Architecture Image Project

James Powell
Scholarly Communications/Library Automation
June 10, 1996

Introduction

In the Spring of 1996, a group of Library faculty met with photographic services (the imaging group: Annette Burr, Gail McMillan, Gary Worley) to develop a plan to support the digitization, archiving, identification and long-term storage of slide collections. This group decided that they would launch a pilot project during the summer. Photographic Services would scan 200 slides place them on a server in Scholarly Communications. SCP staff would design and implement a prototype workflow, storage, retrieval and annotation mechanism using a metadata scheme developed by this group.

This document describes a preliminary proposal for implementing the requested prototype. The system must be fully in place by late summer 1996 so that teaching faculty can use it to develop course materials for the fall.

• ***Implement new image delivery mechanism***
Images are currently dropped into an FTP archive on scholar2. This should be changed as soon as possible to an HTML file upload form. This will notify staff of a new image, allow us to archive it in a permanent location, assign it a URN, and add an entry to the Oracle Image Database.

• ***Assign URNs to each image upon delivery***
Use PURLs to uniquely identify images. A plain image URN points to the image itself, while a modified version of this URN points to the metadata record. For example, if an image is assigned the URN http://purl.lib.vt.edu/IMAGES/SLIDES/ARCH/0012060596.gif, the URN http://purl.lib.vt.edu/IMAGES/SLIDES/ARCH/0012060596.meta points to a PL/SQL query such as http://quasar.lib.vt.edu/cgi-bin/secwow/images.getimage?query=0012060596. A record containing the URN of an image should be added to the image database after submission.

• ***Refine metadata scheme and map to an Oracle relational database design***
There is a fairly good match between the metadata scheme designed by the group and OCLC's Dublin Core. It will be necessary to add fields to the scheme in order to record all of the data required. Here is a review of the elements allowed by the Dublin Core (from OCLC/NCSA Metadata Workshop Report at http://www.oclc.org:5046/oclc/research/conferences/metadata/dublin_core_report.html):

Dublin Core - *Element Description*

Subject: The topic addressed by the work
Title: The name of the object
Author: The person(s) primarily responsible for the intellectual content of the object
Publisher: The agent or agency responsible for making the object available
OtherAgent: The person(s), such as editors and transcribers, who have made other significant intellectual contributions to the work
Date: The date of publication
ObjectType: The genre of the object, such as novel, poem, or dictionary
Form: The data representation of the object, such as Postscript file or Windows executable file
Identifier: String or number used to uniquely identify the object
Relation: Relationship to other objects
Source: Objects, either print or electronic, from which this object is derived, if applicable
Language: Language of the intellectual content
Coverage: The spatial locations and temporal durations characteristic of the object

With one exception (language), each element can be readily mapped to the metadata requirements of the imaging group. The following table relates Dublin core elements to one or more fields required by the group - each element name is followed by the proposed Oracle field name and data type and the content description:

Element	Field Name	Oracle data type	Description
Subject	subject	varchar(240)	Keywords (eg impressionism, mothers, children)
Title	title	varchar(240)	Title/Subject of object pictured
Author	author	varchar(120)	Personal or corporate name
Publisher	publisher	varchar(120)	Archiver Agent (eg Scholarly Communications)
OtherAgent	otheragent	varchar(120)	Digitized by ... (eg Photographic Services, VPI)
Date	epub_date	varchar(24)	Date published online
ObjectType	objecttype		Type (form/genre)
Form	form_filetype		Electronic file type
Identifier	urn_id	varchar(240)	URN (handle or PURL) of image
Relation	relation		Collection membership
Source	source		Source: vendor, book, publication, donor
Language	lang		Language used: usually not applicable here
Coverage	cov_spat_size cov_spat_loc cov_spat_date		Object's spatial (size, location) and temporal (date, or ranges of dates) characteristics

Some fields required by the group are clearly related to elements in the Dublin Core, but are not present. These are listed below along with the elements to which they are related:

Other Proposed Meta-data Fields

Element	Field Name	Oracle data type	Description
Related to *subject* element	period	varchar(24)	Period/Century
Related to *subject* element	medium	varchar(24)	Medium/Materials
Related to *author* element	birthplace	varchar(24)	birthplace/nationality
Related to *form* element	access	varchar(24)	Access restrictions
Related to *ObjectType* element	img_view	varchar(24)	Image View
Related to *ObjectType* element	img_type	varchar(24)	Image Type

There will be additional fields in the Oracle table for maintenance and administration. Each administrative field will include an adm_ prefix. They should not be considered part of the metadata describing an object.

The Image Database
Metadata will be recorded in an Oracle relational database. The Image database will consist of three main tables and two linker tables. The object table describes the image. It includes the following fields: subject, title, publisher, otheragent, epub_date, objecttype, form_filetype, urn_id, relation, source, lang, cov_spat_size, cov_spat_loc, cov_spat_date, period, medium, access, img_view, img_type.

The collection table is identical to the object table except it lacks a relation field. In fact, few collections will require all the elements used to describe an object, but they will be available if needed. Finally the author table includes the following fields which describe the artist: author, birthplace, urn_id. The urn_id of a collection or author should be registered with the URN resolver as an SQL query that will return all of the images belonging to a collection or all images created by an artist.

A link table will associate an author to an object and an object to a collection.

• ***Implement a simple patron user interface and a library staff interface***
Most users will get at the image collections either using a URN they've previously recorded, or by searching the Oracle Image Database. A simple search interface will suffice for this pilot project and the implementation of such an interface is trivial. However, the library staff will require a more complex interface that will allow them to add collection and author records and modify object record for new images. One feature of the Dublin Core is that all elements are optional. Our implementation requires that at least one element, the identifier field of the object table, be mandatory. A URN must be assigned to an image in order for it to occur in the database. In addition, an image will require several other non-empty fields in the object, collection and author tables before it is accessible by patrons. The interfaces will be HTML forms generated by PL/SQL procedures accessed via the secwow package.

Conclusion

The proposed project can be completed in a matter of weeks if sufficient staff time can be made available. Much of the work will have to be completed by someone with Oracle and PL/SQL experience. The upload form and PURL registration can be handled by other staff. The resulting product will utilize several components of a digital library including metadata and URNs. This could be the first real product to come out of the Digital Libraries project. Currently, there is no reliable mechanism for archiving and retrieving digital images in the library. Scholarly Communications is amassing hundreds of images as they are digitized by Photographic Services but does not yet have a plan for making them available to the public. Many library units could ultimately benefit from this prototype system.