ETANA-DL: A Digital Library for Integrated Handling of Heterogeneous Archaeological Data

Unni Ravindranathan, Rao Shen, Marcos André Gonçalves, Weiguo Fan, and Edward A. Fox
Digital Library Research Laboratory
Virginia Tech
Blacksburg, VA 24061
{unni, rshen, mgoncalv, wfan, fox}@vt.edu

James W. Flanagan
Department of Religion
Case Western Reserve University
Cleveland, OH 44106
jwf2@case.edu

ABSTRACT
Archaeologists have to deal with vast quantities of information, generated both in the field and laboratory. That information is heterogeneous in nature, and different projects have their own systems to store and use it. This adds to the challenges regarding collaborative research between such projects as well as information retrieval for other more general purposes. This paper describes our approach towards creating ETANA-DL, a digital library (DL) to help manage these vast quantities of information and to provide various kinds of services. The 5S framework for modeling a DL gives us an edge in understanding this vast and complex information space, as well as in designing and prototyping a DL to satisfy information needs of archaeologists and other user communities.

Categories and Subject Descriptors
D.3.7 [**Information Systems**]: Information Storage and Retrieval – *Digital Libraries*

Keywords
Archaeology Digital Libraries, Open Digital Libraries, OAI, 5S Modeling.

1. INTRODUCTION
Archaeological research results in the production of vast amounts of heterogeneous data – field records, images, GIS records, VR models, etc. To handle these, most projects use some kind of information system. These are tailored to meet the individual requirements of projects, are monolithic in nature, and most importantly use different schemes to organize and store the data. This makes it very difficult for archaeologists to perform collaborative research because interoperability between such systems is limited. There is also a significant time lag between the production of the information and its dissemination, because of reliance on traditional forms of publication.

Our research tries to address the above issues. In this paper, we describe the design and implementation of a prototype digital library (ETANA-DL). We follow the approach advocated by Open Digital Libraries (ODL) [1] wherein a DL is constructed from components, thus leveraging all the benefits of component-based software development. We show that a DL can be very rapidly constructed, given a pool of components that provide the various services required, developing new components as necessary. We also use the metadata harvesting approach of the Open Archives Initiative [2] to tackle the problem of handling heterogeneous archaeological data.

2. ETANA-DL
ETANA-DL is a model-based, extensible, archaeological componentized DL that manages complex information sources using the client-server paradigm of the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [2]. This section describes the basic architecture of ETANA-DL, the services it provides, as well our efforts to model this DL using 5S [3].

2.1 Architecture
Figure 1 shows the integrated system architecture for ETANA-DL as it has been implemented [4].

2.1.1 Data Providers
We have converted portions of the information from an initial set of member archaeology sites (see Table 1) into Open Archives, exposing their data in a uniform metadata format, and then harvesting the records using OAI-PMH [2]. The number of sites and kinds of data are expanding.

Table 1: ETANA-DL Data Providers (current)

Site	Number of digital objects	Types of objects
Lahav [5]	550+	Figurine Images
Nimrin [6]	7800+	Field Records – Bones, Seeds
Umayri [7]	2100+	Field Records – Bones

2.1.2 Union Catalog
We harvest the records from three Data Providers into a central repository. The harvester is run on a periodic basis to ensure that changes made at the data source are reflected in our union catalog. Table 1 summarizes the contents of the union catalog and the types of objects harvested from the sites. Our union catalog utilizes a flexible XML-based schema design to maintain the integrity of the original data as well as provide the basis for even more integrated services.

2.1.3 Services
We use pre-existing ODL components to provide DL services such as searching, annotation, recommendation, etc. These ODL components communicate with each other using lightweight XML-based protocols (XOAI) [1]. We also developed other components for browsing, personal lists, workflow management,

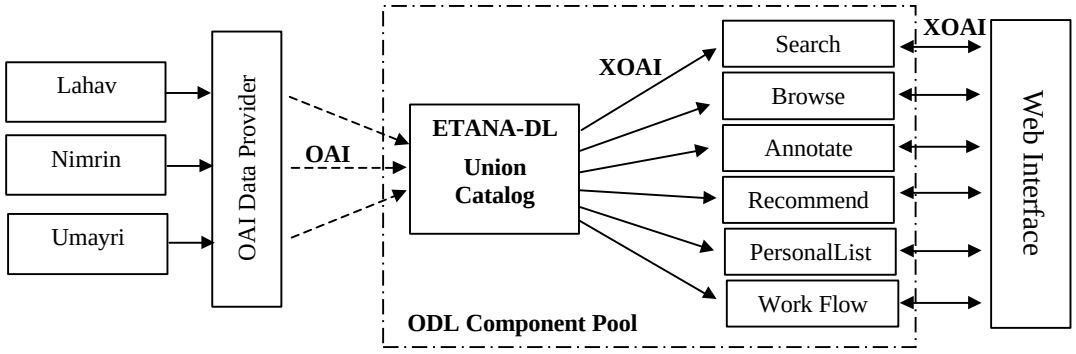

Figure 1: ETANA-DL Architecture

and recording of most recent queries issued. The web interface serves as the glue that binds these components together.

2.2 5S Modeling of Archaeological Systems

In order to solve the problems of heterogeneous data schemes and processing in archaeology, we are creating a unified (meta)-model for archaeological systems based on the 5S framework for information systems. 5S (Streams, Structures, Spaces, Scenarios, and Societies) has already been used to create meta-models for education-oriented digital libraries [3]. In the archaeological setting, streams represent the enormous amount of dynamic multi-media information gathered and created by specialists. Structures represent the ways archaeological information is organized along several dimensions, for example, site organization (e.g., region, site, locus, container), temporal organization (e.g., portion of the Early Bronze Age), and taxonomical organizations (e.g., pottery types and bone types). Spaces capture spatial and geographic distribution of found artifacts and 3D models of the past. Scenarios and Societies capture and model not only the services used by archaeological experts (e.g., artifact analysis and interpretation, visualization) and the general public (search, browse, annotate, recommend), but also hypotheses that correlate different types of historical and pre-historical societies, with the found objects, and with their spatial/temporal/physical properties.

So far, we have concentrated on the structural aspects regarding the archaeological sites and have used the resulting partially extended meta-model to allow integrated browsing and advanced searching across multiple collections/sites.

3. CONCLUSIONS AND FUTURE WORK

The speed (4 months) with which the ETANA-DL prototype [4] has been put together from various ODL components developed at Virginia Tech [8], and its success in providing basic DL services without compromising on performance, argue in favor of this approach to building DLs. ETANA-DL can be easily customized to meet particular requirements of various partner archaeology sites because of the loose coupling between components. Since our components also have proven useful in other DLs [9], we believe our approach to creating DLs handling heterogeneous data can be easily extended to other application areas.

Future work includes analyzing and ingesting data from additional sites, building a more complete 5S metamodel, developing more components/services for archaeology researchers and other user communities, and user interface research.

4. ACKNOWLEDGMENTS

This research is supported in part by National Science Foundation (NSF ITR-0325579). Marcos André Gonçalves is supported by a fellowship from AOL. We also thank Douglas Clark, Joanne Eustis, Paul Gherman, Larry Herr, Paul Jacobs, Douglas Knight, Oystein LaBianca, David McCreery, and Randall Younker for their support.

5. REFERENCES

[1] Hussein Suleman. *Open Digital Libraries*. Ph.D. thesis, Dept. of Computer Science, Virginia Tech, Blacksburg, VA, 2002. *http://scholar.lib.vt.edu/theses/available/etd-11222002-155624/.*

[2] OAI – Open Archives Initiative. *www.openarchives.org*.

[3] Marcos André Gonçalves, Edward. A. Fox, Layne T. Watson, Neill A. Kipp. *Streams, Structures, Spaces, Scenarios, Societies (5S): A Formal Model for Digital Libraries.* ACM Transactions in Information Systems (TOIS) (in press), 2004.

[4] ETANA-DL: Managing complex information applications: An archaeology digital library. *http://feathers.dlib.vt.edu.*

[5] Lahav Research Project. *http://www.cobb.msstate.edu/dig/lahav/.*

[6] Tell Nimrin Project Report. *http://www.cwru.edu/affil/nimrin/.*

[7] Madaba Plains Project – Tall al-'Umayri. *http://www.wwc.edu/academics/departments/theology/mpp/.*

[8] DL-in-a-box. *http://dlbox.nudl.org/.*

[9] CITIDEL. *www.citidel.org.*