MetaDB v3.2 User Guide

(revision 1.0)

Eric Luhrs
Digital Initiatives Librarian
Lafayette College
April 12, 2011

0. MetaDB Development and Support

MetaDB was designed to facilitate Digital Initiatives projects at Lafayette College, and we feel that our digital collections have been strengthened by its use. In the hope that other institutions may benefit from our work, we have decided to release the source code for MetaDB v3.2 under the General Public License (GPL). MetaDB carries no warranty whatsoever, is used completely at one's own risk, and comes with absolutely no guarantee of support.

That said, we intend to answer general questions by email so far as time allows, and we welcome community support from others who want to improve MetaDB, either by pointing out problems, suggesting new features, or by contributing new or improved code. Please direct all correspondence to metadb@lafayette.edu, or visit our project page at http://digital.lafayette.edu/projects/metadb.

1. General Overview

MetaDB is a web application that provides tools to automate and distribute the creation of metadata for digital image collections. Typically, an administrative user creates and configures projects and manages user access to the system. Once a user account has been created, the holder can login and edit metadata for projects to which they have been assigned. After all metadata for a given project has been entered, an admin user exports project data from MetaDB, and loads it into the organization's forward-facing image repository.

Most image repository platforms provide basic tools for creating collections and adding metadata. MetaDB makes it easier to create strong data records by providing remote access to users, by allowing multiple collaborators to work concurrently, by including tools to improve the speed and quality of metadata entry, by adding the ability to quickly sort, narrow, and search image sets, by centralizing the management of project settings and controlled vocabulary lists, and by automatically generating technical metadata and access derivatives.

MetaDB has been used extensively at Lafayette College over the past few years, so we hope that the User Interface is mostly self explanatory. This document is meant to provide a high-level overview of how the system works, and will explain some of MetaDB's main features. For a general overview of installing MetaDB, please see MetaDB-v3.2-general-install.txt. For detailed instructions about installing MetaDB on

RedHat Enterprise Linux 6, including configuration of all required software and services, please see MetaDB-v3.2-RHEL6-install.pdf.

The following sections are organized by the titles that appear on tabs and sub-tabs in MetaDB's user interface.

2. User Management

MetaDB has two user types: *admin* and *worker*. Admin users have access to all projects, to the "User Management" and "Project Management" tabs, and to all "Data Management" sub-tabs. Using these restricted features, admin users have the ability to create and delete worker accounts, specify authentication type, reset local passwords, and grant/revoke worker access to individual projects and features. Worker accounts require explicit permission to view or edit data for each project, as well as permission to access restricted features for each project.

A. Create User

Each new user requires a username, user type, and authentication type. Note that LDAP authentication will only work if network settings are correctly configured in metadb.conf. Use of LDAP authentication requires settings appropriate to your environment in metadb.conf. This feature uses the supplied username to authenticate against an LDAP password, which means that MetaDB and LDAP usernames must match. Only LDAP users specifically added to MetaDB will be allowed to login. Local passwords provide a way to provide external users with access to the system, which can be useful for collaborative projects with outside researchers. After a new user has been created, the user interface will automatically load the "User Management" → "User Settings" sub-tab.

B. Delete User

The "Delete User" sub-tab is self-explanatory. Simply select the user and click "Delete." Note that this action is permanent, and cannot be undone. For security reasons, admin accounts must be deleted with the command-line `utils/create-delete-admin-users.sh` script on the server.

C. User Settings

After selecting a username from the dropdown menu, admin users can modify user type, authentication type, change local passwords, and grant/revoke permission to view or edit data, and access to restricted features for each project.

Figure 1 shows a typical user configuration.

Each row from the table displayed in *Figure 1* is used to control user access to specific projects and features:

project — Selects the project for which access settings will be configured for the selected user.

desc metadata — Specifies if selected user will have read only or read/write access to descriptive metadata for the selected project.

admin metadata — Specifies if selected user will have read only or read/write access to administrative metadata for the selected project.

table edit — Specifies if selected user will be able to view metadata for selected project from "Data Management" → "Table Edit" sub-tab. If table access is granted, fields with read/write access will be editable in table mode. Note that the table edit interface does not check against controlled vocabulary lists.

vocab admin — Specifies if selected user is allowed to manage controlled vocabulary lists from "Data Management" → "Controlled Vocabulary" sub-tab. If vocab access is granted, user will be able to create, modify, and delete controlled vocabulary lists associated with the specified project.

import/export Specifies if selected user is allowed to access import and/or export tools from "Data Management" → "Import Metadata" OR "Expoprt Metadata" sub-tabs. Be aware that granting import permission allows users to permanently overwrite all metadata for the selected project.

3. Project Management

1. Create Project

Each new project requires a name for identification within MetaDB. User input is restricted to alphanumeric characters and hyphens. After creation, project names are displayed in the project dropdown box to the right of the main tab set that appears on every page.

The supplied project name value also tells MetaDB where to look for master images within the MASTERS_DIRECTORY (as defined in metadb.conf). So, for example, if MASTERS_DIRECTORY is set to /assets/metadb/master and the project name is "test", then MetaDB will look for master images in /assets/metadb/master/test. It is important to note that all master files must end with a sequential four-digit number, such as filename-0001.tif, filename-0002.tif, and that generated derivatives will take the same form, and based on the above metadb.conf settings, will be saved to /assets/metadb/access/test.

The project notes field provides space for a brief description. Note that these fields cannot be edited after the project has been created. The use project as template checkbox allows for the creation of a new project with the same field-configuration as an existing project. After a new project has been created, MetaDB automatically loads the "Project Management" → "Edit Project" sub-tab.

2. Delete Project

The "Delete Project" sub-tab offers three levels of deletion:

"Delete all data and images" permanently deletes the entire project, including all descriptive, administrative, and technical metadata, and all derivatives created for the selected project.

"Delete administrative/descriptive data only" permanently deletes all supplied metadata, but leaves intact technical metadata and all derivatives created for the selected project.

"Delete technical data and images only" permanently deletes all technical metadata and all derivatives created by MetaDB but leaves descriptive and administrative metadata intact. Note that after derivative images have been deleted, they can be re-parsed from the "Project Management" → "Project Settings" sub-tab.

3. *Edit Project*

a. *Project Information*

Field values in this section cannot be edited after project creation.

b. *Descriptive & Administrative Metadata*

These identical sections allow admin users to configure metadata fields and settings to be used for the selected project. Fields defined as descriptive will appear on the default tab while editing metadata, with administrative fields, and automatically generated technical metadata on adjacent tabs.

Fields can be reordered within or between descriptive and administrative sections by grabbing the up/down arrow and moving the field to a new location. The "-" sign deletes the field associated with it, and the "+" sign creates a new field. Multiple fields can be added at once by specifying the number of fields desired and clicking the button. It is important to click each section's update button before navigating away from this page.

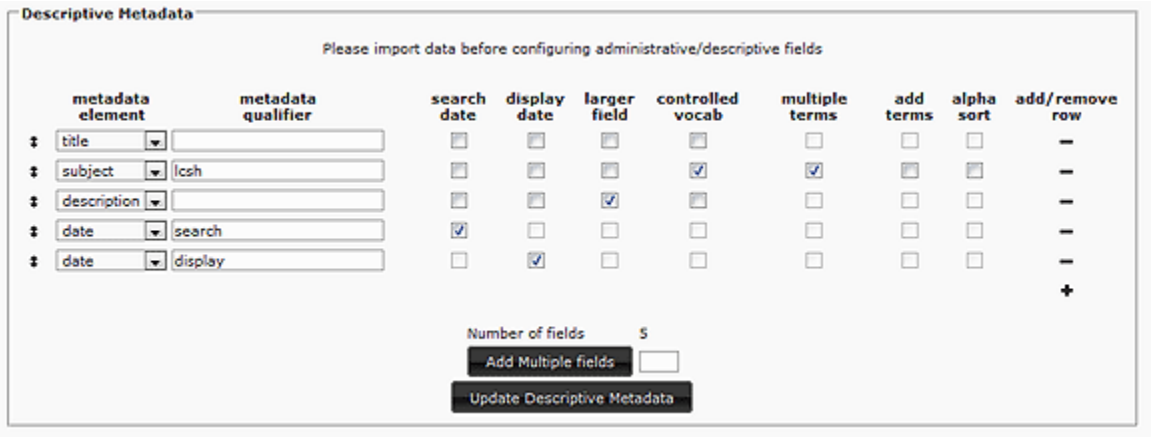

Figure 2 shows an example field configuration.

The **metadata element** dropdown includes the list of elements defined in metadb.conf (currently Dublin Core) and the **metadata qualifier** text field allowing these metadata fields to be modified for specific types of content. For example, qualifying the subject element with lcsh would yield a metadata field called "subject.lcsh". Note that duplicate field names are not allowed.

The following settings can be applied to each field:

search date Field will only allow machine readable dates in the following formats: YYYY, YYYY-MM, YYYY-MM-DD, YYYY-YYYY

display date Field will only allow human readable dates in the following formats: YYYY, YYYY-MM, YYYY-MM-DD, YYYY or YYYY, circa YYYY-MM, YYYYs, before YYYY, after YYYY

larger field Display eight rows for text, instead of one.

controlled vocab Force selection from a list of authorized terms. Choosing this option will display a popup window for selection/creation of controlled vocabulary list. A controlled field has the following options:

multiple terms Allow users to choose more than one controlled term.

add terms Allow users to create new controlled vocabulary terms during data entry.

alpha sort Display and export applied vocabulary terms alphabetically. Default is order entered.

c. Derivative Settings

This section defines the derivatives that will be generated for the specified project, and sets size and branding options. Note that URLs for zoom and download JPG derivatives are publically available outside of MetaDB, with links provided in technical metadata fields called identifier.url.zoom, and identifier.url.download (if "Download JPGs" checkbox is selected).

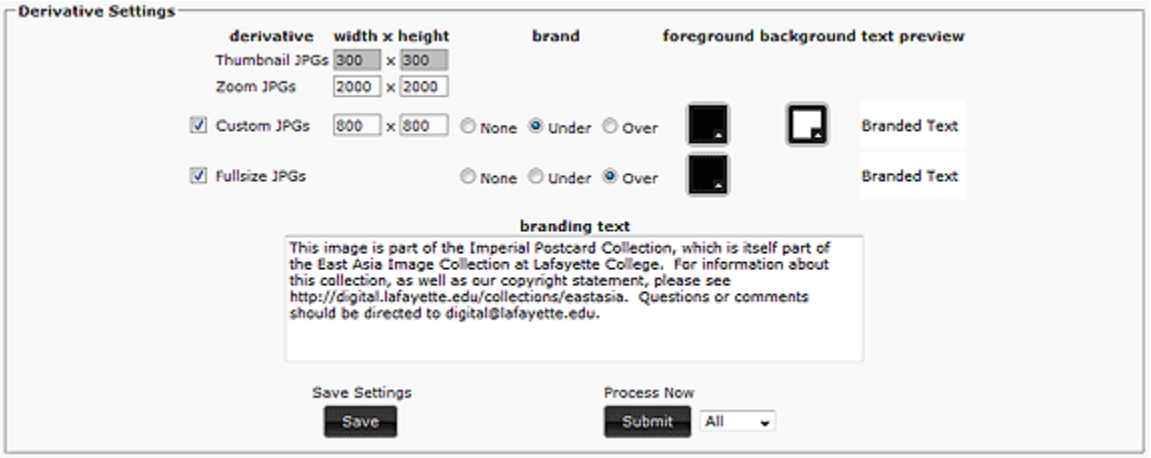

Figure 3 shows an example derivative configuration.

Thumbnail and zoom images are required for the "Item Metadata" tab and the pan/zoom interface that is automatically created for each item within MetaDB.

zoom image size can be customized in the UI, but thumbnail size must be changed in metadb.conf.

These numbers are constraints, not specific sizes, so if you intend to reduce both portrait and landscape oriented images to fit on a 1024x768 screen, you would restrict both height and width to 1024 pixels. This will reduce the longest edge, whether it is the height or the width, to 1024 pixels, with the shorter edge being reduced proportionally to the same ratio.

Checkboxes allow admin users to specify if smaller download-sized and fullsized derivatives are desired. If created, optional text can be branded over or under these derivatives. The save button records settings without parsing images, the submit button parses all or specified images right away. Technical metadata is automatically generated whenever images are parsed. Note that it is necessary to parse master images for each collection to generate derivatives and technical metadata.

4. Data Management

The "Data Management" tab provides specified users with access to the following data management tools:

a. Table Edit

Unlike the "Item Metadata" tab, which displays all fields associated with one particular item on a single screen, the "Table Edit" feature provides an interface that allows users to view specific fields from multiple items in tabular form. Upon loading the "Table Edit" sub-tab, users are presented with a list of all fields in the selected project, from which up to four fields can be display at once. Technical, administrative, and descriptive metadata sections will indicate if the current user has read-only or read-write access to related fields.

Figure 4 shows selected fields in tabular format.

When data is first loaded, all selected fields from the current project are displayed 50 at a time. This number can be adjusted using the left-hand dropdown menu, and chunks of data can be navigated using the right-hand pager. Above the pager is a search box which will narrow displayed records to those that include the provided search term(s).

This example shows all items that include the keyword "military". Clicking a column heading will sort all data in ascending order, by that column; clicking again switches to descending order. Filenames appear first since they provide a convenient way to link images with field data. Clicking a filename loads a popup version of the full "Item Metadata" editor for that item.

When metadata is displayed in table view, read-only fields are printed gray and read-write fields are printed black. Black read-write fields can be clicked for inline editing, gray read-only fields cannot. So in this example, format.technical.DPI fields, the data for which was automatically extracted by the image parser, cannot be edited.

It is important to note that this interface does not check against controlled vocabulary lists, which makes it possible to enter unapproved terms.

b. Import/Export Metadata

Access to the import and export features is controlled separately, but the "Import Metadata" and "Export Metadata" sub-tabs look almost identical. Both allow approved users to specify if certain character types should be preserved or encoded for safety, and to choose between comma separated values (CSV) and tab separated values (TSV) data formats.

There are two main differences between the import and export sub-tabs. One is that the import feature requires users to specify a file to upload, while the export feature expects a data rage (all records is the default). The biggest difference between MetaDB's import and export features concerns technical metadata. Because MetaDB parses master images to generate technical metadata, including an MD5 checksum, it does not allow existing technical metadata to be imported into the system. For this reason, all imported data is assumed to be either descriptive or administrative. After importing metadata, any duplicate or non-Dublin Core fields will be marked red on the "Project Management" tab. It is important to note that importing new metadata deletes all existing metadata, so access to this feature should be reserved for advanced users.

Export, by contrast, allows users to specify if technical metadata should be included. Keep in mind that the filename is a technical metadata field, and will likely be required when importing data into other systems.

c. Controlled Vocabulary

The "Controlled Vocabulary" sub-tab allows approved users to manage lists of vocabulary terms from which controlled metadata values must be chosen. Users with access to the Controlled Vocabulary Management tool are not able to change controlled vocabulary field settings or the associations between metadata fields and vocabulary lists, but they are able to add, delete, and modify terms for vocabulary lists used by any project to which they have been granted access.

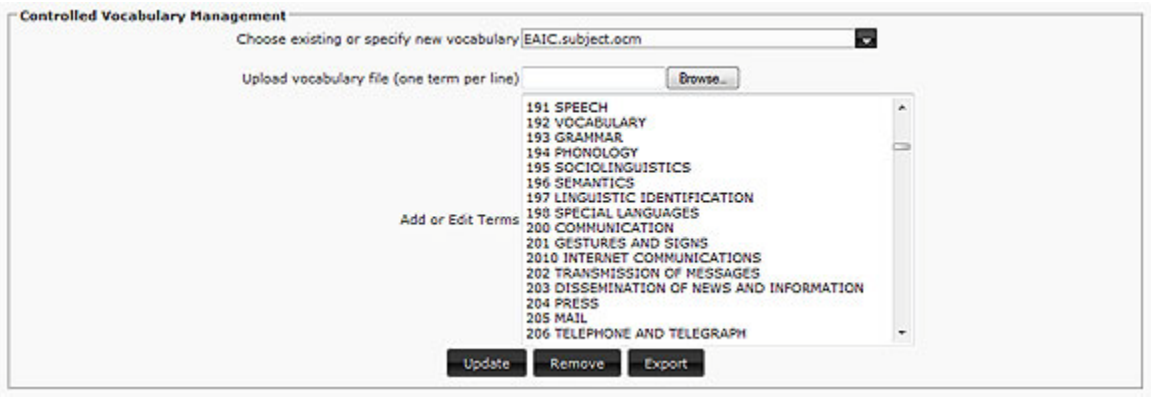

Figure 5 shows a sample list of controlled vocabulary terms.

To edit an existing list, choose the list name from the dropdown menu. To create a new list, provide a new list name in the same text box, then upload a text file that contains the approved terms, or enter them into the text box. The "Update" button saves changes, "Remove" deletes the selected list and "Export" creates a text file of approved terms that can be saved locally.

5. Editing Project Metadata

The "Item Metadata" tab is the primary metadata editor for all items managed by MetaDB, and is where general users will spend a majority of their time. This page displays metadata fields and derivative images that an admin user has configured using the "Project Settings" tab. Even though many of MetaDB's core features are used to create and interact with the "Item Metadata" page, we have tried to hide as much complexity as possible in order to provide users with an interface that is uncluttered and easy to use.

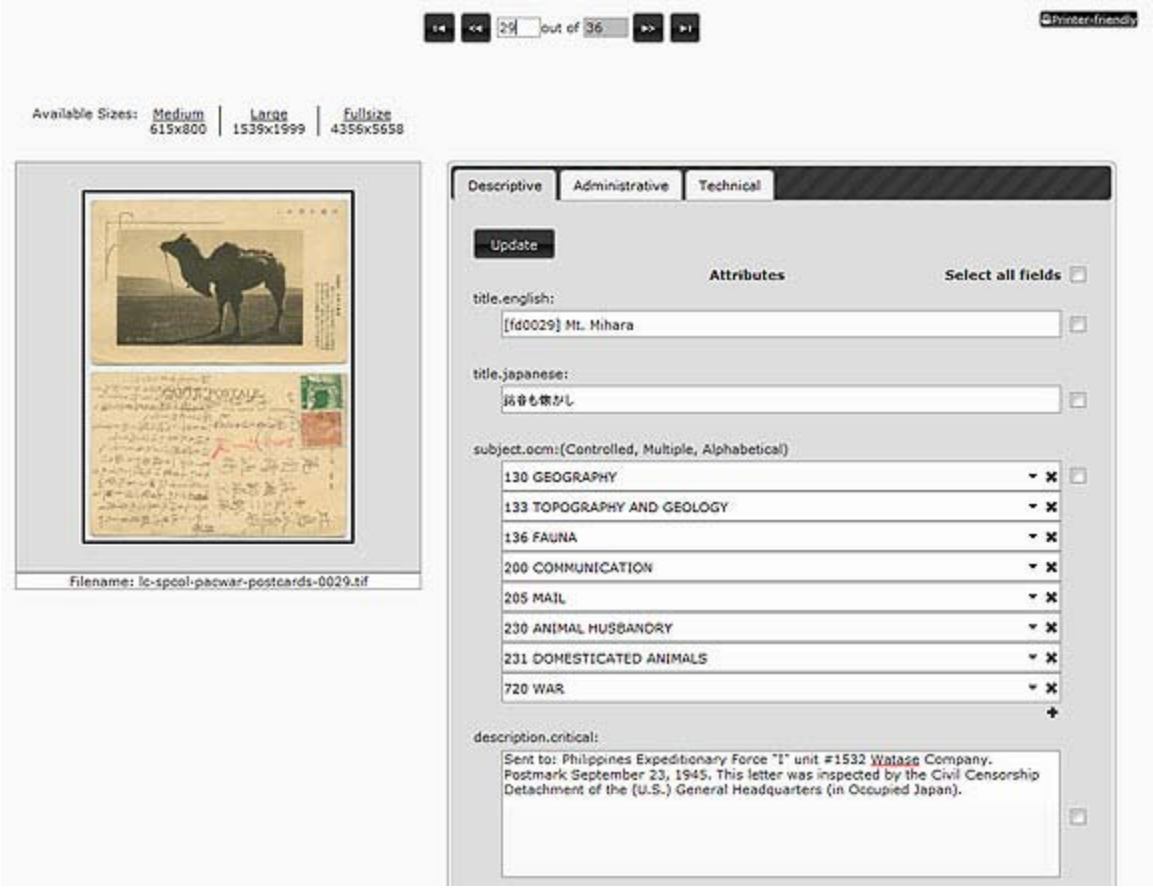

Figure 6 shows an example record from the "Item Metadata" tab.

The "printer-friendly" button displays the thumbnail and a simple list of fields without the user interface. Record navigation is handled by the MS Access-style controls at the top of the page. The first button returns to the first item in the current collection, the second button goes back one record, the third button moves forward one record, and the fourth button jumps to the last record. Each button saves field data before moving to the next record. Between the buttons are two small number boxes, the first which allows users to jump to a specific record, and the second which displays the total number of images in the current collection.

The left side of the page provides access to derivative images. Clicking the thumbnail loads a new tab with the zoom/pan interface. Because terms like Download JPG, Zoom JPG, and Fullsize JPG (as specified on the Project Settings tab) mean little to most users, they are respectively called Medium, Large, and Fullsize here. The master filename is listed below the thumbnail. It is important to understand that record numbers correlate to master image files, which are required to end with matching sequential four-digit numbers, such as filename-0001.tif, filename-0002.tif, etc.

The right side of the page is dedicated to a tabbed metadata interface that allows users to toggle between Descriptive, Administrative, and Technical metadata fields. Note that Technical metadata is generated automatically, and is not editable.

Metadata field names indicate which settings have been selected on the "Project Settings" tab. Allowing the cursor to hover over a configured field displays a brief tool tip with data entry instructions. Controlled fields are presented with a simple dropdown menu. Controlled fields that allow multiple terms add +/- controls, and controlled fields that allow additions use a combo dropdown box from which existing terms can be selected, or new terms added. Search and Display date fields will be checked for required syntax. The height of large fields can be adjusted by dragging the bottom edge of the text box.

Checkboxes to the right of the metadata input fields allow users to select values from the current record and apply them to other records within the same collection. Once some or all checkboxes have been selected, a new menu is revealed next to the "Update" button, with options to apply values from selected fields to "all records" or "these records" within the collection. Choosing "these records" reveals a text box that allows users to specify record numbers within the current collection to which selected data should be copied. Accepted syntax includes an individual number, multiple numbers separated by commas, and ranges of numbers separated by dashes. Multiple numbers and ranges may be combined.

6. Search

The "Search" tab allows users to query metadata from projects to which they have been granted access. The search options are simple: current collection or all collections. Search results are displayed using an interface that is similar to that used for the "Table View" sub-tab.

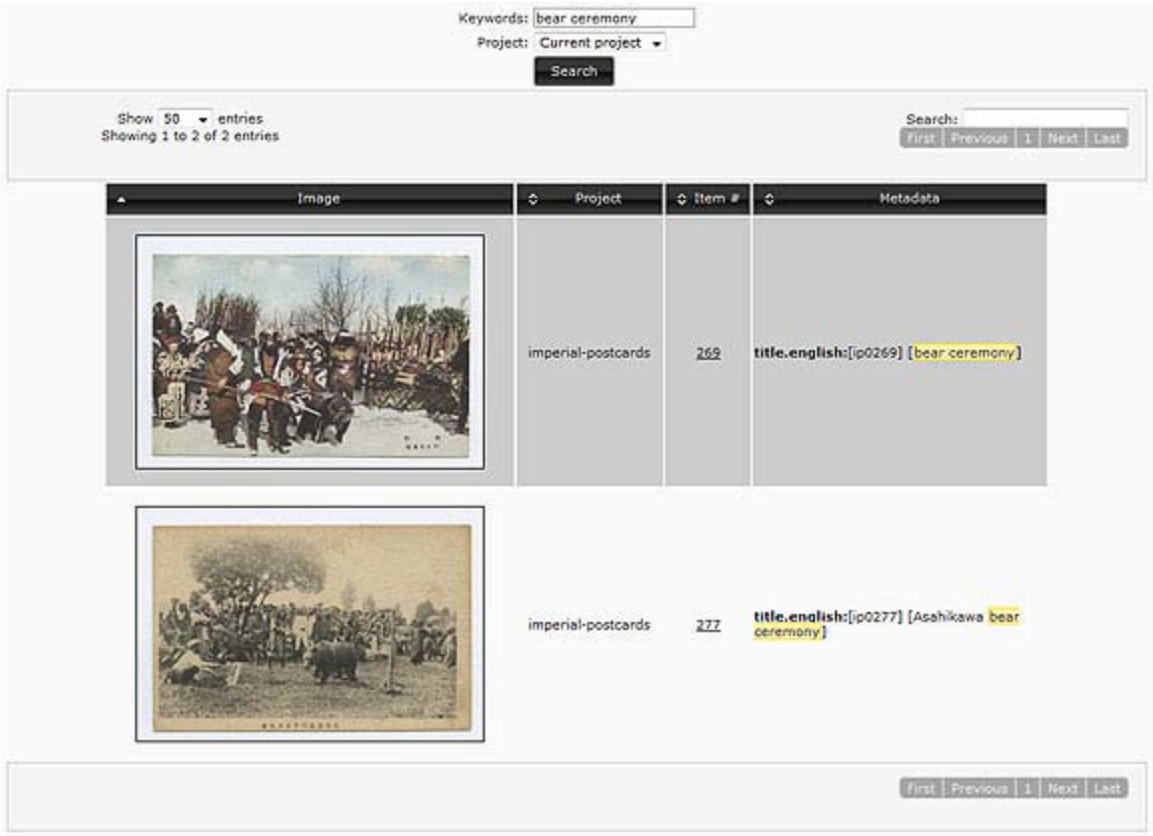

Figure 7 shows two records return by a sample search.

Each row shows a thumbnail, the project name, record number, and any fields that contain the search terms, with query text highlighted yellow. Clicking a thumbnail launches the zoom interface in a new tab, and clicking the record number loads a popup version of the full "Item Metadata" editor for that item. Like the "Table View" sub-tab, the search interface allows users to set the number of records displayed at one time, page though these groupings, and further narrow results based on additional keywords. Search results can also be sorted by ascending or descending order by clicking on any column heading.