

MetaScholar Metadata Migrator User Guide Version 1.0

An application for migrating metadata from
databases to OAI data providers.

Urvashi Gadi, Liz Milewicz, Katherine Skinner

May 18, 2005

Version 1.0 Manual

Document Information

Version:	1.0
Created:	09/10/2004
Last Modified On:	05/18/2005
Author:	Urvashi Gadi, Liz Milewicz
Technical Lead:	Martin Halbert, Katherine Skinner
Contributors:	

Revision History

S. No.	Revision	Date Modified	Modified By	Comments
1.	1.0	10/29/04	Urvashi Gadi	Initial Draft
2.	1.8	12/10/04	Elizabeth Milewicz	Changed Format, Revised Draft
3.	1.9	04/10/05	Urvashi Gadi	Added Confirm Screen
4.	2.0	05/18/05	Katherine Skinner	Revised Draft

Table of Contents

- Getting Started 1
 - Logging In 1
 - First-Time Users 2
 - Password Help 2
- Step 1: UPLOAD 3
- Step 2: VALIDATE 5
 - First Data Record 5
 - Field Names 7
 - Start Over..... 8
- Step 3: CONFIGURE 9
 - Required Fields 9
 - Archive Identifier 9
 - Administrator Email 10
 - Optional Fields..... 10
 - Repository Name 10
 - Record Limit 10
 - Data Filename 10
 - Metadata Formats 11
 - Other Information 11
- Step 4: CROSSWALK 12
 - Discard Option 14
 - Derived Title Option..... 14
- Step 5: PRODUCE..... 15
 - Confirm 15
 - Produce..... 16
- Appendix: Data Files 17
 - Types of Data Files 17
 - Xbase Data File (.dbf) file 17
 - Comma-Separated Value (.csv) File..... 17
 - Tab-Delimited (.tab) File 17
 - Exporting Files as .csv, .tab, or .dbf 17
 - Access instructions for exporting to dBase (.dbf) 17
 - ProCite instructions for exporting to .csv or .tab format 17

Getting Started

The Metadata Migrator tool allows institutions such as museums, archives, research centers, and small libraries to make their locally stored records available for online searching. Using the Metadata Migrator, collections specialists can map or crosswalk the field names of their institution's records into Dublin Core elements to create OAI-compliant XML records. They can also create a data provider that allows OAI harvesters to serve out these records within larger digital library structures, including such sites as OAIster and AmericanSouth.org. Because institutions can select which fields from each record to "migrate," they retain control over sensitive information while making general information about their collection available to scholars conducting web searches.

Logging In

The Metadata Migrator was created by the MetaScholar Initiative at Emory University General Libraries, through an IMLS (Institute of Museum and Library Services) grant. It is located online through the Emory University website, and can be accessed through the following link: <http://metascholar.org/metadata>.

The log-in screen (Figure 1, below) should appear in your window. This page lets you log in with the username and password you received from the MetaScholar Metadata Migrator administrator when you registered. Once you've successfully logged in, the UPLOAD screen will appear and you can proceed with Step 1.

[Help](#)

M E T A S C H O L A R
M E T A D A T A M I G R A T O R

Step 1
UPLOAD

Step 2
VALIDATE

Step 3
CONFIGURE

Step 4
CROSSWALK

Step 5
PRODUCE

Please log in below using the username and password provided to you by Emory University's MetaScholar Initiative.

Username :

Password :

[Log-in help?](#)
[Forgot your password?](#)

Figure 1. Login Screen

First-Time Users

Users must have a username and password in order to use the Metadata Migrator. If you are a first-time user, please contact the current Metadata Migrator administrator (mdmsupport@metascholar.org) to register and receive your username and password. This email address is also provided through the "Log-in help?" link.

Password Help

If you've forgotten your password, click on the link "Forgot your password?" You will be asked to provide the email address that you submitted when you registered (see Figure 2). Your password will then be emailed to this address.

Metadata Migrator Password Help

Please enter the email address you provided when you registered:

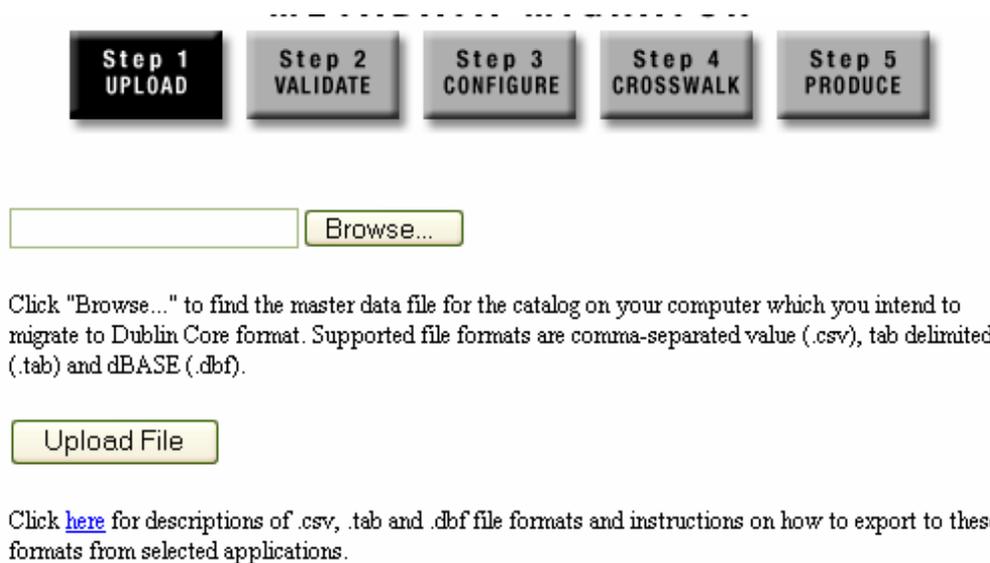
Submit

Figure 2: Password Help screen

Step 1: UPLOAD

From the UPLOAD screen (Figure 3), you can upload a data file to begin the process of migrating the file to Dublin Core XML. The data file should be formatted as a .csv (comma separated value), .tab (tab delimited) or .dbf (dBase) file, and should be located on or mapped onto your computer. These files organize and separate the data and the records in predictable ways, making it easier for the Metadata Migrator Tool to identify each record and its data and eventually to map data with Dublin Core elements.

If the data file is not already in one of the three required formats, it may be easily converted. Click on the “here” link (just below the “Upload File” button) for instructions on how to convert data files and for more information on the three data file formats. You can also find this information in the Appendix to this manual.



Step 1
UPLOAD

Step 2
VALIDATE

Step 3
CONFIGURE

Step 4
CROSSWALK

Step 5
PRODUCE

Browse...

Click "Browse..." to find the master data file for the catalog on your computer which you intend to migrate to Dublin Core format. Supported file formats are comma-separated value (.csv), tab delimited (.tab) and dBASE (.dbf).

Upload File

Click [here](#) for descriptions of .csv, .tab and .dbf file formats and instructions on how to export to these formats from selected applications.

Figure 3: UPLOAD screen

Clicking on the browse button enables you to browse files on your local machine. Figure 4 (below) shows what this might look like on your computer.

Step 1
UPLOAD

Step 2
VALIDATE

Step 3
CONFIGURE

Step 4
CROSSWALK

Step 5
PRODUCE

Click "Browse..." to find the master data file for the catalog on your computer which you intend to migrate to Dublin Core format. Supported file formats are comma-separated value (.csv), tab delimited (.tab) and dBASE (.dbf).

Click [here](#) for descriptions of .csv, .tab and .dbf file formats and instructions on how to export to these formats from selected applications.

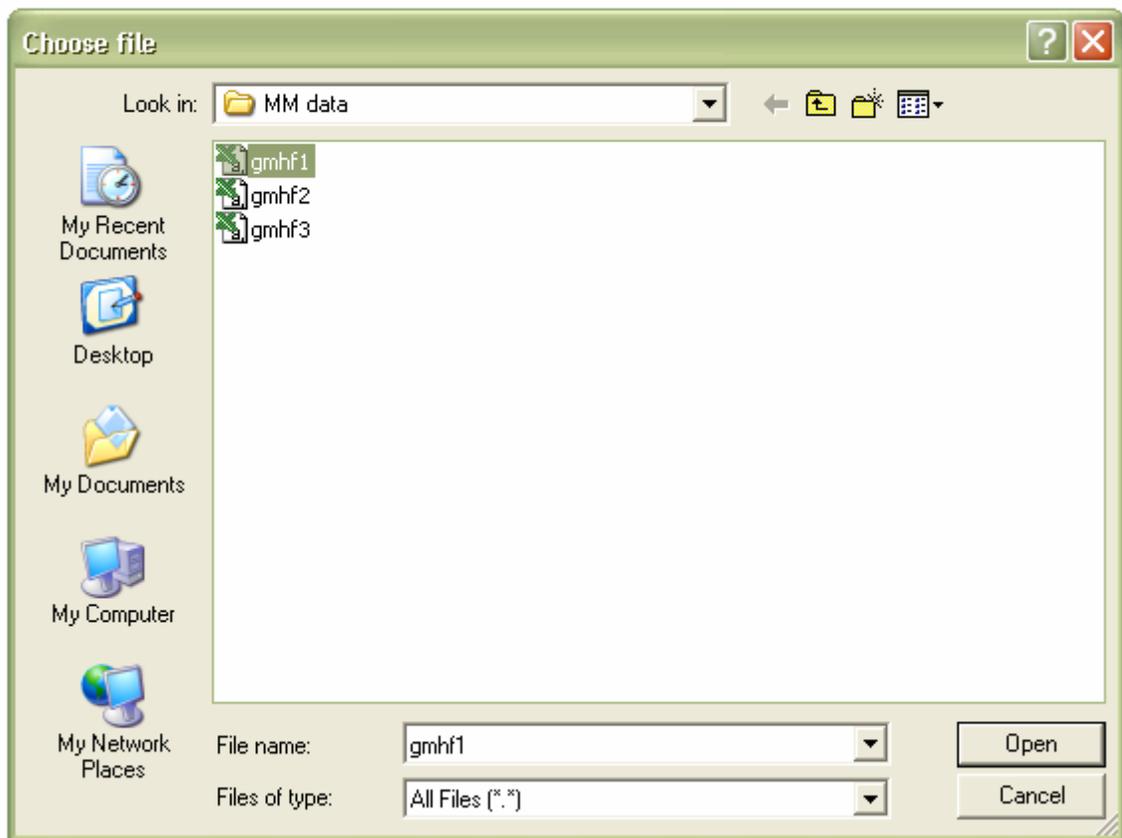


Figure 4: Browsing and selecting a local file to upload

Once you have selected and opened a data file, it appears in the "Browse" box. Clicking on the "Upload File" button uploads the selected file and opens the VALIDATE screen (Step 2).

If the Metadata Migrator cannot open your file, an error screen will open, indicating that there was a problem with the file. After a few seconds you will be directed back to the UPLOAD screen, where you can select a different file to upload or simply log-out. Make sure that the file you are attempting to load is one of the three file types supported by the Metadata Migrator: .csv (comma separated value), .tab (tab delimited) or .dbf (dBase).

Step 2: VALIDATE

The VALIDATE page displays the first line of information from the uploaded file, and asks, “What type of information is this?” You may indicate what type of information is displayed by clicking one of three buttons: “First Data Record” (if it appears to be information about a specific item), “Field Names” (if it appears to be headings or general labels), or “Start Over.” If you realize that you have uploaded the wrong file, you may go back to the UPLOAD screen by clicking on the “Step 1 UPLOAD” button.

The following subsections illustrate the “First Data Record” and “Field Names” pathways, showing the steps that will be involved with each before you may proceed to Step 3. The last subsection, “Start Over,” describes why you might need to select “Start Over” and what to do before you attempt to upload the file again.

First Data Record

Figure 5 (below) provides an example of a first record with data. If the information on your screen contains information about a specific item (i.e., the first data record), click the “First Data Record” button. The Metadata Migrator will then assume that there are no field names for the data elements, and that this is the first record.

Step 1
UPLOAD

Step 2
VALIDATE

Step 3
CONFIGURE

Step 4
CROSSWALK

Step 5
PRODUCE

The first record of the uploaded file is:

Deffa, Chip
Reference book
GMHF-LIB-19
Voices of the Jazz Age

Library

What type of information is this?

First Data Record Field Names Start Over

Figure 5: VALIDATE screen, with “First Data Record” information

Once you have validated the type of information displayed on the screen, the Metadata Migrator will display the first three records, *without* field names (Figure 6). You will need to provide field names which correspond to each line of data. Creating field names not only provides a specific label for the data, which can then be mapped to Dublin Core elements, but it also helps you to check that the data in each line are all the same type and to clarify for yourself how they are conceptually linked.

As Figure 6 illustrates, the first three data records are displayed to the right of the field name boxes; “discard” is automatically given as the Field Name for each line of data. Select which data to migrate by supplying appropriate field names for the data. Click in the box to the left of each line of record data, highlight the word “discard,” and type in the appropriate field name for that data. Dotted lines within each record indicate that no data was available for that particular field. You should refer to your master files to determine what field name should be associated with these lines of data. **By default, data will be discarded unless you provide a field name.**



Please provide field names for each data element:

Field Name	Record 1	Record 2	Record 3
discard	Deffa, Chip	Finn, Julio	Wigginton, Eliot
discard	Reference book	Reference book	Reference book - folk art and music
discard	GMHF-LIB-19	GMHF-LIB-7	GMHF-LIB-65
discard	Voices of the Jazz Age	The Bluesman	Foxfire 2
discard	-----	-----	-----
discard	-----	-----	-----
discard	-----	-----	-----
discard	-----	-----	-----
discard	-----	-----	-----
discard	Library	Library	Library

Next>>

Figure 6: Display of first three data records without field names

If these records are not displaying correctly, you may need to start over by clicking on the “Step 1 UPLOAD” button. Before proceeding again, be sure to check the record file to make sure it is one of the three accepted types (.csv, .tab, or .dbf) and that the information in the file is entered correctly.

When you have finished entering the field names, click the “Next” button at the bottom of the screen. This will take you to the CONFIGURE screen (Step 3).

Field Names

Figure 7 (below) provides an example of a first record that contains field names. If the information on your screen is a list of field names, click the “Field Names” button. The Metadata Migrator will then assume that these are the field names for the data elements of each record.

Step 1
UPLOAD

Step 2
VALIDATE

Step 3
CONFIGURE

Step 4
CROSSWALK

Step 5
PRODUCE

The first record of the uploaded file is:

artist
item
catalog
title_make
label
no__serial
composer
producer_a
performer
location
year_induc
venue
city_count
connection

What type of information is this?

First Data Record Field Names Start Over

Figure 7: VALIDATE screen, with “Field Names” information

Once you’ve validated the type of information displayed on the screen, the Metadata Migrator will display the first three records *with* their field names (Figure 8, below). This display gives you an opportunity to check that the records will be migrated correctly. Compare each field name with the data in the first three records, as displayed to the right of the “Field Name” box. If the field names correspond to the records’ data, click the “Next” button to proceed to Step 3.



First three data records are:

Field Name	Record 1	Record 2	Record 3
artist	Deffa, Chip	Finn, Julio	Wigginton, Eliot
item	Reference book	Reference book	Reference book - folk art and music
catalog	GMHF-LIB-19	GMHF-LIB-7	GMHF-LIB-65
title_make	Voices of the Jazz Age	The Bluesman	Foxfire 2
label	-----	-----	-----
no_serial	-----	-----	-----
composer	-----	-----	-----
producer_a	-----	-----	-----
performer	-----	-----	-----
location	Library	Library	Library
year_induc	-----	-----	-----
venue	-----	-----	-----
city_count	-----	-----	-----
connection	-----	-----	-----

Next>>

Figure 8: Display of first three data records with field names

If these records are not displaying correctly, you may need to click the “Step 1 UPLOAD” button or the “Step 2 VALIDATE” button to return to an earlier step in the process. Before proceeding again, be sure to check the record file to make sure it is one of the three accepted types (.csv, .tab, or .dbf) and that the information in the file is entered correctly.

Start Over

If the information displayed on the first UPLOAD screen is neither data for the first record nor the field names for your file, you may click on the “Start Over” button to return to the UPLOAD screen. If this happens, please check the following features of your file before uploading it again:

- 1) Is the file one of the three accepted types: .csv, .tab, or .dbf?
- 2) Are the fields in the record properly delimited (i.e., with a tab or a comma)? Make sure that your field names and field data are properly entered and separated before you attempt to map them to Dublin Core.
- 3) Is the first record in your file blank? If there is no information displayed, the first record in your file may be blank.

Step 3: CONFIGURE

The information you provide on the CONFIGURE screen, shown in Figure 9, is used to set up your OAI data provider. This information will become part of your migrated metadata's unique identity, and will be used by OAI metadata harvesters when they harvest your data.

The CONFIGURE screen asks a series of important questions. Answers are required for two of these questions: you must provide a unique identifier for your converted data (see the subsection on the "Archive Identifier" below), and you must provide an email address that will serve as a contact point for those who access your migrated records. The other questions are optional, but we strongly encourage you to complete all of these questions. When you're done, click the "Next" button to go to the CROSSWALK screen (Step 4).



Please answer the following questions to configure the OAI data provider that you are creating.

REQUIRED FIELDS

Archive identifier
[unique identifier for this archive]

Administrator's email

OPTIONAL FIELDS

Repository Name
[for identification purposes]

Record Limit
[number of records to be harvested before issuing a resumption token]

Data filename
[select a data element to create unique file names]

Metadata format

Other Information
[additional comments about this OAI data provider]

Figure 9: CONFIGURE screen

Required Fields

Archive Identifier

This information is **mandatory**. The archive identifier is a machine-readable string of numbers, letters, or a combination of both, which is unique to the data. This unique number/name will be used as part of the universal resource locator (URL) assigned to the migrated metadata. Keep track of the archive identifiers you have used for your metadata, and do not repeat identifiers from previous sessions or data. If you do repeat an archive identifier, a message will alert you that you have used the identifier before and must select a different

one. As an example of one way to create an archive identifier, you might use the numerical date and your name to uniquely identify a set of migrated records, adding ordinal numbers at the end if more than one set is migrated in one day (see Figures 10 and 11).

REQUIRED FIELDS

Archive identifier
[unique identifier for this archive]
Administrator's email

051805mhalber

mhalber@emory.edu

Figure 10: An Archive ID created using a six-digit numerical date, first name initial, and last name

REQUIRED FIELDS

Archive identifier
[unique identifier for this archive]
Administrator's email

051805mhalber2

mhalber@emory.edu

Figure 11: An Archive ID created for another batch of records, migrated on the same day, by the same person

If you have migrated a set of records and then decide that you would like to make changes to it and migrate it again, **give this data a new, unique identifier**. Do **not** use the data's original archive identifier to "overwrite" the previous data, as this cannot be done without creating errors. If you need to clear a directory, contact the administrator: mdmsupport@metascholar.org.

Administrator Email

Providing an administrator's email address makes it easier for problems to be reported and corrected. Providing this information is **mandatory**. An email address should automatically appear in this box, based on the username you used to log in. You can either use this default email address or enter a more appropriate contact address for reporting problems and corrections.

Optional Fields

Repository Name

By default, the repository name is listed as "OAI Archive." However, we encourage you to provide a more specific name (for example, the name of the institution affiliated with this data). The repository name helps others to identify the source of the information.

Record Limit

Large data sets may need to be broken down into smaller chunks to avoid interruptions or to keep machines from becoming bogged down with data transmission. The record limit is the number of records that will be migrated at a time before the harvester is asked whether or not to continue. The default is 500 records.

Data Filename

By default, the Metadata Migrator will number each record in order to keep them distinct. If your records do not already have unique identifiers, or if you are not sure whether their identifiers are unique, leave this on "default" and do not select a file name.

However, if you are migrating records that already have unique file names, you may want to use those file names to distinguish your migrated records. For instance, if you created unique cataloging numbers for each record and listed them under the field heading "Catalog," you can instruct Metadata Migrator to use data from this field to uniquely identify each record (see example in Figure 12). The pull-down menu lists your data's field names (validated in Step 2). You may use any of these to create unique file names by selecting it from the list.

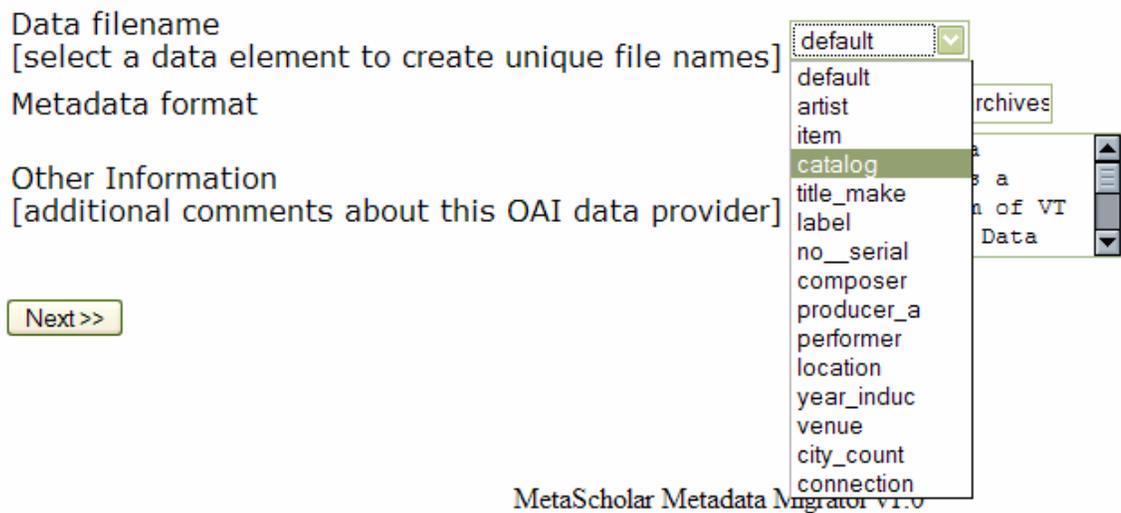


Figure 12: Drop-down menu of possible data filenames

Metadata Formats

By default the Metadata Migrator uses an XML schema for unqualified Dublin Core metadata formats, which replaces your original field names with Dublin Core elements.

Other Information

You may use the comment box to provide additional information about your OAI data repository.

Step 4: CROSSWALK

At the CROSSWALK screen (Figure 13), you connect the original field names to the Dublin Core element set. To the left of each field line is a drop-down menu, listing all the Dublin Core elements (“DC Elements”). Choose the Dublin Core elements that best correspond to each of your field names. **Elements may be used more than once. A complete description of the Dublin Core element set can be found at <http://dublincore.org/documents/dcmi-terms>.**

If you need to revisit any of the previous steps, you may do so by clicking the appropriate “Step” button. Remember that you may only move backward, not forward, and that any information you have entered into later steps will be lost when you go back to an earlier step.

Click the “Next” button when you’re done to complete the migration process (Step 5).

Please map the field names to the corresponding Dublin Core elements.

DC Elements	Field Names	Record 1	Record 2
Discard <input type="button" value="v"/>	artist	Deffa, Chip	Finn, Julio
Discard <input type="button" value="v"/>	item	Reference book	Reference book
Discard <input type="button" value="v"/>	catalog	GMHF-LIB-19	GMHF-LIB-7
Discard <input type="button" value="v"/>	title_make	Voices of the Jazz Age	The Bluesman
Discard <input type="button" value="v"/>	label	----	----
Discard <input type="button" value="v"/>	no__serial	----	----
Discard <input type="button" value="v"/>	composer	----	----
Discard <input type="button" value="v"/>	producer_a	----	----
Discard <input type="button" value="v"/>	performer	----	----
Discard <input type="button" value="v"/>	location	Library	Library
Discard <input type="button" value="v"/>	year_induc	----	----
Discard <input type="button" value="v"/>	venue	----	----
Discard <input type="button" value="v"/>	city_count	----	----
Discard <input type="button" value="v"/>	connection	----	----

Derived Title Option :

Field Names No. of Characters

Figure 13: CROSSWALK screen

Discard Option

The default is for all fields to be discarded. Users must select a Dublin Core element in order for fields to be included in the migrated data set. If the user does not select a Dublin Core element from the drop-down menu, the corresponding field and all the corresponding data items of that field will *not* be included in the generated files, and hence cannot be viewed by the outside world. This feature lets you keep your sensitive metadata safe. **If a field contains sensitive information you do not wish to make publicly available, do not select a Dublin Core element for it.**

Derived Title Option

The derived title option is an optional feature. While many users may not need to use this option, others may find it useful as an alternate title source, particularly if their records are used to catalog many untitled works.

The derived title option lets you choose which field of data to use for the “Title” of an item. You might choose, for instance, to map the Dublin Core element “Title” to a “Description” field that provides lengthy, detailed information about an untitled object. Rather than use all the data from that field as the title, you can use the “Derived Title Option” to specify how many characters (letters, numbers, punctuation, etc.) are used to create the “title.” Choose a field, then indicate how many characters should be pulled from the data field (see Figure 14). Try to choose enough characters to make the title unique. This option is applicable to only one field.

Discard	location	Library	Library
Discard	-----	----	----
Discard	artist	----	----
Discard	item	----	----
Discard	catalog	----	----
Discard	title_make	----	----
Discard	label	----	----
Discard	no_serial	----	----
	composer		
	producer_a		
	performer		
	location		
	year_induc		
	venue		
	city_count		
	connection		

title Option :

No. of Characters

10

Next >>

Figure 14: Limiting the number of characters to be displayed for a field

When you are done, click the “Next” button to complete the migration process (Step 5).

Step 5: PRODUCE

Confirm

The CONFIRM screen, shown in Figure 15 below, asks you to finalize the creation of an OAI Data Provider. If you need to change the information you provided during any of the previous steps, you may do so by clicking the appropriate “Step” button. Remember that you may only move backward, not forward, and that any information you have entered into later steps will be lost when you go back to an earlier step.

[Help](#)

[Log Out](#)



Finalize the OAI data provider

PLEASE NOTE:

By clicking on the finalize link below you will complete the process of metadata migration. After this point you cannot go back and change options. If you would like to go back, please click on any of the first four steps.

[FINALIZE](#)

Figure 15: CONFIRM Screen

By clicking on the finalize link on the screen you confirm that you want to create an OAI data provider.

Produce

If your data was successfully migrated into Dublin Core XML, the PRODUCE screen displays the success message (see Figure 16, below). This page also provides the OAI Interface URL for the OAI repository explorer (which an OAI harvester may use to locate your records) as well as a URL where you can go to actually view your migrated records.



Data successfully formatted to DC format.

OAI Interface URL:

<http://metascholar.org/metadata/provider/test/XMLFile/051805mhalber/oai.pl>

Generated list of Records:

http://metascholar.org/metadata/provider/test/XMLFile/051805mhalber/oai.pl?verb=ListRecords&metadataPrefix=oai_dc

Figure 16: PRODUCE screen with success message

OAI Interface URL

The base URL points to the OAI repository where your formatted records are stored. This repository is located on the Metascholar server and can be accessed by an OAI harvester. Elements of the URL are drawn from the user-provided data: for example, in the URL above, “test” is the username (used to log in) and “051805mhalber” is the archive identifier (provided in Step 3 of the migration process). Use this address to register your archive with an OAI-repository registry.

Generated List of Records

The OAI Interface URL is only readable by harvesters. If you would like to visually review your records, click on the second URL listed on the PRODUCE screen for a display of your OAI-formatted data.

If there was a problem with the migration, the screen displays an error message. If you are having trouble migrating your records, or if you cannot open the link to view your migrated records, please contact the Metadata Migrator administrator (mdmsupport@metascholar.org).

Appendix: Data Files

Types of Data Files

Xbase Data File (.dbf) file

An Xbase, or dBase, data file is the central table in an Xbase database. All other data files are related to this one file. This data file format contains a mix of binary and ASCII data. The header contains binary data. The records are all in ASCII.

Comma-Separated Value (.csv) File

A CSV (comma-separated values) file contains the values in a table as a series of ASCII text lines, organized so that each column value is separated by a comma from the next column's value and each row starts a new line.

Tab-Delimited (.tab) File

A tab delimited file is a special kind of plain text file with a tab between each column in the text. When imported into the desktop publishing application, the tabs allow the columns to line up neatly.

Exporting Files as .csv, .tab, or .dbf

The following sections provide instructions for exporting your files into one of three formats readable by the Metadata Migrator Tool, specifically for two database software programs (Access and ProCite).

Access instructions for exporting to dBase (.dbf)

1. Open the database file you want to export.
2. Select the tables you want to export. [If there's more than one, repeat this process for each table.]
3. Under the "File" menu, select "Export."
4. From the menu of file formats, choose the newest dBase (probably III, IV, or V).
5. Name the file and save it to your hard drive. This is now the file you will want to choose for uploading into the Metadata Migrator Tool.

ProCite instructions for exporting to .csv or .tab format

1. Open the database file you want to export. [If there's more than one, repeat this process for each database file.]
2. Go to the "EDIT" menu on the main menu bar, and select "SELECT ALL" from its auxiliary menu. [This will highlight all of the contents of that database file.]

3. Below the menu bar and above your database file you'll see a submenu with little boxes that can be checked and unchecked. One of the options on the smaller menu bar is "MARK LIST": check the box beside it in order to mark all of the records that are highlighted.
4. Go to the "TOOLS" menu on the main menu bar and select "EXPORT MARKED RECORDS."
5. A box of options will pop up in the middle of your screen. The first option (probably the default) is "comma delimited" or "comma separated" files. If it is already the selected option, proceed to the next step. If this option is not already selected, choose "comma delimited" or "comma separated" from the drop-down menu at the top of the box, and then proceed to the next step. [All of the other information on the page will automatically be correct for comma-separated exporting.]
6. You will see two folder tabs at the top of the box: one says "Delimit Format"; the other says "Export Data." Click on the second tab to open the "Export Data" page. One of the options on this page is "Export Workform Definitions." If the box beside it is checked, UNCHECK this box before proceeding.
7. Now, click "OK" to begin the export. You will see a pop up message about "styles" being removed. Click "OK," and the export will take place.