DIGITAL COLLECTIONS SERVING AS EDUCATIONAL REPOSITORIES

NIKOLAOS V. KARADIMAS
VASSILI LOUMOS

National Technical University of Athens,
Multimedia Technology Laboratory
9 Heroon Polytechneiou, Zografou Campus,
157 80 Athens, Greece
E-mail: nkaradim@central.ntua.gr
loumos@cs.ntua.gr

NIKOLAOS P. PAPASTAMATIOU

Omega Technology
4 El. Venizelou, Kallithea
17676 Athens, Greece
E-mail: nikos@omegatech.gr

Abstract: This work presents an attempt to make digital content more accessible, exploitable and easy to use, throughout Europe, by developing digital collections which will be available free of charge and can be used to further the education and qualifications of all interested parties. At the moment, many museums' collections are almost unknown to the majority of the students and researchers, and even members of these museums do not know about the treasures kept in their neighborhood. Some of the collections are unique all over the world. The work proposed here, offers functions such as multi-criteria retrieval (context-, semantic- and content-based queries) and interfaces for Learning Content Management Systems, that will allow trainers to develop educational scenarios for targeted user groups combining material from digital collections all over the world. The main advantage of this work is that its open architecture can be offered to any museum interested in participating in the network in order to promote its exhibition in digital form. The architecture behind this service utilizes the technology of the web services offering, on demand, knowledge directly from the museum's databases for e-learning purposes.

*Keywords***:** Museums, Digital content, Exhibits, Education, Information sharing, e-Learning, Metadata, Web services.

INTRODUCTION

According to the ICOM code of ethics, all museums should aim to making their collections and the relevant information they hold as accessible to the general public as possible, taking into account restraints arising due to reasons of confidentiality and security (ICOM Code of Ethics for Museums 2006).

To this end, following the evolution of the internet as the primary means of communication and promotion, many museums have deployed projects to digitize their exhibits jointly with the information that describes them that is stored in the form of metadata. In this process, there are two basic principles that museums should take under concideration: to save the material in the correct format so that it can be used for different purposes (preservation, diffusion, print, read, etc) and to maintain interoperability of the published content.

Digitization is the process of converting information into a digital format. Text and images are digitized in a similar way: a scanner captures an image (which can also be a text received as an image in this stage) and converts it to an image file, such as a bitmap. As far as texts are concerned, an optical character recognition (OCR) program analyzes the text image for light and dark areas in order to identify each alphabetic letter or numeric digit, and converts each character into an ASCII code.

Audio and video digitization use one of many analog-to-digital conversion processes in which a continuous variable (analog) signal is changed, without altering its essential content, into a multi-level (digital) signal. The process of sampling measures the amplitude (signal strength) of an analog waveform at evenly spaced time markers and represents the samples as numerical values for input as digital data.

The purpose of digitizing is to make the digitized object more accessible to the public. The Semantic Web, in which information given should be well-defined so that computers and people can cooperate more efficiently, is the platfrom to support this motivation. Semantic Web is based on the idea of having data on the Web defined and linked in such a way, so that its discovery, automation, integration and reuse across various applications become more effective (Hendler et al. 2002).

To achieve this, the quest of matching equivalent fields in different systems describing similar objects is the main barrier. The solution which is interoperability, also presents a twofold problem: Semantic and Syntactic interoperability.

Syntactic Interoperability

The term Syntactic Interoperability refers to the use of common language for data presentation (museum exhibits, in this case). The standard most widely used for metadata representations is Extensible Markup Language (XML) (World Wide Web Consortium (W3C) 2006). XML is a simple and very flexible text format, designed to meet the challenges of large-scale electronic publishing and it plays an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.

An example of a simple but widely used standard that uses a vocabulary of fifteen properties for resource description is the Dublin Core Metadata Inititative. In the Dublin Core, syntax is reduced to arrangements of single terms in the form of entity relationships using triples. The initiators of the Dublin Core use semantics to refer to the definition (or meaning of the elements or fields). They focus on the containers rather than the contents of those fields.

Table 1**:** The 15 metadata elements of Dublin Core.

Title	Contributor	Source
Creator	Date	Language
Subject	Type	Relation
Description	Format	Coverage
Publisher	Identifier	Rights

Dublin Core elements in the corresponding XML file created for each digital material are expressed with the dc: namespace prefix before them, so for example the *Title* element becomes dc:title (all lowercase) (Beckett et al 2001)

<dc:title>My Home Page</dc:title>

The xml syntax used in the previous example is used for representing digital content in a uniform way achieving syntactic interoperability. What is between the tag has to do with semantic interoperability.

Semantic Interoperability

The term Semantic Interoperability refers to the ability of the documented information to be understood by every user or system that retrieves it. Semantic interoperability supports the integration of information offered from distributed resources. Several standards have been developed in order to help museums and other organizations to document their resources in a common way. These standards provide the semantic definitions and clarifications necessary to transform disparate, heterogeneous information sources into a coherent global resource available to Internet users.

Based on the aforementioned principles of interoperability, the proposed solution takes advantage of web services technology and architecture to create a global repository (one stop shop) for students, teachers, researchers and anyone interested in using museum material for educational purposes. The service is free of change. Museum content presentation, along with copyright issues and rights are to be handled by the museums themselves. The advantage of this service is that it minimizes administrative costs, which pose the main barrier in using subscription-based services.

The services offered include:

- Multi-criteria retrieval (context-, semantic- and content-based queries). More specifically, content-based (join and selection) as well as traditional operations will be carried out throughout the service.
- A web service, acting as the medium through which learning objects are found and assessed, before being integrated in e-Learning authoring tools.

The rest of this paper is structured in the following sections: *Section 2* presents a theoretical overview of the related literature. *Section 3* describes the technology behind the proposed service, while in *Section 4* the proposed solution is presented. *Section 5* presents the results while *Section 6* is devoted to discussion and future works.

RELATED WORK

Similar work was undertaken from the Art Museum Image Consortium (AMICO) and ArtsConnectEd. AMICO was a non-profit organization of institutions which posses art collections, collaborating to enable educational use of museum multimedia. AMICO operated from 1997 to 2005 and its members had pooled their collective resources to create a digital library known as The AMICO Library™, which was a licensed digital educational resource available under subscription to universities and colleges, public libraries, elementary and secondary schools, and museums as well. It represented works in the collections of AMICO Members (AMICO 2005). ArtsConnectEd is another relevant product of a partnership between The Minneapolis Institute of Arts and the Walker Art Center. The goal of ArtsConnectEd is to make arts education timely, engaging, interactive, and pertinent for both teachers and students of all ages. ArtsConnectEd offers a user-friendly Web site with access to the combined art collections, libraries, and archives of the Walker Art Center and The Minneapolis Institute of Arts (ArtsConnectEd 2005).

METHODOLOGY

The innovation of the proposed service is that it is open to any museum interested in participating and making its digitized collections available to the web. Special integration to the central service is not necessary, since it is developed in order to communicate with any museum services registered to the system's address book. The development of a web service, offering information on exhibits in the required form (mapped to the systems ontology), as well as the registration of this web service to the central application, are both tasks that should be undertaken by museums interested in participating.

Nowadays, most museums have both the necessary infrastructure and the expertise to support the provision of such a web service. The benefits of this kind of connection are many, while the actual cost of developing of the web service is relatively small. The ultimate goal is that each museum contribution will add to the formation of a perpetually growing knowledge grid, which will perform a specific set of operations and offer information for educational purposes to the general public [figure 1].

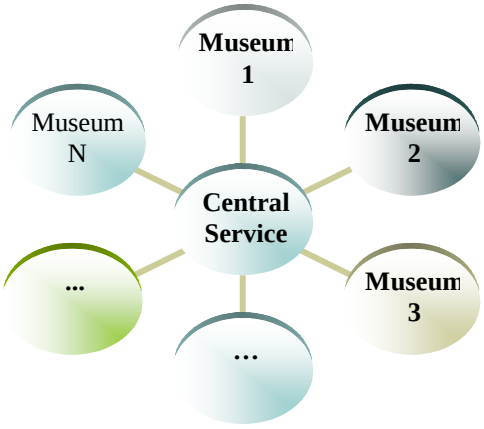

Figure 1: Museums network based on web services and interoperability principles.

The term Web services describes an important emerging distributed computing paradigm that differs from other approaches such as DCE, CORBA, and Java RMI in its focus on simple, Internet-based standards to address heterogeneous distributed computing. Web services define a technique for describing software components to be accessed, methods for accessing these components, and discovery methods that enable the identification of relevant service providers. Web services are programming language-, programming model-, and system software-neutral (W3C Working Group Note 2004).

Web Services interact with other applications by exchanging messages in Simple Object Access Protocol (SOAP) format, while the contracts for the message exchanges that implement those interactions are described via WSDL interfaces.

A Broker server will be set up in order to host the private Universal Description, Discovery and Integration directory (UDDI) so that the museums will be able to register their web services to the proposed system. Each museum web service should be registered to the UDDI directory in order to publish its content. UDDI is an XML-based standard for describing, publishing, and finding Web services. The service consumers can then ask the broker to locate a specific web service and use it to implement a user's request. The proposed system acts as a service consumer and accesses all the museum web services registered to the UDDI registry.

For this specific type of application, the need emerges for stateful web services. When contacting a museum web service registered in the private UDDI, the system has to be informed, if the service is active or not, in order to estimate the time elapsing before it provides the user with the requested information. For that reason, real time communication will be set up (in the form of parallel web services) so that each web service will continuously inform the system about its availability.

To successfully utilize these web services, XML/SOAP messages will be used to exchange information with the main system. SOAP is a lightweight protocol intended for exchanging structured information in a decentralized, distributed environment (Gudgin et al 2003).

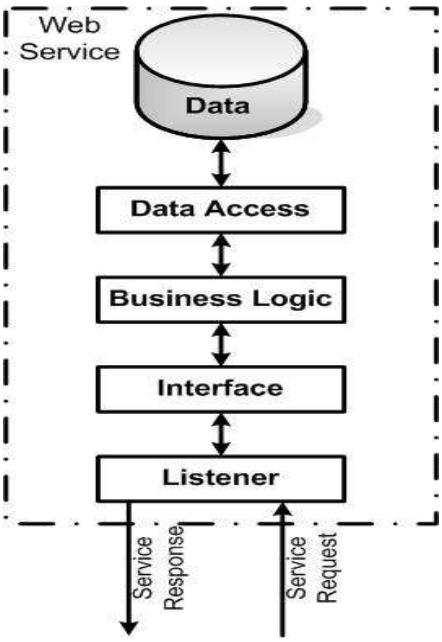

Figure 2: Description of a web service

XML is the technology used with SOAP messages that form the requests and responses from a web service. A client invokes a web service by sending an XML/SOAP message, and then waits for a

corresponding XML/SOAP response. Because all communication is in XML, web services are not tied to a particular operating system or programming language and any museum can build this web service on its current infrastructure.

The content of these messages should comply with the general known standard describing metadata CIDOC Conceptual Reference Model (CRM), now ISO/CD21127[3] which is a common used standard enabling information exchange and integration between heterogeneous cultural heritage relevant sources. Expanding CIDOC's structure with metadata from Learning Object Model's (LOM), a standard that specifies the syntax and semantics to fully and adequately describe a Learning Object, will help to make the identification of the most adequate information on museum exhibits more effective for the indented e-learning use.

PROPOSED SOLUTION

In this work, we propose the development of thematic networks in which the available relevant digital data will be presented in a uniform way. The advantage of the proposed solution is that it uses an open architecture therefore, irrespectively of content size, all information is easily offered for publication. Retrieval techniques and successful collection of related and supplementary information from the museum digital repositories will be designed on metadata schemes, ontologies that will be based on common digital metadata schemes and educational standards.

The main work, should be undertaken at various levels of detail (with examples): character encoding (Unicode, ASCII, etc.), language (English, Japanese, etc.), controlled vocabulary (MeSH, SNOMED CT, etc.), message structure (DICOM, HL7, etc.), learning metadata (LOM, SCORM) and formatting (ASN-1, XML, etc.).

The central service (utilizing the museum web services), designed on this ontology, will be able to offer the requested information about the different exhibits, their characteristics, their creators, the location they were found or created, their use, and so on, and at the same time it will be able to filter this information for the public. This filtering will use LOM's metadata scheme (age, difficulty, education, etc) to result in adequate information which the user will be presented with. This information is provided by the museum that hosts the particular exhibit and it is subject to any restraints arising for reasons of confidentiality and security. The LOM is dependent on data bindings for the interoperable representation, communication and transport of metadata records. "Binding" refers to the expression of the LOM data model --or others-- via a formal language or syntax for the purposes of effective data exchange and processing. In the case of both the LOM and Dublin Core, the general standards used for creating these bindings include RDFS (Resource Description Framework Schema) and XMLS (eXtensible Markup Schema Language). In the case of the LOM, the specific way that the XML Schema language is used to format or "bind" LOM data is itself the subject of standardization (Friesen N. 2004).

The special modules that will serve the digital material to open source authoring tools through web services, will be developed and integrated in these platforms. These web services, such as wizards, will assist professionals to retrieve and make use of the most pedagogically appropriate content in order to create educational packages. The wizard will ask for information input like:

- Who will take part in the course (group profiles)?
- What is the participant's motivation to this specific training?
- What media are the most interesting, the most effective?
- What is the GOAL of the training?
- Will there be another part of the course tought in the conventional way?

The core service, to achieve the aforementioned functionality, includes a smart search engine that will communicate with the web services subscribed in the services private UDDI registry. The results from those "alive" web services will be presented to the user together with additional information offered by the museum. All the nodes depicted in the next diagram [figure 3], describe the architecture between each museum and the central service and communicate with each other through a Transmission Control Protocol/Internet Protocol (TCP/IP).

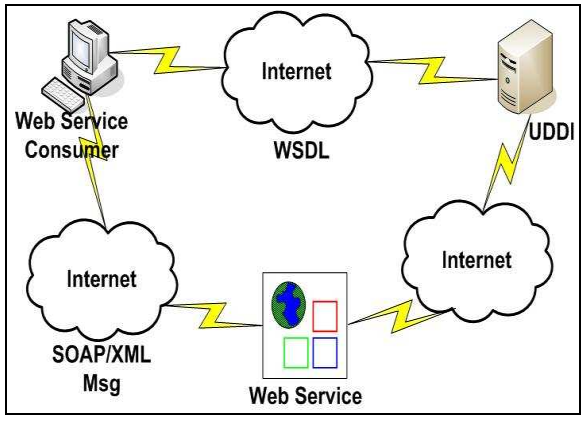

Figure 3: Description of the proposed architecture.

The advantage of the architecture described is that the pool of available exhibits will grow rapidly as

each new museum joins in. The fact that the steps, which the museums have to follow in order to develop and register their own web service to the service's UDDI, are simple secures the viability of the service. Administration costs for the service, often the most common barrier to similar projects, are minimized since the administrator's sole task is to observe the successful integration of the new museum web services to the central system.

RESULTS

The proposed service will offer to many museums the opportunity to communicate their content through larger channels (one stop shops) with minimum effort and cost. The service can be utilized and integrated in the portals of Organizations representing Museums (e.g. ECSITE European Network of Science Centers and Museums) to promote the exhibits of their members. The Web service infrastructures also provide the Museums with the possibility to use their own web services in many different ways (in their portals, in their local area information portals, municipalities, non governmental organizations (NGOs), tourist portals, etc). XML technology has the additional advantage that databases can be upgraded, changed or altered without affecting the overall functionality of the system.

Resulting from the proposed work is the architecture for:

- An online digital museum derived from the thematic combination of digital material from different digital collections.
- A search engine for users to retrieve information based on metadata schemes that will ensure semantic interoperability of the heterogeneous digital content. The engine is based on a comparative study of the related standards for digital libraries and Core Metadata (Dublin Core Working Group), Learning (SCORM) and audiovisual content description (MPEG7) as well as development of an appropriate interoperability framework for their usage.
- An application for museums allowing them to map their digital content on the metadata scheme and offer it as a web service. The application implements a robust model that will allow the use of semantic descriptions of digital content objects and/or segments in order to support the creation of reusable objects that may be used for the collection of personalized learning experiences.
- A Web service offering on demand digital content for the development of e-learning packages.

DISCUSSION AND FUTURE WORK

Future work on the presented service is not only possible and but also necessary. Many additional services can be developed in the described infrastructure that will be based on the shared material. Such services can be, for example, thematic explorations for school children (who will have the opportunity to see specific exhibits from several museums instead of visiting just one museum) combined tourism, multimedia (images, Videos, 3d presentations) presentation of related exhibits to accompany the original exhibits in the museum halls and so on.

Moreover, the architecture presented can integrate the benefits of the Open Grid Services Architecture (OGSA) and OGSA-DAI. OGSA describes an architecture for a service-oriented grid computing environment for business and scientific use. OGSA-DAI is a middleware to assist with access and integration of data from separate sources via the grid.

Grid technologies support the sharing and coordinated use of diverse resources in dynamic "virtual organizations" (VOs)—that is, the creation, from geographically and organizationally distributed components, of virtual computing systems that are sufficiently integrated to deliver desired Quality of Service (QoS) .

The Grid idea emerged initially as a model that combines the resources of many separate computers connected by a network (usually the internet) to solve large-scale computation problems.

Web services, as a new, interoperable and effective way to provide functionality over the internet, caused a new center of attention to the Grid technology, to Grids build on web services (GRID\WEB services).

Following this approach, a Grid can be defined as a layer of networked services that allow users single sign-on access to a distributed collection of computational, data and application resources.

The above lead to the development of The Open Grid Service Architecture that defines standard mechanisms for creating naming, and discovering transient Grid service instances; provides location transparency and multiple protocol bindings for service instances; and supports integration with underlying native platform facilities.

The Open Grid Services Architecture also defines, in terms of Web Services Description Language (WSDL) interfaces and associated conventions, mechanisms required for creating and composing sophisticated distributed systems, including lifetime management, change management, and notification. This architecture can be the next step for the proposed work.

Finally, the exploitation of museum material for learning purposes is an interesting idea that will continue to attract the attention of researchers, especially as technology is constantly evolving and offering additional opportunities.

REFERENCES

AMICO (2005): The Art Museum Image COnsortium (AMICO) http://www.amico.org (Accessed on February 10, 2007).

ArtsConnectEd (2005): http://www.artsconnected.org/ (Accessed on February 10, 2007).

Beckett D.; Miller E. and Brickley D. (2001): [DC-RDF-QUAL] Expressing Simple Dublin Core in RDF/XML. Dublin Core Metadata Initiative. http://dublincore.org/documents/2001/09/20/dcmes-xml (Accessed on August 2007).

Friesen N. (2004): Semantic & Syntactic Interoperability for Learning Object Metadata. In Hillman, D. (ed). Metadata in Practice. Chicago: ALA Editions. http://www.cancore.ca/semantic_and_syntactic_interoperability.html. (Accessed on September 2007)

Gudgin M., Hadley M., Mendelsohn N., Moreau J-J. and Frystyk Nielsen H. (2003): SOAP Version 1.2 Part1: Messaging Framework, W3C Recommendation, http://www.w3.org/TR/soap12-part1, (Accessed on February 10, 2007).

Hendler J., Berners-Lee T. and Miller E. (2002): Integrating Applications on the Semantic Web (english version). *Journal of the Institute of Electrical Engineers of Japan*, Vol 122(10), pp. 676-680.

ICOM Code of Ethics for Museums (2006): http://icom.museum/ethics.html. (Accessed on February 4, 2007).

ICOM-CIDOC (2007): The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC), http://www.willpowerinfo.myby.co.uk/cidoc (Accessed on February 10, 2007).

W3C Working Group Note (2004): Web Services Architecture. http://www.w3.org/TR/2004/NOTE-ws-rch-20040211/. (Accessed on February 10, 2007).

World Wide Web Consortium (W3C) (2006): eXtensible Markup Language (XML). http://www.w3.org/xml (Accessed on February 10, 2007).

AUTHOR BIOGRAPHIES

NIKOLAOS V. KARADIMAS was born in Athens, Greece and he graduated from Patra T.E.I. in 1995 with a Bachelor's degree in Electrical Engineering. He then received a Bachelor's degree with Honours in Electronic Engineering and a Masters degree in Computer Science from Glasgow Caledonian University, Scotland in 1997 and 1998, respectively. He also received a Masters degree in Distributed and Multimedia Information Systems from Heriot-Watt University, Scotland in 1999. Adittionally, he received his PhD degree in Engineeing from National Technical University of Athens in 2007. Furthermore, since 2001 he is teaching Informatics in Hellenic Army Academy and New York College, since 2003 he is teaching in T.E.I. of Chalkida and since 2005 he is teaching in Technical NCO Academy, as well. He is a member of the Greek Chamber of Engineers, member of IEEE and member of IEE. His research interests are in the fields of Databases, Optimization Techniques, Geographical Information Systems, Decision Support Systems and Multimedia.

NIKOLAOS P. PAPASTAMATIOU was born in Athens, Greece and he graduated from Aegean University, Samos in 1998 with a Bachelor's degree on Mathematics. He then received a Masters degree in Distributed and Multimedia Information Systems from Heriot-Watt University, Scotland in 1999. He holds also a vocational training certificate on project management from National and Kapodistrian University of Athens. From 2002 he is working for a software house as a web analyst/developer. He has followed several seminars on web technologies (XML Web services, Programming with XML in the .Net Framework, ASP.NET) and he has participated in many European and National research projects. His research interests are in the fields of Web Development, Databases, Web Services, Grid Computing, Natural Language Processing and Multimedia Applications. He is a member of the PMI.

VASSILI LOUMOS received the Diploma of Electrical Engineering and the PhD degree in Computer Science from the National Technical University of Athens (NTUA), Greece in 1975 and 1989. From 1975 to 1978 he was a research engineer at the Centre National de Recherhe Scientifique, Paris, France. From 1980 to 1982 he was with CGEE ALSHTOM as a field engineer. In 1990 he joined the faculty of the School of Electrical & Computer Engineering of the NTUA where he is presently a professor. He is teaching Multimedia Technologies and Computer Graphics. His research interests are in the fields of Multimedia, Computer Vision and Internet Navigation.