

Adding a User Developed Vocabulary in DSpace

Serhiy Polyakov (sp0055@gmail.edu), Hong Xu (hx0008@unt.edu)

Interdisciplinary Ph.D. Program in Information Science, University of North Texas, Denton, TX

Advisor: Dr. William E. Moen, Texas Center for Digital Knowledge



Introduction

DSpace is a widely adopted open source platform for digital repositories. It provides a self-deposit environment for users and it allows users to supply relating metadata for the items. Some of this metadata may come from vocabularies. DSpace has capabilities of configuring pre-controlled vocabularies. However, these vocabularies cannot be enriched with the necessary terms during the process of submitting items to the repository. This poster describes the method of developing the post-controlled vocabulary and connecting it to DSpace, and the test results of the application.

Methodology

The current released version DSpace 1.4.2 supports controlled vocabularies to limit the set of keywords that users can use while describing items. There are two default vocabularies available with DSpace: The Norwegian Science Index (NSI), and Swedish Research Subject Categories (SRSC). The DSpace adopters can also create their own vocabularies by customizing the vocabulary xml files. But these kind of controlled vocabularies are pre-controlled. These vocabularies do not support development and enrichment during the process of metadata creation. Any changes to the vocabulary xml files are picked up by DSpace only after restart of Tomcat (Java Server Pages container).

The Texas Center for Digital Knowledge (TxCDK) at the University of North Texas (UNT) received a grant from the Texas Higher Education Coordinating Board (THECB) to develop a proof-of-concept application of a learning object repository (LOR). The goal of the awarded project was to demonstrate a proof-of-concept learning object repository's capability to leverage the course content developed through the Texas Course Redesign Project for reuse and repurposing.

Repository contains *U.S. History I* course which consists of units, each of which consists of multiple lessons, which in turn include multiple topics (Figure 1). The components of the courseware can be reused and repurposed in multiple instructional delivery models including online, blended, and face-to-face.

During the process of repository building, based on the user's needs, the team of THECB LOR decided to develop a post-controlled vocabulary which can be developed and enriched when users self-deposit learning objects. This vocabulary is stand alone application that links to DSpace metadata entry interface.

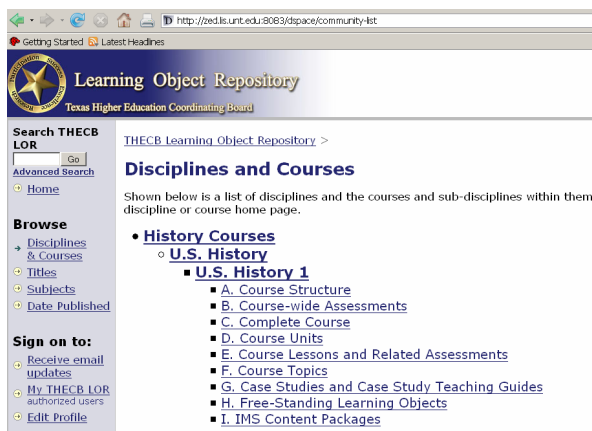


Figure 1. Learning objects repository

Developed vocabulary tool provides the following functions:

- Provide an online version of a controlled vocabulary that metadata creators can link to and have open when they are determining what terms to put into the Subject field;
- Let metadata creator click on a term in the online controlled vocabulary and have it automatically pasted in to the Subject field;
- Allow the metadata creator to add new terms to the controlled vocabulary through a simple to use submission form;
- Make newly submitted terms immediately available for all users working on items submissions.

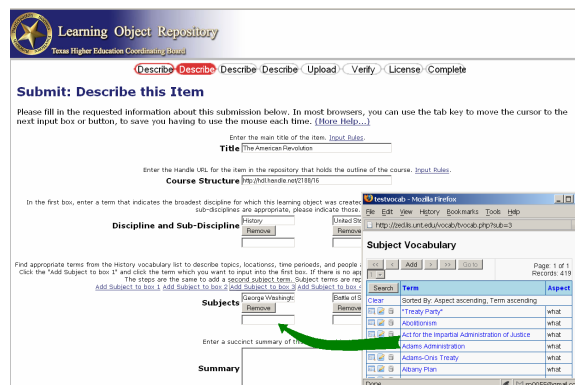


Figure 2. Metadata entry page of the repository and vocabulary

The vocabulary was developed using open source database management system MySQL and PHP scripting language as well as Java Script.

When a user submits an item to the repository he or she needs to assign subject terms. Clicking on the link *Add Subject to box #* brings subject vocabulary application window (Figure 2). A user can search this vocabulary for the terms and add terms to the Subject box by clicking on the term.

A user can add new term by clicking on the *Add* button if appropriate term is not found in the vocabulary. Entry form dialog will come up (Figure 3) and after a user adds and saves the entry, new term will be immediately available for use. This is especially important for the environment where more than one user is working on the uploading items to the repository.

The vocabulary was tested while submitting hundreds of items to the repository by the group of six people working concurrently from different locations. Application proved to be effective and useful tool.

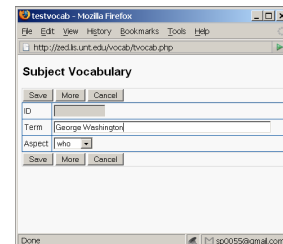


Figure 3. Interface for adding new terms to the vocabulary

Conclusion

DSpace is highly customizable open source software package for digital repositories. There are many add-ons and tools available in DSpace for building repositories that meet specific requirements. However, in some instances standalone applications are necessary to add more features. Standalone subject vocabulary that was described in this poster is an example of such application.