Classical Artifact Research Repository in DSpace: An LTS - Faculty - Student Partnership

Brandeis University
Speakers:
Lois Widmer, Library & Technology Services
Dai Wei, Library & Technology Services
Jenny Zimmer, ’08, Classical Studies & Anthropology
<table>
<thead>
<tr>
<th>Outline</th>
</tr>
</thead>
<tbody>
<tr>
<td>» Set the stage: Brandeis, CLARC, &amp; LTS</td>
</tr>
<tr>
<td>» Institutional Repository (IR) initiative -- marketing, policies, partnerships</td>
</tr>
<tr>
<td>» DSpace software and CLARC</td>
</tr>
<tr>
<td>» Lessons learned</td>
</tr>
<tr>
<td>» Future plans</td>
</tr>
<tr>
<td>» Q&amp;A and discussion</td>
</tr>
</tbody>
</table>
Introduction to Brandeis University

» Founded in 1948 as a liberal arts school
» Strong doctoral program
» Carnegie classification: Research Universities (very high research activity)
» About 5,000 students
» Located in Waltham, MA
Merged organization since January 2005 -- successful and still maturing.

Matrix organization:

Combines the advantages of the pure functional (traditional) structure and the product organizational structure.

Temporarily group together specialists from different units to work on special projects.

Environment of change requires that we be agile and proactive. We’re using more Project Management techniques and approaches.
Brandeis Institutional Repository

- Officially launched March 2008
- Platform: DSpace Open Source software
- 12 months to analyze business, functional & technical requirements, select software, migrate existing content
- Philosophy: IR belongs to the Brandeis community and is supported by LTS

Business, functional & technical requirements – 12 months may not seem very agile; however,

Had been early adopters of beta version of Ex Libris’s Digitool repository software under an IMLS grant to scan our collection of Daumier lithographs.

Were testing VFinity software for development of video content.

Thoughtful, detailed process.

IR belongs to the Brandeis community:

In process of forming an advisory committee that will represent the entire Brandeis community.

In DSpace, which is organized by community, sub-communities, and collections, we expect to minimize our role in collection administration. The LTS role should be support, training, and consultation.
The Department of Classical Studies is a small, but enthusiastic and intellectually active department, whose dedicated faculty encourages students to explore their interests in one of three tracks: Classics, Greek or Latin Literature, and Classical Archaeology and Ancient History. The Classical Artifact Research Center (CLARC) mostly falls under the latter track; however students from the other tracks have served as interns as well.
The CLARC collection is an eccentric one comprised of approximately 700 donated mostly utilitarian objects of unknown provenance, although research suggests that most of the collection is of ancient Mediterranean origin and dates between 1500 BCE and the Byzantine period (3-14th cent CE). Artifacts range from ceramics, bronze figurines, jewelry, weapons, medical instruments, and bronze fixtures. Such research is ongoing and vital to the collection's purpose as an undergraduate research collection. It is also an integral aspect of the Center's mission and vision of fostering undergraduate research of ancient life and culture, providing interns with hands-on collection management experience, caring for the artifacts, and the goal of creating a physical and digital space for students to make use of the collection in a meaningful and experiential way.

CLARC's story is, to an extent, probably familiar in that until two years ago it had been sitting in storage for over twenty years. The owners of the storage space decided they, justifiably, wanted the space for their own purposes and Classics had to figure out what to do with the collection. Two moves later it found a home in a small, basement room and the Classical Studies Internship was created to curate and research the collection.
The internship is competitive, with three interns per year. This coming semester will see the third group of interns. With the guidance of the department Chair, Classical Archaeologist Ann Olga Koloski-Ostrow, interns spend over 100 hours during the academic year working with the collection and documenting their decisions, thoughts, and tasks in journals.

My year as an intern, the academic year of '06-'07 was the first year of the internship and a time of organizing the collection, discussing goals, and public outreach on campus. As part of the organization and goal reaching process we decided that a digital database was imperative for in-house records, and we hoped to one day create an online presence. We worked closely with the Anthropology Department that year and after much discussion, we decided to be consistent across departments and use the same program, Filemaker Pro. Although we began discussion of fields and design, due to time and budget constraints – we didn't obtain the program, much less a computer, until late in the second semester – creation of the database did not get significantly underway until the past academic year.

My involvement continued in the form of an Independent Study as I worked on the database, policy, and was available to interns. One of the current interns will continue the continuity next year as he researches conservation of the collection as part of an awarded fellowship.

Fields - Loosely based on Anthro Dep.; Beazley, CHIN read various books, consulted archaeologists

Developed into what we needed/wanted
The CLARC – LTS Partnership

» A collaboration among faculty, student interns, and LTS

- Students interns -- key decision-makers
- Faculty relies on their input

Faculty member is Chair of Classical Studies and head of the CLARC internship program. Relies on students’ judgment and requires their full participation in discussions with LTS and in all decisions related to the CLARC collection.
» Convergence:

- LTS seeking early adopters for the IR

- CLARC seeking a way to provide broader access, especially to Brandeis students

LTS seeking early adopters:

In seeking early adopters, initial content for the IR, we also want demonstration collections for further marketing.

These images were of special interest because they’re previously unpublished digital surrogates of objects that are not readily accessible to the wider campus community or to the public, except when they are in special displays. It will be an evolving, growing collection of images with accompanying research.

No copyright or publisher issues to address!

Given our success with this collection, we’re now looking to work with the Brandeis Rose Art Museum, Anthropology Dept., and other programs with cultural objects.
The CLARC – LTS Partnership

» LTS Recommendations versus CLARC Reality

Compliance with best practices & standards

- Image format: TIFF versus JPEG
- Equipment: multiple cameras
- Metadata: Getty AA&T versus CHIN

LTS is interested in adhering as closely as possible to existing standards and best practices such as the Visual Resources Association’s metadata element set for cultural heritage objects and their Cataloging Cultural Objects guide to describing cultural works and their images; however, CLARC had already created their FileMaker Pro database, before LTS became fully involved with the project. Expecting them to retrofit to our standards and redo considerable work was not reasonable or practical.

--We recommended TIFF format for preservation standards - they had already settled on jpeg.
--We preferred a single camera, possibly a single photographer.
--We thought that the Getty thesauri (Art & Architecture Thesaurus and Geographic Names Thesaurus) would be suitable.

We compromised:
JPEG because the work was already in progress in this format.
Multiple cameras provided by the students because this is what’s available.
Canadian Heritage Information Network (CHIN) as the source for descriptive metadata. As it turns out, CHIN is based on the Getty Art and Architecture thesaurus, supplemented with French language equivalents.
Example compiled by Jenny to demonstrate the impact of 3 separate cameras: Canon, Samsung, Panasonic.
In keeping with the philosophy that the Brandeis community owns the IR and individual departments take responsibility for collections/communities, the CLARC faculty adviser is now the DSpace administrator for the CLARC collection. For technological reasons and time management, we’re discussing the possibility of adding one of the student interns or the student manager as a co-administrator each academic year. Of course, this will require training on the DSpace software each year, good documentation, and LTS support.
Now I will talk about some technological aspects of this project, such as:

What are the determining factors in selecting DSpace as our institutional repository to hold CLARC and other digital collections.

What is the process to export FileMaker Pro records to DSpace, including mapping locally developed metadata elements in FMP to standard qualified Dublin Core elements, and how we use XML files as a tool to transport data from FMP to DSpace.

I will also talk about how we customized the DSpace interface for the CLARC collection and what are the enhancements we need.
Why DSpace?

» Library-focused standards for metadata, intercommunication, federation, and harvesting: Dublin Core, OAI, SRU

» Open source, large and vibrant community of shared users in the academic library space.

» Flexible user interfaces (Manakin)

» Easy integration with other library and academic tools

» Library focused standards for preservation

So why we selected DSpace as our institutional repository. Here are some determining factors:

1. We're concerned with library-focused standards for metadata, intercommunication, federation, and harvesting. By default, DSpace is optimized for the qualified Dublin Core, and the records in DSpace can easily be harvested by using the open archives initiative protocol for metadata harvesting. DSpace is also integrated with SRU (Search/retrieval via URL), a standard xml-focused search protocol by the Library of Congress to replace the library standard Z39.50 protocol. These protocols will enable us to share our metadata and links globally.

2. Secondly, DSpace is open source, and has a large and vibrant community of shared users in the academic library space. Currently, there are well over 300 production DSpace sites at major universities in North America and worldwide who share our needs and provide each other with help and support tools.

3. Next, DSpace has flexible user interfaces, especially with its latest version, 1.5, which allows communities and collections to establish a unique look and feel distinct from the default installation of DSpace.

4. DSpace can also be integrated easily with other library and academic tools, because it uses standard, open, and documented APIs for interaction and intercommunication. For example, it will work with Metalib, our Federated Search Engine, as well as other tools.

5. We are also concern with library focused standards for preservation. DSpace is integrated with the Handle System (the corporation for national research initiatives) to assign and resolve persistent identifiers to objects, and it generates and tests checksums for ingested objects.
Next I will talk about how we map the local FMP elements to the standard Dublin Core elements. First I will open a screen shot of our FileMaker Pro database to show you a original CLARC record. As you can see this record includes locally developed descriptive metadata fields, such as catalog number, material, region and storage. To map these local fields to the standard qualified Dublin Core elements, we referred to the existing metadata standards crosswalks already done by others as on the next slide.
We found the metadata standards crosswalks on Getty’s web site was very helpful. (open the web page). As you can see, this page includes crosswalks between Dublin Core and many other metadata standards, such as MARC, VRA and CDWA.

Although the metadata set in FMP doesn’t follow any metadata standards on this page, some of the local elements can be found in one of these existing standards. For example, the local element material is also a basic element in the CCO standard for cataloging cultural objects. By checking how the element in CCO are mapped to the dublin core, we emulated it in our mapping.

The other three web resources we referred to are the CCO web site, the VRA core 4.0 by the visual resources association data standards committee and CDWA metadata standards. These resources helped us to better understand the requirements of describing CLARC collection images and how to map the local elements to Dublin Core.
In addition to doing research in existing metadata standards, we also met with the CLARC department to clarify the meaning of some local elements before we finalized our crosswalk. Let me show you a screen shot of our final crosswalk with details (open Excel file). As you can see the first column has the local elements, and the second column shows what DC elements they map to.

Some of the FileMaker Pro elements don’t have exact matches in qualified Dublin core, and we mapped them to the closest DC fields, for example:

Culture is mapped to subject, and region is mapped to coverage spatial.
Using XML files to transport data

Why xml?

» Designed to transport data

» Endorsed W3C standard and interpreted with a wide variety of tools

» Complements Java

After the mapping work, we started to transport the records. We decided to use XML files as a transportation tool, because XML was designed to transport and store data, although not to display data. XML is the endorsed industry standard of the World Wide Web Consortium (W3C) and is supported by all leading software providers. Both DSpace and FileMaker Pro can import and export records in XML formats. Also XML complements Java very well which is the programming languages used in DSpace.
Here is a screen shot of the XML file exported from the FMP. Let me open the file and show you more details.

This file includes information about the export itself, such as from which FMP instance the records are exported and if the export was successful. It also shows what elements are exported for each record. As you can see, basically all the descriptive metadata, such as title and culture, were exported from the FMP and included in this XML file.
Using XML files to transport data
Transform records to DC by using XSL

To convert the FMP XML file to a standard Dublin Core xml file, we wrote a xsl file. Here is the XSL file. Basically, each record in the FMP xml file is mapped to a “dublin_core” record. We also added some functions in the XSL file to capitalize the first letter in the title field and to normalize spaces in the records, for example removing extra line breaks from the description field.
Using XML files to transport data
DC XML file for DSpace

This is the final dublin core record converted from the File Maker Pro record before we loaded it to DSpace database. One element I want to mention here is the field to hold the dimension information. In the original FMP record, it displays the objects dimensions in separate fields. When we did the conversion, we merged those fields to two qualified dc format elements, one showing centimeters, the other inches.
This shows a full CLARC record in DSpace. (open the web page) In addition to the elements from FMP, there are also few elements added by DSpace during import, such as the identifier.uri field which serves as this item’s permanent URL.
We customized the DSpace interface for the CLARC collection. For example, the default theme shows title and author fields. But since author fields don’t apply to the Clarc collection, you can see “Unknown author” is displayed for all the records. We replaced the author field with type and material fields which are relevant to this collection.
We also updated the CLARC collection home page on DSpace. We added a picture and also a description on this collection, including links to the Classical Studies department and CLARC.
DSpace Enhancements

There are still things in DSpace that need to improve. Let me show recent enhancements applied in DSpace 1.5 on our test instance so far. (open the page) In the previous versions, by default, DSpace only allows users to browse by title, author, subject and date. We added a new browsing by medium function here that is very useful for the CLARC collection.
Handling of multipart objects with hierarchical structure (sample)

Image presentation and manipulation
– Jpeg2000

Currently DSpace supports organizational structure, such as communities, sub-communities and collections, but it doesn’t support the content structure and there is no multipart object viewer. For example, we have a record in CLARC collection that describes three similar seals and each seal has three images. Ideally the images of these three seals should be displayed in a multipart object viewer like this in a hierarchical structure so that end users can easily compare them in one page.

We also need a tool, for example an integrated JPEG2000 server, to efficiently deliver large image files, and to allow users to zoom and manipulate images in browser. Currently it is not easy to integrate this kind of tool in DSpace.
Having CLARC work directly in DSpace

» Web based – more access points

» Avoid the need for a crosswalk

» Single system and database for maintenance

We are also considering having CLARC work directly in DSpace. The advantages of doing it are:

First DSpace is web based and has more access points. Currently the student interns have to use the machine in CLARC, which has FMP installed to enter or edit a record. By working directly in DSpace, they can do it where they have internet access.

Second, it would avoid the need for a crosswalk but would still need to modify the qualified Dublin Core to accommodate concepts.

Also, it will be a single system and database for maintenance and will be done by the university network system department. CLARC doesn’t need to do backup any more.
Lessons Learned: Working with Customers

» Communications

» Documentation

» Decision-making

» Flexibility

» Project Management techniques

It’s challenging to work with a group of customers, who in some cases have differing views, need to make internal decisions and reach consensus before the project can advance. In the case of student interns, the repository work was one more obligation among the many commitments.

**Decisions** – be clear on who the decision-makers are!

When possible, we need to get in on the ground floor of projects in order to help guide key decision-making that will facilitate the partnership and assure technological compatibility with minimal extra work.

Document everything, including how you reached your decisions -- it will be an enormous aid to you, your customers and future staff. We followed up each meeting with detailed minutes, action items, and decisions, which were distributed to everyone involved or affected by the project, all of whom could provide feedback.

**Flexibility**

Standards are ideals -- accept that it isn't always possible to apply them or meet them. However, in making decisions, be clear about why you’re deviating from established standards and document your decisions.

**Project Management:** PM typically defines scope, duration and budget for any given project. Define and work within the constraints of scope, quality, cost, time, and resources. Beware of scope creep and feature creep. Document the project scope and develop a project plan - can always renegotiate and re-evaluate. Stay realistic and focused, consider what you can accomplish within current time
Next Steps & Future Plans

» Document thoroughly for knowledge continuity and transfer
  • Publish guidelines and best practices
» Meet early with the new interns to establish schedule, best practices and workflows
» Set plan and deadlines for the next AY
» Apply our lessons learned, both to the CLARC collection and to others
Next Steps and Future Plans

» Build LTS Metadata team

» Assessment
• How do we measure success for the IR?
• Is distributed collection administration viable, scalable?

» Enhancements:
• Thesauri, authority control & controlled vocabularies
• 3D visualization of objects
• More content!

Metadata team:
½ training session for 12 LTS staff members, then hands on work with the Daumier lithograph collection (our legacy, first IR collection).
Further training through OCLC programs.

Collection administrators:
It’s too soon to know the viability/scalability of this work model, although at this point we have 6 different collection administrators, 4 in active collections, the other 2 working on collections that are in development.
Lessons Learned: Student Perspective

» Communication and Continuity

» Standardization and Metadata
  • The earlier and more thoroughly documented, the better
Next Steps & Future Plans cont.

» CLARC’s physical collection
- Continue research, photography, and conservation
- Update data
- Increasing accessibility
- Improving quality control

» On-line presence
- Would like ability to browse by material and type, culture, time period and region.
- Incorporate exhibits
- Present student research as available

For CLARC
Agree upon data dictionary
Agree upon metadata/standards
Continue the photographing, measuring, and research of objects
build a library of research related to the collection
engage in conservation efforts on select objects and refurbish storage
increase accessibility and aid research with a digital catalog
provide appointment-based sessions where students and interested parties can have supervised access to the Center and collection for research
engage the Brandeis campus and public in outreach programs, fostering interest in ancient life and culture, classical archaeology, and fields such as metallurgy

For collaboration/online presence
Eventual ability to browse/search by type, culture, time period, and region
Incorporate exhibits (perhaps online exhibits on CLARC website eventually)
Increased communication/more regular meetings
Continue data entry
Rewrite many of the physical descriptions
Proofread entries for copied mistakes, transcription errors, typos, and inaccuracies
Provide information online related to collection
Update CLARC website
Continue to update location of objects as exhibits change
Provide student research online with consent as it comes up
Redo some photos (blurry, additional angles, incorporate color bar)
Improve quality control with photos (and everything)
Record exhibits (in-house at least)
Eventual 3D Blog
Acknowledgements

» CLARC Photography

- Jenny Burgman
- Kevin Goscinak
- Alex Smith
- Jenny Zimmer
- Mike Lovett, University Photographer
- Ann Olga Koloski-Ostrow, Faculty Adviser
Questions and Discussion?

http://dcoll.brandeis.edu

© 2008 Brandeis University
All Rights Reserved