logo.tif

Metadata Services Toolkit

User Manual

Table of Contents

Setting up the MST 3

Login 3

My account 3

Email Configuration 3

LDAP configuration 4

User Management 4

Repository Management 5

Harvesting metadata from a repository 6

Adding/Updating Metadata Services 7

Adding a Metadata Service 7

Updating a Metadata Service 8

Controlling the flow of data 9

Browsing Metadata Records 9

Viewing Logs 10

Glossary 11

The Metadata Services Toolkit (MST) is a web application which harvests metadata records from one or more OAI repositories, processes the harvested data with one or more ­metadata services to refine it into a more useful form, and expose the processed metadata with the OAI repositories connected with each of the metadata services. This document assumes you have already installed the MST by following the instructions found in the [MST Installation Manual](http://code.google.com/p/xcmetadataservicestoolkit/downloads/list). This user manual covers the following:

* how to create additional user accounts and manage their permissions
* schedule harvests of OAI repositories
* control the flow of metadata through the metadata services
* view the metadata before it was processed and after each service processed it
* transfer the processed metadata into other applications using the OAI protocol

In this document, the URLs of pages within the MST are written assuming that you have installed the MST on a Tomcat Server running on localhost port 8080. You will have to adjust the domain and port of these URLs if your install does not match these values.

## Setting up the MST

Only users with “Configuration” permission can perform the operations described in this section other than logging in and editing their account settings. Permissions are described in detail in the “User Management” section, but for now it is enough to know that all administrator accounts have this permission.

### Login

When the MST is installed it comes with an administrator account with the username/password combination “admin/mst4world.” The first thing you should do after installing the MST is to login to the MST at http://localhost:8080/MetadataServicesToolkit/index.jsp using this account.

### My account

Click the “My Account” link above the main menu on the left side. We recommend that you change the password of the admin by clicking the “Change Password,” then filling out and submitting the form. You should also change the email address associated with this account to that of the person responsible for maintaining the MST using the form on the “My Account” page. This is the address to which password recovery emails will be sent, so it is important to configure it correctly. To recover a forgotten password click the “Forgot Password” link from the login page, enter the account’s email address in the textbox and click the submit button.

### Email Configuration

Next, hover over “Configuration” in the main menu and select “Email Configuration” from the menu that drops down. From this screen you can define the SMTP server that the MST should use to send email notifications. Enter the location of the SMTP server in the “Outgoing Mail Server” box. The “From Address” and “Password” are the credentials used to login to the account on the SMTP server which will send notifications, and the “Port Number” is the port on which the SMTP server can be accessed. You can select the type of encryption the SMTP server uses from the “Encrypted Connection” dropdown, and enter the number of seconds the MST should wait for a response from the SMTP server before giving up in the “Timeout” box. After all the information has been entered, click the “Update Email Configuration” button to save the changes.

### LDAP configuration

Optionally, you can configure the MST to use an LDAP server for external authentication. To do so, hover over “Configuration” and select “LDAP Configuration.” Fill in the name your institution uses to refer to the LDAP server in the “Display Name” box (this name will appear in the user interface above the username and password boxes). The “Server URL” box is for the URL of the LDAP server and the “Port” is the port on which it can be accessed. The “Username Fieldname” is the field on the LDAP server which contains the user’s username, and “LDAP start” is the location on the LDAP server which stores the users. If the LDAP server contains a form users may fill out if they forgot their password, you can configure the MST to link to it. To do so, select “Yes” next to “Show Forgot Password URL” and enter the URL in the “Forgot password URL” box. After all the information has been entered, click the “Update LDAP Server” button. To remove LDAP support after it has been configured, click the “Clear LDAP Server” button.

## User Management

In the MST, a “group” is a bundle of one or more permissions which all users belonging to that group have. Groups are the only source of permissions for users, so it is important that they be configured correctly. A group contains one or more of the following permissions which correspond to the tabs on the MST’s main menu:

* Repositories: The user can view, add, edit and delete OAI repositories known to the MST
* Harvest: The user can view, add, edit and delete schedules for harvesting metadata from OAI repositories known to the MST.
* Services: The user can view, add, edit and delete metadata services which the MST uses to process and refine metadata. The user can also view the services and other jobs that the MST is currently running.
* Processing Rules: The user can view, add, edit and delete the rules which govern which metadata services process which metadata records.
* Browse Records: The user can access the MST’s faceted browsing interface for viewing metadata processed by the MST.
* Logs: The user can view logs detailing the actions the MST took while harvesting and processing metadata, as well as which metadata was harvested from the MST by outside OAI harvesters.
* Users/Groups: The user can create new users and groups and can assign and revoke permissions from them.
* Configuration: The user can edit the information about the SMTP Server and LDAP Server used by the MST.

A user with the correct permissions can define a new group by selecting “Add Group” from the “Users/Groups” tab in the MST’s menu. This directs the user to a form which allows them to provide a name for the group, and optional description of the group, and to select one or more permissions that members of the group will have (CTRL+click to select multiple permissions from the list.) Clicking “Add Group” will create the group according to the user’s specifications, and clicking “Cancel” will return the user to the “All Groups” page. The “All Groups” page—which can also be accessed by selecting “All Groups” from the “Users/Groups” tab—shows the basic information about all groups in the MST and gives the option to delete any group except for Administrator, which contains all permissions and cannot be deleted.

Users may be added to the MST in two ways: users with Users/Groups permission can add them directly, or a user can create an account with no permissions and request that a user with the appropriate permissions assign them the appropriate privileges. To add a new user, hover over “Users/Groups” in the main menu and select “Add Local User.” Enter the user’s first name, last name, username, initial password and email address in the form, and select one or more groups to which they belong. (Groups determine what permissions a user has and will be described later in this section.) Click “Add Local User” when the form is filled out to add the user, or “Cancel” to return to the list of users without adding a new user.

A user may create their own MST account on the User Registration screen at <http://localhost:8080/MetadataServicesToolkit/viewUserRegisteration.action>. This form allows them to specify their first and last names, select a username and password for their account, enter their email address, and send a message to the MST administrator. It also allows them to select the server which they authenticate against. At the time of this writing the Local Server is the only option, which means that the user will authenticate directly against the MST.

When a user creates an account in this way, the account will be created immediately but will not have permission to access any functionality of the MST. An email will be sent to the MST administrator group (MST users who have ‘administrator’ permission) informing them that the account has been created and requesting that they assign permissions to the user. A user with Users/Groups permission may then assign permissions by selecting “Users” from the “Users/Groups” menu and clicking on the username of the new user. From this screen, the user may select one or more groups for the new user to belong to and click “Update Local User.”

## Repository Management

Only users with “Repository” permission can perform the actions described in this section.

The first step to processing data with the MST is to define the OAI repositories from which data will be harvested. A user with the appropriate permissions can add an OAI repository to the MST by selecting “Add Repository” from the Repositories tab. Fill out the form with a name by which the MST knows the repository and the repository’s URL, then click the “Save” button to add a new repository. This automatically validates that the OAI repository correctly handles the Identify, ListMetadataFormats, and ListSets OAI verbs, and the results of the validation will be displayed along with buttons to edit the repository information, view the log for the repository (which shows a detailed reason for any validation failures,) run the validation on the repository again, and delete the repository. This information page can also be accessed by clicking on a repository’s name from the “All Repositories” page in the “Repositories” tab. This “All Repositories” page shows the name and URL of each repository as well as their status (whether or not they passed validation) and the timestamp when they were last harvested.

## Harvesting metadata from a repository

Only users with “Harvest” permission can perform the actions described in this section.

A scheduled harvest is a job executed by the MST which harvests an OAI repository at regular intervals. The first time a scheduled harvest runs for a given OAI repository it adds all records in that repository to the MST’s Solr index, and subsequent runs of the same scheduled harvest will add only those records that were changed since the previous harvest began and will remove all records in the index which the OAI repository has marked as deleted. This means that the first run of a scheduled harvest will take several hours for repositories with a few million records, but the following runs will finish in much less time since unchanged records are not harvested.

To add a scheduled harvest, first add the OAI repository to be harvested and then go to the “Add Scheduled Harvest” screen in the “Harvest” tab. Select the repository to schedule a harvest for from the top drop down. Then select how often the scheduled harvest should be run (hourly, daily or weekly) and specify the minutes after each hour, hour of the day, or day of the week and hour when the schedule should be run depending on how often it should run. Optionally, the start and end dates of the scheduled harvestcan also be specified to delay the first run of the scheduled harvest or prevent it from running after a certain time. After filling out this information click “Move to Step 2” to continue defining a new scheduled harvest or “Cancel” to return to the “List Scheduled Harvests” screen without adding a new scheduled harvest.

Step 2 of the “Add Scheduled Harvest” form displays a list of metadata formats and OAI sets that the repository selected in step 1 supports. Select one or more formats and one or more sets to be harvested by the scheduled harvest (CTRL+click to select multiple sets or formats.) Either change the name of the scheduled harvest in the box on the left or leave it unchanged to accept the default name. Optionally, an email address may be provided. This address will be emailed with the results of each run of the scheduled harvest. Click “Finish” to add the new scheduled harvest, “Back to Step 1” to change answers on the previous page, or “Cancel” to return to the “List Scheduled Harvests” screen without adding a new scheduled harvest.

The “List Scheduled Harvests” page in the “Harvest” tab displays information on all the scheduled harvests that have been set up. For each scheduled harvest the name of the scheduled harvest, name of the repository harvested, frequency and time when the scheduled harvest will run, and status of the scheduled harvest are shown. Clicking on the name of a scheduled harvest allows a user to edit it, and clicking “Delete” will delete the scheduled harvest.

While a scheduled harvest is running, its name appears in the green box in the top right corner of the MST on every page. If a scheduled harvest is running, clicking the “Pause” button will suspend it until the “Resume” button which replaces it is clicked. While a scheduled harvest is paused the MST will not run any other jobs, but pausing a job will free up resources for processes other than the MST running on the same server. Clicking “Abort” will cause the scheduled harvest to stop running and the MST will then begin the next job.

## Adding/Updating Metadata Services

Only users with “Services” permission can perform the actions described in this section.

A metadata service is a subroutine of the MST which takes a set of metadata records as input and produces a new set of output records based on the service’s unique functionality. The new set of metadata records are created by combining and improving data from records passed into the service. These output records are added to a Solr index in the MST .

Metadata services for normalizing Marcxml data can be downloaded from the eXtensibleCatalog.org website, or from others in the XC community. Future releases of the MST will also include metadata services for converting MARCXML records to XC schema and aggregating data from multiple XC schema records. Metadata services will be released as a .zip file containing all the files required to run them. Unzipping these files in MST-instances/MetadataServicesToolkit will make them available to be added to the MST.

### Adding a Metadata Service

Steps to add a new metadata service to the MST:

1. Download the .zip file for the metadata service from eXtensibleCatalog.org or a third party website.
2. Unzip the downloaded file to MST-instances/MetadataServicesToolkit.
3. Log into the MST with a user that has “Services” permission.
4. Navigate to the “Add Service” page in the “Services” tab.
5. Select the configuration file that was contained in the downloaded .zip file from the dropdown and click “Add”.

### Updating a Metadata Service

The “All Services” page in the “Services” tab shows information on all services added to the MST. This page shows the name of each service as well as its status and the URL of the OAI repository from which its output records can be harvested. The “Service” and “Harvest Out” buttons download log files for the processing done by a metadata service and the OAI repository of a metadata service respectively.

While a metadata service is running, its name appears in the green “Process Description” box in the top right corner of the MST on every page. If a metadata service is running, clicking the “Pause” button will suspend it until the “Resume” button which replaces it is clicked. While a metadata service is paused the MST will not run any other jobs, but pausing a job will free up resources for processes other than the MST running on the same server. Clicking “Abort” will cause the metadata service to stop running and the MST will then begin the next job.

To update a service with a later version, with the MST in a stable state, stop the Tomcat Server. Individual files can be replaced as needed in the service’s installation directory in MST-instances/MetadataServicesToolkit/services/<your-service>. Restart the Tomcat Server. When the MST restarts it checks each installed service and checks the timestamps of the service’s \*.jar, \*.class, \*.xccfg and \*.properties files. If it finds an updated file time, it reprocesses the service’s files.

## Controlling the flow of data

Only users with “Processing Rules” permission can perform the actions described in this section.

A Processing Rule is a user defined rule which the MST uses to determine which metadata records should be processed by which metadata services and in what order. Processing Rules consist of a source of metadata, a service to process the metadata, a list of OAI sets and metadata formats which the service should process, and optionally the definition of a new OAI set to which processed records should belong.

To add a new Processing Rule, go to the “Add Processing Rule” page under the “Processing Rules” tab. Select either an OAI repository or a metadata service from the radio buttons on the left side of the screen as the source of metadata and select a metadata service from the radio buttons on the right side of the screen to process the metadata. Then click “Continue to Step 2” to finish configuring the Processing Rule or click “Cancel” to return to the “All Processing Rules” screen without adding a new Processing Rule.

Step 2 of the “Add Processing Rule” form displays a list of metadata formats and OAI sets that the source selected in step 1 supports and that the metadata service selected in step 1 accepts as input. Select one or more formats and one or more sets to be processed by the metadata service (CTRL+click to select multiple sets or formats.) Optionally, a name and setSpec for an OAI set into which records processed by the Processing Rule are inserted may be provided. Click “Finish” to add the new Processing Rule, “Back to Step 1” to change answers on the previous page, or “Cancel” to return to the “List Processing Rules” screen without adding a new Processing Rule. For each Processing Rule the “List Processing Rules screen shows the source, metadata service, formats, and input sets that define the Processing Rule. It also contains a button for deleting a Processing Rule.

## Browsing Metadata Records

Only users with “Browse Records” permission can perform the actions described in this section.

The “Browse Records” tab allows users to search all records in the MST’s Solr index using a faceted browsing interface. This interface facets on the name of the repository from which a record was harvested, the name of the service which produced the record, the format of the record, the set(s) which contain the record, the harvest event in which the record was harvested, and the errors reported for the record. To display records, either select a facet or enter a search in the search box. If the checkbox under the search box is checked the content of the record is searched in addition to the facet values.

Clicking on the OAI Identifer of a record will display all the details about that record as well as the XML for the record itself. Clicking on the “Successors” link for a record will display all records which were created when a metadata service processed the record, and clicking on the “Predecessors” link will display all records that a metadata service used to produce the record.

## Viewing Logs

Only users with “Logs” permission can perform the actions described in this section.

The MST keeps four different types of logs. General logs track problems in various categories of the MST’s basic operation. Actions taken and problems encountered while harvesting an OAI repository are logged in Harvest In logs. Actions taken and problems encountered running metadata services are logged in Services logs. The Harvest Out logs track the metadata services’ OAI repositories as they are being harvested by external harvesters.­ Each type of log has a page in the “Logs” tab with more information.

The “General” page in the “Logs” tab shows information for the following eight areas of the MST which are tracked in their own log files:

* Authentication Server Management: Tracks changes made to the LDAP server configured for the MST.
* Jobs Management: Tracks scheduled harvests and metadata services which are configured by the user and executed by the MST.
* MST Configuration: Logs problems with the MST’s configuration file. See the [Installation Manual](http://code.google.com/p/xcmetadataservicestoolkit/downloads/list) for more information on how to configure the MST.
* MySQL: Logs problems accessing the MST’s database and running queries.
* Repository Management: Tracks OAI repositories that are added, edited, and deleted from the MST.
* Service Management: Tracks metadata services that are added, edited, and deleted from the MST.
* Solr Index: Logs problems interacting with the Solr index.
* User Management: Tracks actions taken on the MST that relate to users, groups, and permissions.

For each log, the number of warnings and errors in the log file are displayed on the General Logs page, as well as a button to reset these counts and the date when the statistics were last reset. Resetting a log does not remove the warnings and errors form the log file, it only sets the warning and error counts to 0. The buttons on the top-right and bottom-right of the screen will reset every log on the General Logs page.

The “Harvest-In” page in the “Logs” tab displays information on the OAI repositories in the MST. For each repository a link displaying the OAI request(s) ran during the last harvest schedule is shown, as well as a link to the Browse Records screen faceted to show only records from that provider (“Browse Records” permission is required for this link to work.) The date of the most recent harvest is also shown, as well as cumulative counts of the number of records added during harvests and replaced because a more recent version of the record was harvested. Clicking on the name of a repository will download a log file with more detailed information on actions taken while harvesting the repository. The reset button will reset the records added and records replaced counts, but will not remove the log file. The buttons on the top-right and bottom-right of the screen will reset every log on the Harvest-In Logs page.

The “Harvest-Out” page in the “Logs” tab displays information on the metadata services’ OAI repositories. When an external OAI harvester (such as the Drupal Toolkit) harvests metadata from the MST, it is actually harvesting from and OAI repository for a particular metadata service. For each metadata service the number of records available to harvest and number of times the service is harvested are displayed, as well as the number of warnings and errors generated by harvesting the metadata service. Clicking on the name of a metadata service will download a log file with more detailed information on actions taken while the repository was harvested. The reset button will reset the number of harvests, warning and error counts, but will not remove the log file. The buttons on the top-right and bottom-right of the screen will reset every log on the Harvest-Out Logs page.

The “Services” page in the “Logs” tab displays information about the metadata services which process records in the MST. For each metadata service, the number of records processed by the service, the number of records produced by the service, and the number of warnings and errors the metadata service reported are shown. Clicking on the name of a metadata service will download a log file with more detailed information on the action taken by the metadata service. The reset button will reset the warning, and error counts, but will not remove the log file. The buttons on the top-right and bottom-right of the screen will reset every log on the Services Logs page.

## Glossary

Harvest – Refers to an event when a scheduled harvest was run.

Metadata service – A Java class used by the MST to enhance and refine metadata.

OAI protocol – A protocol for passing metadata between systems. The MST uses this protocol to harvest metadata from outside sources. The OAI protocol is explained [here](http://www.openarchives.org/OAI/openarchivesprotocol.html).

OAI repository – A web service that exposes metadata records using the OAI protocol.

Scheduled harvest – A job in the MST that harvests an OAI repository. A scheduled harvest can be set up to run once an hour, once a day, or once a week. When run it will only harvest changes since the start of the previous run.