

Digitization in the Real World

**Lessons Learned from
Small and Medium-Sized
Digitization Projects**

Edited by

Kwong Bor Ng & Jason Kucsma



Metropolitan New York Library Council

Published in the United States of America by
Metropolitan New York Library Council
57 East 11th Street, 4th floor
New York, NY 10003-4605
p: (212) 228-2320 f: (212) 228-2598
Web site: <http://www.metro.org>

ISBN: 978-0-615-379998-2

Cover Design: Jason Kucsma (*illustration by Smartone Design,
licensed via iStockphoto.com*)

Reviewers Committee: Mark F. Anderson, Jill Annitto, Anna Craft, Jody DeRidder, Renate Evers, Wei Fang, Maureen M. Knapp, Sue Kunda, Mandy Mastrovita, Ken Middleton, Emily Pfotenhauer, Mark Phillipson, Alice Platt, Mary Z. Rose, Stacy Schiff, Jennifer Weintraub, Andrew Weiss.

Copyright © 2010 by Metropolitan New York Library Council. No part of this book may be reproduced in any form or by any means, electronic or mechanical, including photocopying, without written permission from the publisher.

The views expressed in this book are those of the authors, but not necessarily those of the publisher.

Using Omeka to Build Digital Collections: The METRO Case Study

Jason Kucsma (Metropolitan New York Library Council)
Kevin Reiss and Angela Sidman (City University of New York)

Abstract

In September 2008, the Metropolitan New York Library Council (METRO) began work building a directory of digital collections created and maintained by libraries in the metropolitan New York City area. METRO built the directory using Omeka, an open source collection management system, as a test to determine the viability of this platform for member libraries interested in using Omeka to build and deliver their own collections. This paper addresses Omeka's strengths and weaknesses as a software platform for creating and managing digital collections on the web. The analysis includes an examination of original record creation, and the extensibility of the system through the use of plug-ins.

Keywords: Content management system, ContentDM, Digital asset management system, Digital directory, New York City, Digital services, Digital workflow, Metadata management, Non-profit organization, Omeka, Open source software, Plug-in development, User experience, WordPress.

Project Background

The digitalMETRO project was born from the Metropolitan New York Library Council's (METRO) 2007-2010 Digital Library Services Plan. The plan serves as a guiding document for METRO's work in helping member libraries build and maintain unique digital collections and provide access to them online. Since 2005, METRO has provided over \$300,000 in digitization grants to fund over 30 digitization projects. METRO also provides a diverse curriculum of workshops focusing on digital conversion, metadata for digital collections, digital collection management software, and related areas of emerging technologies for libraries. In addition to ensuring these programs continue to provide training and resources to member libraries, the Digital Library Services Plan also recommends the creation of a directory of digital collections in the METRO membership.

With the charge of creating a directory of digital collections, Jason Kucsma, METRO's Emerging Technologies Manager, served as Project Manager and assembled a small project team to begin working on the directory in October 2008. He recruited Angela Sidman, then a Metadata Librarian at the Mina Rees Library at City University of New York's Graduate Center, to serve as the Metadata Librarian for the project. Kevin Reiss, then the Systems Librarian at the Mina Rees Library, was recruited as the project's Web Developer.

Choosing Omeka

One of the first decisions we had to make was which collection management system we would use to build and deliver the directory on the web. We considered three different options. METRO licenses a hosted CONTENTdm (CDM) instance from OCLC, so that was first considered as an option. Sidman and Reiss had worked extensively with CDM to build the METRO-funded Digital Murray Hill Project. While CDM is a system that supports collections with robust metadata and that contain a wide array of digital formats it lacks some characteristics that are desirable in a modern web digital exhibition tool. The out-of-the-box CDM user interface makes it difficult to browse collections by subject and set, and CDM does not support any

of the interactive features that many users expect in a web interface such as tagging, social networking, and a mechanism to accept end-user feedback on the web. Though CDM does provide a basic Application Programming Interface (API) that allows for some modification of default CDM behavior, we felt that this API was an inadequate tool for building the types of features we wanted to include in the digitalMETRO website.

Our interest in deploying a feature-rich digital exhibition tool next led us to consider Omeka, a relatively new open source collection management system that was created by the Center for History and New Media (CHNM) at George Mason University. The Omeka software was the closest option to another open-source web content management system, WordPress, that we considered using for our project. WordPress has well-documented theme mechanisms for customizing the display of WordPress content, and an expanding pool of plug-ins that be used to modify content behavior. Unfortunately, WordPress does not have a well-developed mechanism for supporting the types of collection-building workflows and metadata-creation common to archives and libraries, and creating a plug-in to support these activities was beyond the modest scope of this project. Our exploration of potential content management systems overlapped with the Fall 2008 release of the 0.10 beta version of Omeka, and the growing number of projects using the system in beta made it worth considering.

Additionally Omeka's developers appear to have taken design inspiration from the highly successful general purpose open-source content management system Wordpress. Wordpress is widely known for it's ease of install and high-level of functionality. In this vein, Omeka developers have touted the platform as a "next generation web publishing platform for museums, historical societies, scholars, enthusiasts, and educators." Our project team thought libraries might also fit into that family, particularly smaller libraries with limited technical staff or financial resources to build and deliver digital collections online. The simplicity of installing and configuring the Omeka system rivaled WordPress's ease-of-use, leading CHNM Director Dan Cohen to suggest Omeka is "Wordpress for your

exhibitions and collections.” Omeka also utilizes the same theme and plug-in mechanisms Wordpress utilizes to provide omeka users a means to customize and create new system behavior.

According to the CHNM website: "Omeka provides cultural institutions and individuals with easy-to-use software for publishing collections and creating attractive, standards-based, interoperable online exhibits. Free and open-source, Omeka is designed to satisfy the needs of institutions that lack technical staffs and large budgets."

Omeka was also an attractive choice because the strong and flexible approach to metadata representation built into the software. Additionally, Omeka Developer Dave Lester affirmed an advantage Omeka has for use in libraries and archives over a general-purpose CMS like WordPress on his blog, Finding America, "WordPress doesn't use structured metadata the way that scholars, libraries, and archives do. We have controlled vocabularies, 50 ways of classifying the same thing, and need a system that allows us to easily do that."

While Omeka does not currently support controlled vocabularies in the sense that catalogers use to maintain authority control for records, comments like this marked what we felt was a shift within the Omeka developer community from a focus on archives and museums to the digital collection building pursued by libraries. This, combined with the fact that Omeka appeared to be quickly developing a community of users and developers actively creating new features and posting improvements to the software, reinforced our decision to adopt Omeka for the digitalMETRO platform. Also, this relatively low-risk project would allow us to test the viability of Omeka as a collection building and exhibition platform that could be recommended to METRO members and other small and medium libraries and archives.

The Omeka theme and user experience

Choosing Omeka as digitalMETRO's platform allowed us to present digital collections in an interface that provided the features which users expect from a modern, interactive website. Omeka users can easily customize the display and behavior of their websites using a

number of pre-packaged Omeka themes or by manipulating the well-designed basic CSS that controls the basic look and feel of an Omeka theme. Looking at the existing digitalMETRO site (which utilizes the Omeka “Winter” theme) for illustrative purposes, we can explore some of the more useful standard Omeka features. The “Winter” theme is one of 11 out-of-the-box themes available, and the system affords a good deal of flexibility through CSS customization.



The screenshot above displays the main page of our Omeka instance. Visitors to the site are given a number of options which they can use to begin exploring the site. Casual browsers may be drawn to the "Featured Collection," which is randomly generated by Omeka each time the home page is refreshed. Site administrators can determine which items will be included in this rotating cast by simply checking a box in the item record. The site can also be set to feature "Recent Added" items to a collection, and the number of items can be set according to a specific collection's needs. Users may do a simple search using the "Search Box" in the top right corner of the page, or they can conduct an "Advanced Search" using the link just below the

simple search box. While the current search tools offer expected, though limited access to the collection, Omeka 1.1 promises to feature a more robust search tool.

Omeka generates both RSS and Atom feeds to support syndication of item content. A series of tabs affords the user several additional options for navigating the site. In addition to expected administrative tabs that direct people back to the home page or to the about page to learn about the site, the Browse tab is perhaps the most useful. This directs users to a page where they can browse the directory via ordered lists of tags, collection creators, or dates that collections were added to the directory.



Sharing resources from an Omeka collection is easy as well. The clean and simple URLs generated for items in Omeka make it easy for users to share an item from the system via email or other social networking tools. This stands in direct contrast to most proprietary digital content management systems such as ContentDM or Ex Libris Digitool which typically generate obtuse lengthy URLs that are hard to share and parse with many email and social networking tools. Depending on the browser being used, additional sharing options will be featured in the URL window as well. Omeka has a COinS (ContextObject in Spans) plug-in that embeds bibliographic metadata in the HTML of the page, making the site compatible with bibliographic research tools like Zotero. Additionally, a Social Bookmarking plug-in allows users to share a given resource on any number of social bookmarking sites like Delicious, Facebook, Digg, Yahoo!, Technorati, and more. The Omeka theme mechanism makes it easy to deploy and customize these types of features to provide the type of experience most users expect on today's web.

digitalMETRO Scope and Specifications

The digital collections directory project (digitalMETRO) that we executed using Omeka began with a survey of digital collections in the METRO membership that continued as the project took shape.

Kucsma built an initial list of discrete digital collections created and maintained by METRO member libraries. We established a collection policy rooted in the desire to represent as many collections as possible given our limited personnel and time. A collection would be included in the digitalMETRO directory if:

- the creating institution is a member of the Metropolitan New York Library Council
- the institution owns these resources and has permission for them to be accessed freely online
- the collection includes at least 30 resources (images, audio files, finding aids, etc...)

The last criteria was established to exclude small exhibitions that featured only a handful of digitized resources. To facilitate user-submitted collections that could be added to the directory once the project was made available online we anticipated taking advantage of Omeka's "Contribute" plug-in to allow libraries to submit collections to the directory that might not have been included in the first phase of surveying. Additionally, we also created a simple spreadsheet template that libraries could be used in the future to submit multiple collections for inclusion in the directory using Omeka's CSV Import plug-in.

Metadata Management with Omeka

Throughout the project we made extensive use of one of the richest and most fully realized aspects of Omeka: its metadata support. Omeka supports the standard Dublin Core metadata set by default. Omeka also makes it easy to develop and implement project-specific metadata sets as-needed and have them easily feed into the Omeka display and editing infrastructure. There are two ways to do this:

- Through the creation of Custom Omeka Item Types (http://omeka.org/codex/Managing_Item_Types)
- Through the creation of new Element Sets (http://omeka.org/codex/Creating_an_Element_Set)

These two options mean Omeka can support the creation of a locally defined metadata set or easily add support for an existing

metadata standard. An example of this is the inclusion in the Omeka 1.0 release of a plug-in that makes the Extended Dublin Core Metadata Set available as a metadata or element set choice for any Omeka project. A plug-in could be created for any metadata set desired by an Omeka user. This extensible metadata architecture ensures that Omeka will work well with future metadata initiatives since they can be easily supported using either one of these mechanisms, a key factor in deciding whether Omeka is right for an institution's collections. Both Custom Item Types and Element Sets are easily displayed and manipulated in both the Omeka public end-user display theme and from within the backend Omeka administrative environment.

We used this customizable infrastructure to build our own unique collection metadata schema using a Custom Omeka Item Type. Our item descriptions of collections in digitalMETRO are based on the Basic Dublin Core Element Metadata Set and the built-in Omeka Additional Metadata Set. Omeka provides for additional item-type metadata using pre-defined elements for a variety of digital formats including moving image, website, oral history, email, still image, lesson plan, and more. Because we were essentially creating a directory of digital collections, as opposed to creating records for individual items within a larger digital collection, we opted to create a custom “Collection” item type and add the the following project-specific elements to each record:

- Local Project URL
- Alternate URL
- METRO Funded Project (whether or not this project was funded by a METRO digitization grant)
- METRO Funding Year (if applicable)
- Institution Type (academic, public, or special library)
- Full Project Title
- Sample Item URL (for representative images from a given collection)

Building digitalMETRO with Omeka

The digitalMETRO Omeka instance is hosted commercially by Bluehost on an Apache 2 Web Server running on Linux. The project runs in a PHP5/MySQL5 Environment. Omeka is built upon a robust, open-source PHP web application programming framework. It uses the Zend Framework, a Model View Controller (MVC) application which has enabled Omeka to be designed for extensibility out-of-the-box. Omeka has two means for extension: the theme and plug-in. If these sound familiar to many web developers, these are the same constructs used within WordPress.

We modified one of the existing available Omeka themes for the display (Winter), which can be easily downloaded from the themes section of the Omeka site. Our modifications included some minor changes to the text size and color scheme as well as adding a custom header created from royalty-free vector art licensed from iStockphoto. To add additional functionality to the site, we installed the following plug-ins <http://omeka.org/add-ons/plug-ins/> within our Omeka instance:

<i>Atom Output</i>	Adds the Atom Syndication Format to the list of available output formats.
<i>COinS</i>	Adds COinS metadata to various item pages, making them Zotero readable.
<i>Contribution</i>	Allows collecting items from visitors
<i>CSV Import</i>	Imports items, tags, and files from CSV files.
<i>Dropbox</i>	Allows Omeka users to 'batch upload' a large quantity of files at one time, creating unique items in the archive for each file.
<i>Exhibit Builder</i>	Build rich exhibits using Omeka.
<i>Google Analytics</i>	Adds Google Analytics tracking code to the header of every page.

<i>HTML Purifier</i>	Protects Omeka from XSS by filtering HTML/XHTML using the HTML Purifier library.
<i>Lightbox2</i>	Adds functions for themes to add image overlay functionality to Omeka themes using Lokesh Dhakar's Lightbox2 Javascript library.
<i>OAI-PMH Repository</i>	Exposes Omeka items as an OAI-PMH repository.
<i>Simple Contact Form</i>	Adds a simple contact form for users to contact the administrator.
<i>Simple Pages</i>	Allows administrators to create simple web pages for their public site.
<i>Social Bookmarking</i>	Inserts a customizable list of social bookmarking sites below each item in your Omeka database
<i>Sort Browse Results</i>	Allows sorting of results on browse items page.

The growing community of Omeka developers is actively contributing to the functionality of the software through the development of new existing plug-ins and improvement of existing ones. Plug-in development is happening at a rapid rate. Plug-in deployment is also very simple since plug-ins are written as drag and drop extensions to the Omeka application. All that is required to deploy a plug-in is to load it to the "plug-ins" directory of any Omeka application and fill in any required local customization parameters for the plug-in within the Omeka administrative interface. In fact at the time we were compiling this report, a new Social Bookmarking plug-in was made available, and we were able to add it to digitalMETRO in less than five minutes. There are a number of additional Omeka plug-ins that will be of particular interest to libraries, archives, and museums including OAI-PMH Harvester (OAI Harvesting Support) and Geolocation (adds location information and maps to Omeka).

Given our assertion that Omeka's extensible design and the emerging and expanding set of plug-ins that enhance functionality

make it an exciting and potentially empowering tool for cultural institutions who have web development skills on staff, the existing Omeka project developer documentation has the potential to be a major limitation to theme and plug-in development at the moment. The project documentation has lagged quite a bit behind actual development that has occurred in the nine months we have been working with the system. This will need to be improved upon if Omeka is going to serve the needs of the audiences drawn to its flexible infrastructure and low barriers to implementation. There is an active developers' list and help forums on the Omeka site, but the Omeka Codex <http://omeka.org/codex/Documentation> needs to improve to the level where a competent departmental Web Manager or Web Librarian can easily start hacking the code using clearly defined examples to get started. At the moment, much of the customization work requires combing through source code to find a suitable example or an explanation of a given function's parameters. This is a stark contrast to the much larger pool of documentation and examples available for systems like WordPress.

Plug-In Modification

Our plug-in work to-date with Omeka falls into three categories. First, we deployed a number of plug-ins that required no modification aside from basic configuration within the administrative interface. Second, we modified a few of the plug-ins. The Lightbox plug-in written for Omeka 10 required some slight modification to run on the Omeka 1.0. We also modified the Contribution plug-in so that libraries could recommend collections to be added to the directory. These modifications allow the plug-in to record basic metadata for the "collection" Omeka Item Type we defined for the digitalMETRO directory. Lastly, we are engaged in some new plug-in development that will attempt to increase the findability of digital objects within Omeka collections.

As written, the contribution plug-in was not particularly suited to our project design since it was built to facilitate self-archiving by object authors. For digitalMETRO, we configured the form to solicit

some basic metadata about collections that visitors would like to see added to the directory. Once the form is submitted, a skeleton "non-public" omeka item record is created and queued for the Metadata Librarian to complete and make public on the site. It is easy to see how a feature like this can allow cultural heritage institutions to quickly and easily create sites to collect information and resources from individuals around a certain topic or event. For example, The April 16 Archive site, which is dedicated to collecting and preserving stories about the Virginia Tech tragedy, uses the contribute feature to solicit stories, pictures, and audio files related to the campus shooting.

This type of modification to an Omeka plug-in is relatively easily achieved, with a few caveats. Omeka has been implemented as an MVC application within the PHP Zend framework, so each plug-in conforms to the same framework familiar to today's web application developers. Plug-ins are composed of a directory structure that mirrors the MVC framework with a "view", "controller", and for more complicated plug-ins a "model" directory. The modification for the contribution plug-in mostly happened at the view level, though the internal data structure that stores information about contributions needed to be adjusted to be capture the data from submitters that we wanted. The "controller" file also needed to be modified in order to map contributions directly to the custom "Collection" Omeka item type we used to code each digital project description.

New Plug-in Development

In order to advance the project's goals of highlighting the collections within digitalMETRO we are developing two plug-ins, a Sitemap generator and an Category Browser that turns any Omeka metadata element into a browsable category. The first will enhance the site's discoverability by exposing Omeka content in a simple XML format defined by the sitemaps protocol <http://www.sitemaps.org/>. This file can then be submitted formally to search engines like Google or Yahoo!. The major challenge in building this plug-in was ensuring all Omeka content, exhibits, items, and static content were represented in the sitemap output. A plug-in hook was also added so that the creation

of new content or updates to existing data was recorded in the sitemap.

Browse Collections (7 total) Sound recordings			
Browse by Title	Browse by Tag	Browse by Creator	Most Recent
Sample Image	Title	Creator(s)	Tags
	Academic Podcast Initiative	St. John's University (New York, N.Y.)	Queens, Colleges, Sound recordings, Education, Outreach, St. John's University, Podcasts
	Jewish People's University of the Air	Touro College	Judaism, Jewish life, Sound recordings, Education, Touro College, Jewish People's University of the Air, Radio programs, Outreach
	Childhood in the Bronx	Herbert H. Lehman College	Photographs, Bronx, Lehman College, Sound recordings, Social conditions, Childhood, Oral history

The second plug-in in development will enable Omeka to behave in a fashion similar to the "category" mechanism. The plug-in will allow any Omeka element to become an active browsing option and ensure that the option is represented with a clean URL that can be used effectively by search engines and end-users alike. We took a first pass at this by manipulating the Omeka advanced search screen to enable the Creator, Format, Subject, and Type, elements to be browsable. This delivers cumbersome URLs that look like this:

```
http://nycdigital.org/dmetro/items/browse?search=&advanced[0][element_id]=51&advanced[0][type]=contains&advanced[0][terms]=Sound%20recordings&range=&collection=&type=&tags=&submit_search=Search
```

The plug-in will instead produce URLs that will take the following form:

```
http://nycdigital.org/dmetro/items/browse/format/sound+recordings/
```

The plug-in will provide an administrative screen that will enable an Omeka administrator to select which specific elements will become part of the new browsing options made available through the plug-in. The long-term goal is to make functions available to theme authors so

they may easily create tree-like navigation structures for desirable browsing elements. This will substantially increase the number of different access points users have to items in an Omeka collection. The resulting "categories" will also be printed out in the sitemap generated by the aforementioned sitemap plug-in.

digitalMETRO Workflow and Technical Issues

Creating and adding item content within Omeka occurs primarily in the web browser via the password-protected administrative interface. Like the public-facing side of the Omeka instance, the admin interface is implemented using a theme architecture that can be customized to meet the needs of a particular Omeka project. While we haven't explored this option it is certainly available and could be a means to address some of the administrative user interface issues we encountered during our collection building and metadata creation activities that will be discussed in the next section.

The screenshot displays the Omeka administrative interface for editing a collection item. At the top, the page title is "Edit 'Columbia Historical Corporate Reports Online Collection: Digitized Annual Reports from Watson Library's Collection'", with "Public: Featured: on the right. A left sidebar contains a vertical menu with the following items: Dublin Core, Contribution Form, Additional Item Metadata, Item Type Metadata, Collection, Files, and Tags. The main content area is titled "Dublin Core" and includes a descriptive paragraph: "The Dublin Core metadata element set. These elements are common to all Omeka resources, including items, files, collections, exhibits, and entities. See <http://dublincore.org/documents/dces/>." Below this, there are two metadata fields. The "Title" field has a value of "Columbia Historical Corporate Reports Online Collection: Digitized Annual Reports from Watson Library's Collection" and includes a "Use HTML" checkbox. The "Subject" field has a value of "Columbia University" and includes "Use HTML" and "Remove" options. Each field has an "Add Input" button with a plus icon.

There are low barriers to getting started with record-creation in Omeka. Individual records may be authored in the administrative interface by filling out a tabbed customizable metadata template. There is a "files" tab where media associated with a given item can be

uploaded using a web form with a file input form. Larger groups of records and media may be batch loaded from CSV files, which we discuss below. However, in the earliest stages of the METRO project, the Metadata Librarian entered individual records directly into the administrative interface using the web template. This template is divided into six sections which reflect the intellectual organization of the project's metadata:

1. Dublin Core: Basic Dublin Core Metadata Element Set common to all Omeka projects
2. Contribution Form: Contains data elements pulled from the Contribution plug-in used by member libraries
3. Additional Item Metadata: Custom metadata fields added by project staff to augment the standard DC data elements
4. Collection: Project-specific metadata associated with the custom item type "Collection"
5. File: Associates digital image files with metadata records for display purposes
6. Tags: Allows for the addition of uncontrolled or loosely controlled vocabulary to aid in subject analysis and collection description

The bulk of the Metadata Librarian's work occurred in the Dublin Core (DC) tab of the item creation form. It was here that project description (Title, Creator, Format, Description) and subject analysis (Subject) took place. For librarians with experience in creating records using DC the item creation template is familiar and intuitive. For librarians accustomed to creating records using the rigid structure and numeric field tags of MARC, the Omeka interface offers a soft entry into new types of metadata creation. The web form provides a blank field for each data element and users may build from there, adding fields through a simple click on a "plus" button labeled "Add input." Under each data element is a definition of the element's purpose and examples of proper usage. These data element definitions would be particularly useful for projects in which there are multiple record

creators, including students or interns, and for librarians with limited experience in creating non-MARC metadata.

While it was easy to begin authoring records in the administrative interface, the limitations of the system became clear as we started working with a greater volume of records. For example, there is no mechanism for populating related records with common data within the administrative interface. To ensure consistency between records, the user must rely on time-consuming tactics such as copying and pasting from previously created records or from source sites like the Library of Congress authority files (LCSH for subject authorities and LCNAF for names) or the Art and Architecture Thesaurus (AAT). Additionally, navigating within a single record was time-consuming, as the template's layout required a burdensome amount of scrolling. This scrolling could not be avoided as the "Save Changes" button appeared only at the bottom of the screen and regular saving was key to successfully authoring records in Omeka. The item creation template was, of course, web based and if a user accidentally navigated away from the active record, either to another tab within Omeka or to one of the many sites from which data had to be copied and pasted into the form, then any unsaved content within the record disappeared. The system does not offer a preventative warning asking if the user would like to save any unsaved changes before closing the record. On several occasions we lost time and detailed intellectual work by accidentally navigating away from the item creation form.

These issues make Omeka problematic as a serious choice for a project that will involve large-scale and or detailed metadata creation, since using it in the fashion digitalMETRO attempted revealed serious data consistency issues. However, considering the strong, extensible MVC architecture Omeka is built on top of, all of these issues seem to be within reach of being fixed if the development community would take notice. As better documentation of the Omeka administrative theme and Omeka plug-in develop, these issues could be addressed. For example, support for controlled vocabularies could come via a plug-in developed for a given vocabulary. A more streamlined editing interface that supports both record templates and data-typing of metadata fields could be accomplished through rewriting some of the

MVC code within the administrative theme. A more robust search and retrieval mechanism using the Lucene search engine available within the Zend framework is currently under development by the core Omeka development team.

Project staff quickly moved to batch loading records as a means of bypassing as many of the administrative interface's shortcomings as possible. The project manager created an Excel spreadsheet and populated it with key data elements which could be determined through a preliminary viewing of a potential site: Creator, Title, URL, METRO funding status, etc. The Metadata Librarian then reviewed the spreadsheet and ensured that the completed fields complied with LCNAF before the web developer loaded the file into Omeka. This process created a queue of skeleton records which the Metadata Librarian could then enhance through description of subjects, formats, additional authors, and other data before making the records viewable on the public side of the site. The benefits of moving to batch loaded CSV files were immediately clear. The use of spreadsheets allowed the Metadata Librarian to easily populate fields with shared data, which improved consistency within and between records and also decreased the time spent on cutting, pasting, and scrolling. This, in turn, led to a decrease in the number of records lost through browser navigation errors.

Intellectual Control and Tagging within Omeka

Even with batch loads, control over data within Omeka is loose. Catalogers more comfortable with the tight intellectual control offered by a traditional library information system or a digital content management system like CONTENTdm may be disappointed by the weak search and retrieval features and the inability to carry out basic find/replace tasks or other global changes within the record-creation module. The weakness of the intellectual control provided by Omeka means that for a project of any size or scope it is crucial for librarians to get their records right the first time around, when they author them. This is not necessarily easy, as Omeka does not support controlled vocabularies such as LCSH, AAT, or the Thesaurus of

Geographic Names. Metadata authors must work around the system to ensure that they enter correct, consistent, and authoritative data into their records. For this project multiple strategies were employed to control headings, including cutting and pasting directly from authority files, loading standardized data through batch loads, cross-checking through keyword searches on the public site, and exporting alphabetical lists of subject headings and tags into tab delimited files for review by hand. In addition to the difficulties faced in creating consistent data content, there was an additional hurdle in that Omeka does not include a record validation function to make sure that core fields are populated.

The tagging mechanism built into the item creation module offers a bright contrast to the difficulties of working with controlled data. With the move to Omeka 1.0 an unlimited number of tags can be added to each record. Our project staff opted to take full advantage of the tags, building them into a robust companion to the more structured DC data. The same core metadata fields found in the item records are repeated in the tagging, only this time instead of relying on controlled vocabularies, natural language terms are used. Compare, for example, the structured and unstructured terms used to

describe John Jay College's collection "Crime in New York 1850-1950," <http://nycdigital.org/dmetro/items/show/18>.

	Controlled vocabularies	Tags
Creator	John Jay College of Criminal Justice	John Jay College
Additional creators	Turkus, Burton B.	Burton B. Turkus
	Lawes, Lewis Edward, 1883-1947	Lewis Edward Lawes (1883-1947)
Subjects	Homicide	Homicide
	Organized crime	Organized crime
	Criminal investigation	Criminals
	Sing Sing Prison	Sing Sing Prison
	Death row inmates	Inmates
Materials	Forensic photographs	Crime scene photographs
	Identification photographs	Mug shots
	Transcripts	Trial transcripts

A keyword search on the public site will pull up content added through controlled vocabularies as well as the tags. In the absence of traditional see/see also references, which are built into library catalogs and databases, the hope is that this combination of structured and unstructured data will help lead non-expert users to the information they seek. Even an experienced librarian might not immediately think to search for "Identification photographs" but the more commonly used "Mug shots" will guide users to the same authoritative data and with a deal less frustration or confusion.

Experiments in Exhibition Building

Omeka was developed with the express needs of museums, historians, and educators in mind. The software's origins in the arena of cultural institutions could account for what is, by library standards, weak

intellectual control within the admin site. However, its development at George Mason's Center for Social History and New Media also directly led to one of Omeka's greatest strengths: exhibition building. Once users have created records within the system they can, with great ease, put together exhibits which shine a brighter light on related groups of material. A site editor fills out a web form to determine top level data such as Title, Slug, and Theme, then assigns section metadata, and finally adds individual records to the section. A given exhibition may have multiple sections, thus allowing for more complex relationships between digital objects. Templates make the layout of individual records semi-customizable within exhibits. Some templates give precedence to images while others allow for a greater expanse of exhibition-specific didactic text.

Exhibition-building fell on the margins of METRO's project goals and digitalMETRO does not fully exploit this feature's true potential. Nevertheless, project staff did create two exhibits which appear on the lower left side of the homepage as "Featured Exhibits." The "Art and Design" collection offers a good example of how the exhibition feature can be used for educational or curatorial purposes. This exhibit gathers together design drawings and finished pieces from across the applied arts. Site visitors may explore fashion, furniture, and theater sets through multiple libraries' collections, including the Cooper-Hewitt National Design Museum, Columbia University, Yeshiva University Museum, and Parsons the New School for Design. In this case, the exhibition was built around the unifying theme of design history and illustrated the range of materials owned by METRO's member libraries. The strength of the exhibition feature is in its ability to convey a curatorial perspective in a dynamic, interactive setting which guides users from object to object while providing expert contextual information alongside standardized metadata and social tagging.

Conclusions

Omeka has great potential to effectively and efficiently support small and medium-sized digital collection building and online exhibitions

for libraries and archives. The software is well-suited to enable librarians, archivists, and curators to work effectively in the context of their disciplines. The core Omeka software architecture and design are well-suited to allow the software to expand and improve as the user community grows. It is also well-positioned to serve as a tool institutions can use to repackage existing digital collections in a new, modern web exhibition framework with the availability of a robust CSV import option that can bring both metadata and media into the system.

Continued improvement of the Omeka administrative interface within this architecture will help it become a more viable digital collection building solution for libraries and archives. We experienced a good deal of frustration navigating through the interface to create records. We also had serious concerns about the ability to create consistent and detailed metadata within the administrative interface. The ability to create and manipulate metadata in a spreadsheet and populate skeleton records in Omeka works around some of the administrative interface issues, but there is still room for improvement in how a user creates records online. Omeka's search and retrieval capabilities also need to improve for it to become a more fully realized digital collection management tool. At this writing the Omeka development is working to utilize the Apache Lucene support with the Zend framework to provide a more effective full-text search interface for Omeka. It is hoped this effort will result in a new "Lucene" plug-in for the Omeka 1.1 release.

Despite these limitations Omeka is certainly well-positioned within its target market of small to medium-sized institutions that need an easy-to-deploy, effective, professional tool to make digital library, archival, and museum content available on the web. It is important to measure Omeka's functionality, features, and limitations against some of the same functionality, features, and limitations of a proprietary system. We hope that our experience will help other institutions evaluate Omeka as a possible collection management system for their digitization projects and anticipate and overcome any of the obstacles we encountered.

References

- 2006_digsurveyreport.pdf*. (n.d.). Retrieved September 22, 2009, from http://www.metro.org/images/stories/pdfs/2006_digsurveyreport.pdf
- 2007_digplan.pdf*. (n.d.). Retrieved September 22, 2009, from http://www.metro.org/images/stories/pdfs/2007_digplan.pdf
- Cohen, D. (2008, February 20). *Introducing Omeka*. Dan Cohen's Digital Humanities Blog. Retrieved September 22, 2009, from <http://www.dancohen.org/2008/02/20/introducing-omeka/>
- Dave Lester's Finding America. New Omeka Release 0.10 Beta*. (n.d.) Retrieved September 22, 2009, from <http://blog.davelester.org/2008/11/12/new-omeka-release-010-beta/>
- digitalMETRO*. (n.d.). Retrieved September 22, 2009, from <http://nycdigital.org/dmetro/>
- Explore Murray Hill through images and maps. Digital Murray Hill*. (n.d.) Retrieved September 22, 2009, from <http://murrayhill.gc.cuny.edu/>
- Model-view-controller - Wikipedia, the free encyclopedia*. (n.d.). Retrieved September 22, 2009, from <http://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller>
- Omeka | Documentation - Omeka How To*. (n.d.) Retrieved September 22, 2009, from <http://omeka.org/codex/Documentation>
- Omeka | Home*. (n.d.). Retrieved September 22, 2009, from <http://omeka.org/>
- OpenURL ContextObject in SPAN (COinS)*. (n.d.) Retrieved September 22, 2009, from <http://ocoins.info/>
- The April 16 Archive*. (n.d.) Retrieved September 23, 2009, from <http://april16archive.org/>
- WordPress Blog Tool and Publishing Platform. (n.d.) Retrieved September 22, 2009, from <http://wordpress.org/>
- Zend Framework. (n.d.) Retrieved September 22, 2009, from <http://framework.zend.com/>

