of various types of content (text, images, film, video, etc.) from various sources (museums, broadcasters, archives, libraries, etc.). The objective is to identify specific issues that content holders need to have addressed before they make the content in their care available for research purposes.

Although the Dublin Core Metadata Initiative (DCMI) [13] does not deal specifically with museum archives, the Dublin Core Metadata Element Set (DCMES) [14] is of importance for resource discovery across domains and hence of great relevance to any system proposing information retrieval over the Internet. It is also pertinent to issues of interoperability and information exchange.

A metadata review undertaken by the ARCO project [2] revealed that no single existing metadata element set was suitable for the range of processes envisaged in the ARCO system. More specifically, there are no metadata standards which cater for the digitisation, storage and management, and dynamic creation of virtual exhibitions of cultural heritage objects. For a more detailed account of these processes see [15][16].

In developing the ARCO metadata element set (AMS) [17], the initiatives described above were taken into account and in fact, the AMS draws on elements from a number of these, in particular the DCMES and SPECTRUM using the concept of application profiles [18][19][20]. An application profile is a metadata schema which draws on existing metadata element sets, adapting and customising specific elements for a particular local application, further details appear in Section 4.

2.2. Other European Projects

The 3D-MURALE project [21] is developing and using multimedia tools to measure, reconstruct and visualise archaeological ruins in virtual reality, using as a test case the ancient city of Sagalassos in Turkey. A conventional text based database is being developed for storage and retrieval of text, 2D image and VRML image information on archaeological content such as buildings, artefacts, parts of artefacts, drawings, documents, and research papers. The project will make this content available by remote Internet access for other archaeological researchers and members of the public.

The ARTISTE [22] project aims to build a tool for intelligent retrieval and indexing of high-resolution images. This project aims to have a core component that is compatible with existing standards such as Z39.50.

ARCHEOGUIDE [23] has built a system providing information access at cultural heritage sites through the use of augmented reality, 3D visualisation, mobile computing, and multi-modal interaction techniques.

The COVAX project [24] is developing a network of XML-based repositories as a distributed database to be accessed as a single database and will act as a meta-search engine, offering access to book references, finding aids, facsimile images, museum items, etc.
3. Overview of the ARCO system

Figure 1 shows the architecture of the ARCO system. It is divided into three conceptual areas according to the interaction of the user with the data. The producer side, which is responsible for the creation of the multimedia content. The server side that is responsible both for storing the multimedia representations and associated metadata as well as preparing dynamic presentations to be displayed on the consumer side.

4. Metadata in the ARCO System

One of the key components of ARCO is the design of the ARCO Metadata Schema (AMS). AMS is used throughout the whole ARCO process from digitisation to visualisation for describing cultural archives stored in the ARCO database.

4.1. The ARCO Data Model

The ARCO data model (see Figure 2) consists of several related entities. We define a class Cultural Object (CO) as an abstract representation of a physical artefact in the ARCO system. There are two non-abstract entities, which are subclasses of the Cultural Object: the Acquired Object and the Refined Object.

- **Acquired Object** (AO) is a digitisation of the physical artefact used in the ARCO system;
- **Refined Object** (RO) is a refinement of an Acquired Object or another Refined Object. There may be more than one RO created from a single AO or RO.

A Cultural Object (i.e. AO or RO) may be composed of one or more Media Objects (MO). The Media Objects are representations of the Cultural Object in a particular medium. Examples of Media Objects are 3D Model, Simple Image, Panoramic Image, and Description—each with differing MIME types. A Refined Object inherits all Media Objects from the Cultural Object it refines, and may add new ones. For example, a museum curator may create an RO from an AO by adding a 3D Model or Description.

There are two main categories of Media Objects: **simple** and **composite**. Simple Media Objects correspond to media that can be represented in one data object—such as Description and Simple Image. Simple Media Objects contain the object data directly. Composite Media Objects do not contain the data directly, but instead are associated with a number of other Media Objects (either simple or composite). Composite Media Objects represent data objects with complex structure such as a VRML Model, Multiresolution Image or Panoramic Image. The Media Objects that are children of a composite Media Object may have additional attributes that result from the parent-child Media Object relationship.

The association of AMS metadata descriptions to the ARCO data model entities is also shown in Figure 2. The AMS descriptions are used at two levels, Cultural Object and Media Object.

At the Cultural Object level there are three sub-schemas:

- **Cultural Object AMS** describes the physical artefact. CO AMS is common for all COs (AOs and ROs) and is inherited between COs in the refinement hierarchy;
- **Acquired Object AMS** describes the Acquired Objects. The AO AMS is inherited by AO refinements (ROs);
- **Refined Object AMS** describes particular Refined Objects.

As a result, each Acquired Object is associated with two sub-schemas: CO AMS and AO AMS, and each Refined Object is associated with three sub-schemas: CO AMS, AO AMS and RO AMS.

At the Media Objects level there are two sub-schemas:

- **Media Object AMS**, which is common for all types of Media Objects, and
- **Media Object Type AMS**, which is optional and specific for a particular Media Object Type (e.g. Simple Image, Panoramic Image).

Therefore, each Media Object instance may be associated with two sub-schemas: the generic MO AMS and the MOT AMS specific for the particular type of the Media Object.

4.2. ARCO User Groups

Several different user groups have been identified as users of the ARCO system. These user groups need to use AMS metadata in different ways.

- **Cataloguer** is a person cataloguing cultural artefacts in the ARCO database. This user group may either store cultural artefact metadata in the database or search the database by using metadata terms.
- **Digital Photographer** is a person producing digital photographs of artefacts to be used in the ARCO system. The digital photographer may store metadata associated with digital photographs in the database.
- **Object Modeller** is a person creating 3D models of artefacts used in the ARCO system using digitisation and modelling tools. This user group may store metadata in the database associated with the created 3D models.
- **Object Refiner** is a person refining the original 3D models of artefacts. This user group may either store metadata in the database associated with the refined 3D model, or search the database by using metadata in order to retrieve the desired 3D model.
- **Exhibition Designer** is a person responsible for creating content for the end-user interfaces – either 2D or 3D exhibitions. The designer selects a number of cultural objects and a visualization template by querying the database using metadata elements.
- **End User** represents the end-users of the ARCO system. These are the museum visitors, Internet users and
the museum researchers. End users are allowed to browse or search the virtual exhibitions using AMS elements and also create new exhibitions by querying the database with AMS.

4.3. Metadata Requirements in ARCO

ARCO uses metadata to describe four main aspects of the museum objects: curatorial, technical, resource discovery and administrative.

**Curatorial Metadata** records curatorial knowledge about the object and is composed of descriptive elements drawn from the DCMES, SPECTRUM, AMICO and REACH. The terms that have been chosen serve mainly a descriptive purpose to provide basic information about the original physical object from which the digital artefact was created. It is possible to gain additional information by cross referencing the value of the *source* element with a museum’s own holdings information.

**Technical Metadata** records the types of hardware, software and various technical parameters used in the creation, storage and visualisation of the cultural object representation. This type of metadata is often required to address issues relating to digital preservation.

**Resource Discovery Metadata** adopts appropriate elements from the DCMES to enhance the potential for cross-domain discovery of museum artefacts.

**Administrative metadata** adopts appropriate elements from the DCMES to keep track of the creator of a metadata record and any modifications that are made to it.

4.4. Interoperability and Standards

While considering the metadata requirements in ARCO we have found that no single metadata schema is adequate. This is because there is no metadata standard that describes all the processes applied to a cultural object from digitisation to visualisation as in the ARCO system.

Nevertheless, in order to facilitate uptake of the ARCO system by the museum community, it was important to take into consideration current practice in metadata usage by the museum databases. These were analysed based on case studies in the project’s pilot sites: the picture library of the Victoria and Albert Museum in London and the MODES database of the Sussex Archaeological Society. After completion of these studies, it was found that SPECTRUM and DCMES were the standards most compatible with existing metadata used by museum content management systems.

In the interests of interoperability and to avoid reinventing metadata elements, the elements in the AMS are drawn from SPECTRUM, the DCMES, AMICO and REACH base schemas.

4.5. The ARCO Metadata Element Set

The AMS consists of a number of schemas which may be viewed as sub-schemas of the whole element set. In this section we list the elements used in the ARCO element sets indicating their origin. In addition to the Cultural Object AMS, Acquired and Refined Object AMS, and Media Object AMS, six example Media Object Type specific schemas are presented. A lack of adequate space inhibits us from considering the full specification of the AMS here, this is however available on the ARCO project website [17].

**Table 1. Cultural Object AMS**

<table>
<thead>
<tr>
<th>DCMI</th>
<th>Source</th>
<th>Contributor</th>
<th>Format Medium</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Date Created</td>
<td>Format Extent</td>
<td></td>
</tr>
<tr>
<td>Title Alternative</td>
<td>Type Coverage Spatial</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Description Rights</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 2. Acquired Object AMS**

<table>
<thead>
<tr>
<th>DCMI</th>
<th>Identifier</th>
<th>Contributor</th>
<th>Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Date Created</td>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Description Format Extent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Relation Is Version Of</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 3. Refined Object AMS**

<table>
<thead>
<tr>
<th>DCMI</th>
<th>Identifier</th>
<th>Contributor</th>
<th>Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Date Created</td>
<td>Format</td>
<td></td>
</tr>
<tr>
<td>Publisher</td>
<td>Description Format Extent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creator</td>
<td>Relation Is Version Of</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 4. Media Object AMS**

<table>
<thead>
<tr>
<th>DCMI</th>
<th>Identifier</th>
<th>Contributor</th>
<th>Rights</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Description Format Extent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type</td>
<td>Date Created</td>
<td>Rights</td>
<td></td>
</tr>
<tr>
<td>Subject</td>
<td>Creator</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 5. Simple Image AMS**

<table>
<thead>
<tr>
<th>ARCO</th>
<th>Technique</th>
<th>Resolution Compression factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image Size</td>
<td>Compression method Colour Depth</td>
<td></td>
</tr>
</tbody>
</table>

**Table 6. Description AMS**

<table>
<thead>
<tr>
<th>ARCO</th>
<th>Technique</th>
<th>Length Character set</th>
</tr>
</thead>
</table>
5. AMS Implementation

The element sets of the AMS are encoded in XML Schemas [25]. XML provides extensibility and ease of data processing. In the implementation of the AMS, several issues related to interoperability and ease of use have been considered. These include user-friendly display names for elements, source identifiers, element definitions, and content guidelines. Also, a system of auto-generated elements has been used for all elements that can be generated by ARCO components and controlled vocabularies are used to avoid description inconsistencies between objects.

**Display Name:** The element name used in the AMS encoding is not always appropriate for display in ARCO modules. To make the system more user-friendly, additional display names have been included in the AMS XML Schema.

**Source Identifier:** To enable AMS metadata interoperability with other systems and applications, each element in AMS has a source identifier [26]. In ARCO we have developed the AMS by adopting elements from various standards. Each adopted element is encoded with its associated identifier to provide information to indicate from which standard the AMS element was taken. Source identifiers maximize the AMS’s potential for interoperability.

**Element Definition** is used to provide a statement that clearly describes the concept and essential nature of the data element. It can be used by an end-user interface in order to facilitate the user’s work.

**Element Content Guideline** provides the end-user with context information that advises the user on how to fill in the value of the metadata element.

**Auto-generated Elements:** Values of these elements are set automatically by the application facilitating the creation of metadata.

**Dictionary Elements:** Values of these elements are taken from a special dictionary stored in the ARCO database. This provides an additional layer of data consistency.

All of this additional information is provided in the `<appinfo>` section of each element annotation in the AMS XML Schema, as shown in Figure 3.

```
<xsd:element name="name" type="xsd:string">
  <xsd:annotation>
    <xsd:appinfo>
      <arco:autoGenerated>MOName</arco:autoGenerated>
      <arco:displayName>Name</arco:displayName>
      <arco:identifier>
        http://purl.org/dc/elements/1.1/title
      </arco:identifier>
      <arco:definition>
        A name given to the Media Object
      </arco:definition>
      <arco:content>
        This should be a name appropriate for ...
      </arco:content>
    </xsd:appinfo>
  </xsd:annotation>
</xsd:element>
```

Figure 3. Encoding of additional information in AMS XML Schema

5.1. AMS Implemented as Local Schemas

The AMS is implemented as a set of distinct schemas that correspond to the AMS element sets briefly described in Section 4.5 and detailed in ARCO Deliverable D8 [26]. The schemas include: Cultural Object Schema, which defines the cultural object metadata, Acquired Object Schema, which defines the acquired object metadata, Refined Object Schema, and so on, see Figure 4.

As described in ARCO public Deliverable D9 [27] the list of Media Object Types supported by the ARCO system is extensible. A user can add new Media Object Types without the need to modify the database schema or tools. A number of pre-configured Media Object Types are installed within the ARCO system: Simple Image, 3ds max Project, Description, Multiresolution Image, Panorama Image, VRML model, Image Collection, QuickTime VR.

For some of these Media Object Types the ARCO system provides specific sub-schemas. If custom types are added to the system it is possible to either use only the general MO AMS or load into the database a new schema specific to the particular Media Object Type and use both.

5.2. AMS Implemented as Application Profiles

The ARCO consortium has been investigating the concept of application profiles for AMS implementation. An application profile is defined as a schema that consists of data elements drawn from one or more namespaces, combined by implementers, and optimised for a particular local application [18]. The ARCO fourth prototype system will adopt application profiles to implement the AMS.
6. AMS and ARCO components

As discussed previously, AMS is used throughout the whole ARCO system in different ARCO components. The components and their relation with AMS are discussed in the following sub-sections.

6.1. AMS and the ARCO Database

The ARCO database is a central element of the system architecture. ARCO components store data and associated metadata in the database and retrieve them when needed. The ARCO database has been implemented on top of the Oracle 9i Object-Relational Database Management System, which provides advanced functionality needed to store and efficiently manage XML data.

To assure data type extensibility, the database design uses a meta-schema approach where data model entities are not implemented by fixed structures in the database, but dictionaries of data types, their attributes and relationships are stored as data. This approach allows flexibility in adding or modifying types of objects supported by the database without the need to modify the ARCO database schema.

6.2. AMS and ACMA

To allow efficient and user friendly management of data contained in the ARCO database a content management application has been designed and implemented. This application is called ARCO Content Management Application (ACMA). The main purpose of ACMA is to manage data and metadata in the ARCO database.

In brief, ACMA consists of four main data managers and eight additional tools. The main data managers are: Cultural Object Manager, Presentation Manager, Template Manager, Template Object Manager.

The additional tools are XSL Manager, Configuration Manager, Data Type Manager, Template Domain Manager, AMS Schema Manager, Media Object Type Manager, Template Object Type Manager, and Security Manager.

One of the most important data managers is the Cultural Object Manager. It is used for managing data associated with Cultural Objects. This manager is equipped with a built-in Metadata Editor for creating and updating metadata records. Screenshots of the ARCO Metadata Editor are presented in Figures 5 and 6.

The ARCO Metadata Editor is independent of any particular structure of metadata. Instead, it uses the AMS XML Schema which defines the structure of the AMS. The Metadata Editor uses it to enforce insertion of only those metadata elements that are allowed in a given context (see Figure 5) and provides hints and content guidelines for users as shown in Figure 6.

The values of metadata elements entered by a user are instantly validated against the type definition in the AMS XML Schema. For elements constrained by enumerations, all possible values are presented in a drop-down menu. For dictionary elements, a special panel is displayed allowing access to the dictionary contents. The AMS Metadata Editor also fills-in values for auto-generated elements.

Another important ACMA tool related to AMS is the AMS Schema Manager. It allows administration of multiple versions of the AMS schema stored in the ARCO database; new AMS schemas for new data types may also be created.
All ACMA tools are independent of the particular schema version and can still be used even if the schema evolves. Also, with the AMS Schema Manager, each museum may create its own version of the AMS to suit its needs and practices. The AMS Schema Manager is presented in Figure 7.

6.4. AMS and XDE

In order to keep the ARCO architecture extensible and enable data interoperability between ARCO and other systems and applications, all data exchange interfaces of ARCO components are based on XML. For this purpose, a single XML format called the XML Data Exchange (XDE) is used. XDE guarantees that the data chunks exchanged between ARCO components are structurally valid and complete and that appropriate metadata is attached.

The database plug-in illustrated in Figure 8 allows the user to visualize and update the necessary metadata associated with Cultural Objects and Media Objects to enable efficient searching and browsing of the cultural object archive.

6.5. AMS and Augmented Reality Interfaces

The Augmented Reality Interface (ARIF) is used by ARCO end-users. ARIF interfaces allow access to both 2D and 3D virtual exhibitions containing virtual representations of cultural artefacts. These exhibitions are visualized in virtual and augmented reality worlds either inside museums or through the Web. Virtual exhibitions are created dynamically based on the database contents and user interaction.

Metadata are an integral part of the information visualized in this component. Figures 10 and 11 illustrate visualization of the metadata of the same cultural object within the museum and over the Web. Figure 12 illustrates the concept of searching the database content using elements from the AMS.
7. Conclusions and Future Work

The ARCO metadata element set, or AMS, addresses all the processes that make up the full ARCO system; these range from digitisation and modelling, through refinement and storage, to visualisation of cultural museum artefacts in virtual and augmented reality exhibitions. AMS extends the application areas of metadata in the cultural heritage sector beyond the standard cataloguing of museum collections towards multimedia Internet publishing and dynamic virtual reality systems, while at the same time conforming with current museum best practice in the interests of interoperability.

The AMS is currently undergoing a process of evaluation and assessment undertaken by museum pilot sites, which are participants in the ARCO project. Additional work in this area involves further investigation and implementation of the AMS as an application profile as well as consideration of the CIMI.
SPECTRUM XML binding which has recently become available.

Exploration into the generality of the AMS for areas such as e-commerce, computer games and other forms of entertainment would prove a fruitful exercise in establishing its value with regard to other types of 3D applications.

Acknowledgements

This work has been funded by the EU IST Framework 5 Programme under Key Action III Multimedia Content and Tools, Action Line III.1.6 “Virtual Representations of Cultural and Scientific Objects”. The authors also acknowledge other partners of the ARCO consortium for their considerable efforts in the project, in particular Jessica Lindholm, formerly at UKOLN, who undertook the majority of the metadata review for the project and Joe Darcy, University of Sussex who was responsible for much of the work on the IMRR.

References

Figure 1. Major components of the ARCO system

Figure 2. AMS descriptions of ARCO data model entities