

How to Harvest Metadata Records

You can harvest metadata records from repositories (in OAI-DC format) and spreadsheets (in CSV format) by enabling the optional Harvester Module (Administer > Modules > Islandora Tools > Islandora Harvester). The module will parse the source file and create a new object in your repository for each item in the metadata record.

The source file must use headers that map to Dublin Core.

The following tutorial will show you how to harvest items from OAI-DC sources. To harvest from a CSV file, simply ensure that the column headers map to Dublin Core, and follow steps 3-5 of the following tutorial (replacing the OAI-DC request with your CSV file).

How to Harvest OAI-DC Metadata Records

OAI-DC is a metadata format supported by the OAI-PMH ([Open Archives Initiative Protocol for Metadata Harvesting](#)) framework. Repositories can choose to expose their metadata using this framework, and harvester tools can make requests to download some or all of these metadata records. Once you have a base OAI-DC request for a collection, you can follow these steps to harvest the records into your repository.

1. Construct a Valid OAI-DC Request

All OAI-DC requests must have the following components:

1. **A base URL:** This includes the internet host and port, and (optionally) a path, of the repository.
2. **One or more keyword arguments:** These take the form of *key=value* pairs.

For more information on OAI-DC requests, please see the [OAI-PMH documentation](#).

In this tutorial we'll be using a request for metadata records from Dalhousie University: http://dalspace.library.dal.ca:8080/oai/request?verb=ListRecords&metadataPrefix=oai_dc

The base URL is <http://dalspace.library.dal.ca:8080/oai/request>.

The arguments are *verb=ListRecords* and *metadataPrefix=oai_dc*.

This request will harvest all available OAI-DC metadata records. However, you can limit the request in various ways; for example, by set.

2. Limit the Request to a Specific Set (Optional)

Navigate to http://dalspace.library.dal.ca:8080/oai/request?verb=ListRecords&metadataPrefix=oai_dc to view the full list of metadata records in XML. You can limit this request by, for example, choosing one of the available sets:

```
<setSpec>hdl_10222_11202</setSpec>
```

You can add this set to your request as an argument like this:

http://dalspace.library.dal.ca:8080/oai/request?verb=ListRecords&metadataPrefix=oai_dc&set=hdl_10222_11202

This will produce a list of records limited to that particular set.

3. Create a Collection to Harvest Records Into (Optional)

You can create a new collection to harvest metadata records into by using the Collection Manager. For detailed instructions, see [How to Create a New Islandora Collection](#).

Alternatively, you can simply harvest the metadata records into an existing collection.

4. Harvest Metadata Records into the Repository

Click the 'Harvest Items for this Repository' link (or navigate to <http://your.site/islandora/harvest/>). Choose 'OAI-DC' from the 'Source Type' selection and enter the request URL: http://dalspace.library.dal.ca:8080/oai/request?verb=ListRecords&metadataPrefix=oai_dc&set=hdl_10222_11202.

Use the drop-down menus to select the collection you want to harvest into and the content model to associate with the newly created objects.

The screenshot shows a web form titled "Harvest Items for Repository" with a "Home" link. Under the "Ingest Metadata" section, there are two main parts: "Source Data" and "Destination".

Source Data:

- Source type:** Two radio buttons are present: "CSV" (unselected) and "OAI-DC" (selected).
- Harvest from an external repository (OAI Set):** A text input field contains the URL "http://dalspace.library.dal.ca:8080/oai/request?verb=ListR". Below it is a note: "Enter full url to retrieve entire OAI resultset."

Destination:

- Add to collection:** A dropdown menu shows "Harvester Test Collection".
- Choose content model:** A dropdown menu shows "Islandora strict PDF content model".
- Submit:** A button at the bottom left.

Annotations (Red boxes with arrows):

- A box labeled "Copy the OAI-DC request URL into this field" points to the URL input field.
- A box labeled "Choose a collection to harvest into" points to the "Add to collection:" dropdown.
- A box labeled "Choose a content model for the new objects" points to the "Choose content model:" dropdown.

While you will only be harvesting metadata records at this point, you can always add datastreams to the objects (such as images and PDFs). Choosing a relevant content model will make this process easier in the future.

Click 'Submit' to begin harvesting metadata.

5. Verify Your New Objects


Navigate to the collection you chose to harvest into and verify that your new objects were created successfully. You should see a new object for each metadata record in the OAI-DC request.

[Home >](#) [Digital repository >](#) [Harvester Test Collection >](#)


Digital Repository

[Add](#) [View](#) [Object Details](#) [Manage This Collection](#)


1-12 of 12




[Contributions to the ichthyology of Nova Scotia](#)




[Gold and its separation from other minerals](#)




[Inaugural Address](#)



[Notice of the occurrence of a trilobite in the lower carboniferous limestone of Hants Co.](#)



[On magnesia-alum, or pickerinigte, containing a little nickel and cobalt, occurring in slate in Hants Co.](#)



[On the characteristic different coal seams i](#)