Building digital collections for archeological sites: Metadata requirements and CIDOC CRM Extension

Georgios S. Gkrous[‡†] and Mara Nikolaidou[†]

† *Department of Informatics and Telematics, Harokopio University of Athens, 17671, Athens, Greece.*

‡ *Department of Librarianship and Information Systems, TEI of Athens, 12210, Athens, Greece*
gtgkrous (at) yahoo.gr *mara (at) hua.gr*

Abstract: *Cultural heritage information management and provision can be improved by the usage of Internet and related technologies. The research effort presented in this paper focuses on the creation of a digital collection representing an archeological site. Such a collection is composed of digital entities representing monuments and describing them using proper metadata. Such a metadata model should provide information for the monuments, ease the visitation of the site by Internet users and promote it. It can be based on CIDOC CRM, which is a formal ontology for the uniform description of cultural resources. CIDOC CRM's specification focuses on museum documentation, thus it is necessary to extend the ontological model to effectively describe archeological sites, emphasizing spatial characteristics and facilitating the integration of the archeological site monuments with exhibits hosted in other sites, for example museums. Thus, metadata fields describing the structure of physical objects, e.g. monuments, and spatial and conceptual correlations between exhibits are explored.*

Keywords: *digital collection, CIDOC-CRM, ontology, cultural heritage, archeological site*

DOI: It would be provided by publication house

I. INTRODUCTION

Utilizing information and communication technology (ICT) to promote cultural heritage remains a hot topic (Papaconstantinou, 2008). The explosion of Internet usage indicated an alternative way to advertise the visitation of archeological sites and museums significant to every nation's history. Therefore, especially for countries such as Greece being characterized by cultural richness, promoting archeological sites by enhancing their digital presence on the Web is a task of significant importance.

Information regarding a visit to an archeological site can be categorized to pre-visit and during-visit data (British Museum, 2006). Pre-visit information is usually provided via tourist guides or a Web site, targeting to attract and provoke all interested people to actually visit the site. During-visit information consists of the material used to ease the visitor's walk around in the site and enhance his/her experience by additional information. Special brochures, which include a site map and a short description of all monuments, or audio devices, playing recorded messages related to the monument in multiple languages according to input keys, are usually available. Such tools have specific restrictions. Few pages of a tourist guide or general data displayed on a web page may not be considered as satisfying by the interested visitor. In such case, he/she may search for additional information after the visit has ended. The attendance of professional guided tours is always an alternative, but as the technology has penetrated every day life, the usage of mobile devices as smart phones, may be considered as a way to provide information to visitor during the visit. In such case, advanced navigation services around the site may be provided along with all kinds of information regarding monuments and exhibits that might be moved away from the site for numerous reasons (for example preservation, security, etc).

A world wide trend is observed regarding the use of telecommunications in the cultural field. Related research efforts exposing the benefits of such cooperation are the creation of a cultural information network on Canada (CHIN), the virtual outdoor museum in Latvia and the preservation of English and French landscapes using image processing (Hemsley *et al.*, 2005).

The benefits of exploring ICT to promote cultural heritage are expressed by the numerous research projects conducted all over the world. This effort is supported by European funded projects regarding digital libraries, digitalization techniques and Internet archives. Digital cultural experiences field refers to the exploitation of edge technology tools for increasing the knowledge and experiences that cultural heritage sources (museums and archeological sites) share to the public and is explored by projects are ARtSENSE, CINeSPACE and CHESS (European Commission, 2011).

Regarding faced restrictions on a archeological site, by using new technologies it is feasible to represent a monument in a digital form and portray its original or current shape (visualization, animation) (Vlachakis *et al.*, 2001). Furthermore, multiple geographical and conceptual connections between monuments can be established in order to present an integrated aspect of History (for example objects with the same creator or similar use). Advanced services such as navigation for mobile users can be provided guiding visitors among the site's monuments (Bonfanti *et al.*, 2007) and giving all additional information according to their position. The provision of such a service can be facilitated by the creation of a digital collection for archeological site, utilizing digital library technology. The collection should contain monument related information and enable navigation. Additional material, for example digital animations representing monuments in their original form, as well as

digital representations of their current state, may also be included. The digital collection may be useful as during-visit, pre-visit and post-visit data, since it might constitute a continuous source of information for the visitor. Museums, for example the Hermitage Museum in St. Petersburg (http://www.hermitagemuseum.org/html_En/12/hm12_0.html) or the Tate Gallery in London (http://www.tate.org.uk/collection/) have successfully employed such projects during the last decade.

Such digital collections should be described in a common fashion, supporting their integration in larger collections, as Europeana pan-european cultural digital collection (http://www.europeana.eu/portal/), and enable searching their content in a unified fashion. Thus, CIDOC Conceptual Reference Model was proposed by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM), as a formal ontology created to support information exchange between heterogeneous cultural sources.

In this paper, we explore the potential of building a digital collection for an archeological site, focusing on digital object structure to depict monuments and the metadata requirement for their efficient description. The adoption of CIDOC CRM is suggested for this purpose, while the proposed extensions are discussed.

II. DIGITAL COLLECTION CHARACTERISTICS

The creation of a digital collection for an archeological site and the rendering of an advanced navigation service using its material is the focus of our effort. Every monument will be represented by a digital entity described by all related data. Multiple connections between objects of the same site (spatial relations) and between objects that are placed on another site or museum (conceptual relations) must be supported. The provision of the advanced navigation service relays upon maintaining all necessary information related to monuments (Gavalas *et al.*, 2005). The benefits of such an effort relate to both the strengthening of the site's digital presence (pre-visit information) and the improvement of services' quality when visiting the site (during-visit information).

The information structure and management must be conducted taking into consideration information heterogeneity, principles of semantic web, simplification of information and interoperability (Ravindranathan, 2004). For this purpose, we identified the requirements the supported metadata scheme should satisfy and explore the potential of using CIDOC CRM .

III. METADATA FOR ARCHEOLOGICAL SITES

A. CIDOC Conceptual Reference Model

The CIDOC Conceptual Reference Model is the outcome of a long term development work carried out by the International Committee for Documentation (CIDOC) of the International Council of Museums (ICOM). Since 2000, the development group responsible for CIDOC, named CIDOC CRM Special Interest Group, in collaboration with ISO aims to form an international standard (ICOM/CIDOC, 2010).

The main objective of CIDOC CRM is to provide all the semantic clarifications in order to allow the enlargement of local resources to global ones. It provides a formal language concerning cultural heritage, especially museum documentation, for the definition of data relations. It aims to be used as a guide for all interested parties, when structuring and relating cultural information assets, to support associative queries by providing a basic model of associations and to ensure that the implementation of data transformation algorithms will be performed without loss of meaning. The intended scope of CRM is to cover not only museum but all types of collections (arts, archeology, ethnography) and contribute to the harmonization between cultural information, libraries and archives.

The CIDOC CRM model contains classes, properties and inheritance rules. Classes and properties can be identified by their initial code (E for classes and P for properties). By its structure, CRM model is extensible and users are encouraged to create new instances according to their needs (Cripps *et al.*, 2004). The latest released version (Version 5.0.2 January 2010), CIDOC CRM includes 90 classes and 148 properties.

An example representing reasoning about spatial information is shown in Fig.1. The relations between main classes and their subclasses are shown as arrows and properties as rectangles. An exhibit represented by Physical Thing is connected to Place via the action-property Has current location. Place can be identified by Place Name, Address or Spatial Coordinates.

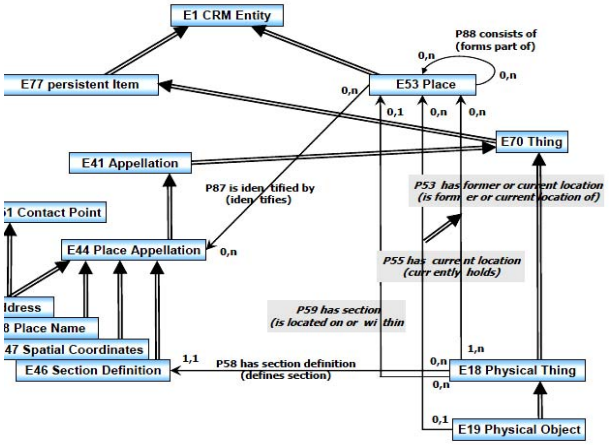

Figure 1. CIDOC CRM spatial information

B. Metadata Requirements

In a digital collection for an archeological site, every monument is represented using a digital object (complex or simple), which encompasses all related information and corresponding connections with the other monuments. The set of digital objects constitute the digital collection of the archeological site. All the information required to fully describe each monument are grouped into two basic categories: *Physical Info*, related

to the monument as perceived in the physical world and *Digital Representation Info*, related to the digital representation of the monument. Each of them, is further decomposed into subcategories, as represented in Fig.2 based on CIDOC CRM concepts and notation.

Subcategories of information consist of metadata, describing the monument or its digital representation and digital material representing the monument. They are discussed in the following:

- *Structural info*: It is divided into two subsets specifying whether the object is part of a monument (Part Of info) or it is contained by/contains other exhibits (Contain Info). The introduction of these relations is necessary for sufficient description of situations, which are very common in cultural heritage, where a monument is divided into constituents or it may contains other artifacts, as sculptures. The location of digital object parts should be specified. This way, in the case where a sculpture found within a building during the excavation of the site is moved to a museum, the visitor may be informed and have access to it during his/her visit in the site.
- *General info*: It contains descriptive info about the exhibit regarding its creator, its cause of creation, its use, the submitted procedures (excavation, reconstruction, restoration, modification) and the responsible persons that have performed these actions, the historic periods that all actions have taken place, the material, the used tools and its owner.
- *Multimedia info:* It is composed of two subsets (images and videos). The subset of Images includes digital processed images that show the original or current form of the exhibit and contribute to an integrated aspect of the object through visualization. The subset of Videos is a gallery of video files from excavation, object modifications and animated representations, related to the monument.
- *Correlation info:* It is consisted of two subcategories that refer to geographical and conceptual correlations (*Spatial info, Conceptual info*). The set of spatial information include navigation data to all objects that belong to the monument's surroundings and the corresponding path to be followed. The path is defined by parameters as direction, distance and height that can be processed by a geographical information system (GIS) for assisting the provision of the navigation service (Coors, 2004).

  The conceptual information includes links to all objects that are semantically related to the specific monument and their location (other archeological site or museum). Additionally, the reason of objects' similarity is provided, for example they belong to the same collection, they have the same creator or use. The role of this set is quite significant, since conceptually related objects contribute to visitors' understanding of the historical sequence.

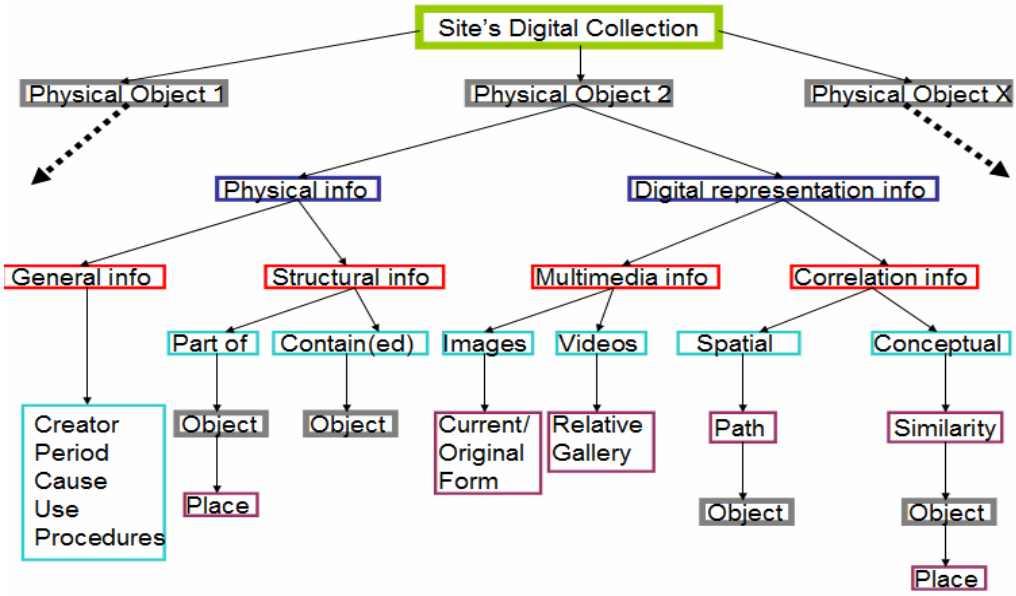

Figure 2. Metadata categories describing archeological monuments

C. CIDOC CRM extension

CIDOC CRM provides concepts for description of museum areas. Supported classes and properties facilitate the administration of museum's content. In the case of an archeological site, CIDOC CRM may support the presentation of some constituents, but is not adequate for the representation of all the information depicted in Fig.2. In order to overcome these restrictions, we suggest the extension of the model, focused on certain directions.

Detailed classes should be introduced regarding the procedures an object is submitted to (related to General info). Current entities supported by CIDOC for this purpose are Activity, Modification and could be enriched with Excavation, Reconstruction and Restoration. Fur-

thermore, new classes must be added for multimedia data including relative videos and images presenting the current/original form of the exhibit.

Important extensions concern structural and correlation information. Since CIDOC CRM was developed for museum oriented purposes, it does not provide the tools for the exploration of the way an artifact is related to its surroundings. Object's structural data representation is limited by two basic restrictions. Firstly, there are no instances to sufficiently represent a physical artifact as a part of a larger monument (for example a part of Parthenon's aetoma). Furthermore, there is no adequate representation of exhibits being contained by other exhibits (for example the artifacts of an ancient arcade). These circumstances are very commonly met on an archeological site and must be effectively represented in the digital world. New classes and properties must be created to specify, if the object is a part of another one, and also allocate the remaining parts. Similar actions will be performed for detailed description of objects containing or being contained by other objects (addition of contain(ed) tag and connection with the other object).

Concerning spatial correlations, CIDOC CRM current release contains only one entity (spatial coordinates) for the definition of object's position. New classes and properties must be defined specifying the surroundings (objects that are in a close distance) of each exhibit and the path that should be followed. Path class will have Direction, Distance and Height as subclasses. This procedure defines a unique path between two objects which is very important for GIS system and the navigation service. Conceptual relations are also not supported by the latest release of CIDOC CRM. New entities and actions must be created to relate similar objects and simultaneously specify the reason of similarity and the location of objects that do not belong to the site.

IV. CONCLUSIONS – FUTURE WORK

The proposed research targets the creation of digital collections for archeological sites supporting pre-visit and during-visit experience. It facilitates the navigation of the visitor on site using a smart mobile phone and enhances the visitor's experience.

Compared to already developed museum navigation services (Cosley *et al.*, 2009), it supports a wider range of provided information, by supporting interrelations to exhibits placed on another site or museum, exceeding the strict boundaries of the archeological site. A wider cultural heritage source is created in order to fulfill all visitor's learning needs.

Digital content organization and management is performed by the usage of the CIDOC CRM ontology model, thus it is extended to describe archeological sites. Metadata requirements for the creation of the proposed digital collection for archeological sites were documented and corresponding CIDOC CRM extensions were identified. New classes and properties must be added to describe structural and correlation information.

We are currently on the detail definition of an ontological model based on CIDOC CRM and its application on an archeological site. The metadata model will be assessed and technical specifications for its support and provision of the advanced navigation service will be defined.

REFERENCES

Bonfanti C., Chiabrando F. and Spano A., "Studies and knowledge fusion in archaeological site maps managed in a 3D GIS", *XXI International CIPA Symposium, 01-06 October*, Athens (2007).

British Museum, "Tours of adult ESOL students", http://www.britishmuseum.org/pdf/pre-visit%20information%20for%20class%20tutors_aug06.pdf [retrieved 29 July 2011].

Coors, V., "Using Wearable GIS in outdoor applications", *Extended Abstract on International Workshop for Interactive Applications of Mobile Computing*, Germany (2004).

Cosley D., Baxter J., Lee S., Alson B., Nomura S., Adams P. , Sarabu C. and Gay C., "A Tag in the Hand: Supporting Semantic, Social and Spatial Navigation in Museums", *Proceedings of the 27th international conference on Human factors in computing systems*, Boston (2009).

Cripps P., Greenhalgh A., Fellows D., May K. and Robinson D., "Ontological Modelling of the work of the Centre for Archaeology", *CIDOC CRM Technical papers* (2004).

European Commission, "ICT Research in FP7", http://cordis.europa.eu/fp7/ict/telearn-digicult/digicult-projects_en.html[retrieved 15 July 2011].

Fox E., Gonçalves M., Watson L. and Kipp N., "Streams, structures, spaces, scenarios, societies (5s): A formal model for digital libraries", *ACM Trans. Inf. Syst.*, vol. 22, pp. 270-312 (2004).

Gavalas D. and Economou D, "Development Platforms for Mobile Applications: Status and Trends", Department of Cultural Technology & Communication, *IEEE Software*, *28*(1), 77-86, Mytilene, Lesvos, Greece (2010)

Hemsley J., Cappellini V. and Stanke G., "Digital Applications for Cultural and Heritage Institutions", Chapter 1, Introduction and International Overview (2005).

ICOM/CIDOC Documentation Standards Group, "Definition of the CIDOC Conceptual Reference Model",http://www.cidoc-crm.org/index.html, [retrieved 20 July 2011].

Papaconstantinou D., "The materiality of cultural information and its absence: Digital curation from an archaeological perspective", *2008 Annual Conference of CIDOC*, Athens (2008).

Ravindranathan U., "Prototyping Digital Libraries Handling Heterogeneous Data Sources - An ETANA-DL Case Study". *Masters Thesis*. Computer Science, Virginia Tech (2004).

Vlahakis V., Ioannidis M., Karigiannis J. et al, "ARCHEOGUIDE: First results of an Augmented Reality, Mobile Computing System in Cultural Heritage Sites", *Proceedings of VAST* (2001).