DCPAPERS

University of Washington Early Buddhist Manuscripts Project in DSpace

Kathleen Forsythe, Alan Grosenheider, Eileen Llona, Jennifer Ward
University of Washington Libraries, Seattle, USA
{forsythe, alang, ellona, jlward1}@u.washington.edu

Abstract

Preliminary findings on the deployment of metadata for a project using DSpace at the University of Washington Libraries are presented. Problems encountered include mapping user-provided metadata for a domain-specific image collection to Dublin Core. Creating collection-specific qualifiers and adding value prefixes were tested in the DSpace configuration. Data inconsistency caused some technical problems, as well as bringing to light other issues involved in deploying a system with user-supplied content.
Keywords: *Dublin Core, DSpace, University of Washington Libraries, Early Buddhist Manuscripts Project*

1. Introduction

The Early Buddhist Manuscripts Project (EBMP) represents ongoing research in a field where researchers, as well as artifacts, are spread around the world. The collection comprises digital images of the ancient manuscript fragments and researcher-applied metadata. The images provide the only viable means for studying fragile scrolls. Given the available infrastucture at UW Libraries, we chose DSpace as the trial storage technology.

Challenges and opportunities presented themselves not only in systems implementation in DSpace, but in various metadata issues. This project has a set of metadata that describes both the digital image and the contents represented. Since our choice of technology required using qualified Dublin Core, challenges of preserving the context of EBMP's metadata have been made apparent in the initial phases of loading the project's files. In addition, EBMP content will change and perhaps transform as research continues. This evolutionary quality is new to project implementers. Other digital collections have been more fixed in scope, basically just adding more of the same kind of object. Here metadata and object relationships may alter as more is learned about the fragments. The architecture of DSpace allows clusters of bitstreams for storage of images and data, and presents some possibilities of archiving metadata snapshots if this is deemed important as the project develops.

Another metadata challenge lies with audience scope. Access is currently limited to a small set of researchers, but will someday be expanded to the public at large. What methods should be employed to encourage good quality metadata that will not only serve the immediate need of researchers, but will be consistent and complete enough to offer an effective interface for cross disciplinary inquiry in the future?

2. Background

The British Library / University of Washington Early Buddhist Manuscripts Project was founded in September 1996 in order to promote the study, editing, and publication of a unique collection of fifty-seven fragments of Buddhist manuscripts on birch bark scrolls, written in the Kharosthi script and the Gandhari (Prakrit) language that were acquired by the British Library in 1994. The manuscripts date from, most likely, the first century A.D., and as such are the oldest surviving Buddhist texts. They promise to provide unprecedented insights into the early history of Buddhism in north India and in central and east Asia [1]. Due to the delicate nature of these ancient objects, digital images have been taken of them, in order to preserve the content which they contain and to reduce the deterioration of the remaining fragments. These images include both fully uncompressed, archival quality TIFF formats, as well as compressed JPG formats for use by the researchers. Manipulations of each image may also be included, representing stages in the interpretation of the images and their contents. The collection also includes digital images of older published photographs of now unavailable fragments.

The UW Libraries was approached by one of the EBMP researchers, requesting that the library provide access and long-term storage for the digital images of the fragments. This required the involvement of the Systems department at UW Libraries, so that the project would be supported within the current infrastructure. Surveying the software tools available to do the job, it was noted that the Libraries is a member of the DSpace Federation and test collections were needed. Library Systems is incorporating DSpace into the Libraries' infrastructure in response to the current goal of providing a trusted digital repository and preservation space. It was decided to test the technology with the EBMP project to learn more about its capabilities. DSpace allows uncompressed storage for large, growing collections and a bitstream architecture permitting the linking of several objects to one set of metadata, both desirable features for this project. In addition, a future implementation of the software will allow for user input through a Web form. The other possible platform currently available to the Libraries, CONTENTdm, requires client installation and support for the inputting interface.

https://doi.org/10.23106/dcmi.952107609

3. Procedure and analysis

The EBMP researcher provided metadata describing image content, the photograph, the digitization process, and publication information for cases where the digital image is of a previously published photograph.

3.1. Importing data into DSpace

The current implementation of DSpace, "out of the box", requires use of Dublin Core elements, including some qualified elements as defined by the Dublin Core Libraries Working Group Application Profile (LAP) [2]. Fitting EBMP's original 37 fields, some of which convey museum specimen attributes, into a flat 15 element Dublin Core structure presented a huge challenge. Even with the qualifiers allowed by the current implementation of DSpace, the context of element values were lost when assigned to generic elements such as "description.other". Thus, we investigated other solutions for accommodating EBMP metadata.

The first method involved defining our own qualifiers to use with Dublin Core elements. These included qualifiers for the Identifier element, such as Frame number and Fragment Number; and qualifiers for the Creator element, such as Photographer, Digital Editor, Discoverer, Scribe. Using DSpace's batch import process to test these qualifiers, however, resulted in load failures, indicating that current software parameters would not allow locally configured metadata elements. Since DSpace is a very new technology undergoing constant development, further testing is needed to determine its capabilities for handling alternate metadata schemes. In order to move the project ahead to accommodate user needs and complete testing, we used the Dublin Core elements and qualifiers as documented by DSpace, and added prefixes to data in various elements to more clearly indicate the values. Here's an example of the display of one element:

Before adding prefixes:
Description: 1
1
J1.1

After adding prefixes:
Description: Fragment Number: 1
Frame Number: 1
Other Number: J1.1

The current metadata architecture supported by DSpace does not accommodate the 1:1 issue well. It is not possible to differentiate which data describe the content of the image as opposed to the image itself vs. its digital manifestation. This in combination with the generic element tags does not present a clear description of objects. Projects from other institutions in DSpace reviewed before our implementation were all text-based and better accommodated by the current DSpace configuration.

3.2. Metadata quality

Other issues discovered in working with the project included quality of user-provided metadata. Consistency in the data was noted to be lacking, especially in fields referring to personal names and dates. Currier and Barton [3] identify research questions pertaining to user-supplied metadata, one of which is the quality for immediate or domain-specific purposes vs. fuller and higher quality to support maximum resource discoverability by a range of searchers. This also became an apparent issue in the EBMP metadata, as those who were outside the "domain" and involved in the Dublin Core mappings tried to ascertain the meaning of some of the EBMP elements. Hopefully, metadata will become more complete as research progresses and more is learned about the scrolls.

4. Conclusion

This project so far has demonstrated the inadequacies of Dublin Core for a highly specialized non-textual collection. However, given other benefits that DSpace offers, our temporary solution of including value prefixes allows the images to be stored safely, to have the multiple versions of the image files linked, and to allow some way for the EBMP research community to access their images. Future testing will include deployment of METS or another appropriate schema within DSpace. METS will enable a richer element set by bringing the capability to reference other metadata schema as well as better representing the relationships among levels. Other element sets such as VRA Core Categories and RLG's REACH will be explored, and/or creating an institution-based application profile that can be used with Dublin Core. Enriching current data with new elements, especially preservation metadata, is under discussion. In terms of improving the quality of metadata, templates encouraging minimal levels and consistency of data will be explored. We will also be looking at OCLC's Web services for authority control as an aid for mitigating inconsistencies and other problems of user-created metadata.

References

[1] *Early Buddhist Manuscripts Project*. Retrieved July 15, 2003, from http://depts.washington.edu/ebmp/.
[2] *FAQ: DSpace at MIT*. Retrieved July 15, 2003, from http://dspace.org/what/faq.html.
[3] Currier, S. & Barton, J. (2003). *Quality Assurance for Digital Learning Object Repositories: How Should Metadata be Created?* Manuscript in publication.

https://doi.org/10.23106/dcmi.952107609