Islandora OAI

Overview

The Islandora OAI module (based on the [oai2forcck Drupal module](#)) provides support for a site to be visible via the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). In short, a site properly configured using this module has its Solr index - and accompanying metadata - visible to other sites that harvest OAI-compatible metadata. These harvesters make various types of requests at a URL that you can specify, and your site responds with metadata information that they in turn can add to massive archival indices. This makes it much easier for researchers to find objects on your site.

For more information on the OAI-PMH, you may consult the official documentation at http://www.openarchives.org/OAI/openarchivesprotocol.html.

Dependencies

- [Islandora Solr](#)
- [Islandora Core Module](#)
- [Tuque](#)

Besides installing the Islandora Solr modules, you will also need to correctly configure Solr and GSearch in order for Islandora OAI to work. The OAI module passes information to metadata harvesters based on results it finds from your Solr index; if Solr is not properly configured, OAI won't function either.

Downloads

[Release Notes and Downloads](#)

Usage

Islandora OAI works mostly autonomously. It gets requests from metadata harvesters in the form of HTTP POST keys that come after your OAI URL. Your site then sends back information, in XML format, based on the values of the keys that were given. You can check that your configuration is correct by manually entering these keys in your browser's address bar, and seeing what comes back.

A simple check you can run involves asking your OAI URL for a list of information about your repository. To do this, you will need to know a few of your site's OAI configuration options. More information on this can be found in the next section of this page.

To check for the first few records, use your browser to access the following site:

> http://**path.to.your.site**/**repository**?verb=**Identify**

Where:

- **path.to.your.site/repository** is the URL found on the Islandora OAI configuration page, in the Configuration section, under 'The path of the Repository'
- **Identify** is a verb that is designated by the OAI-PMH to return basic configuration information about your OAI metadata repository.

If your Solr Index is set up correctly, and you entered the URL properly, you should see an XML file containing information about your OAI setup.

Configuration

Configuration options for the Islandora OAI module can be found at http://**path.to.your.site**/admin/islandora/tools/islandora-oai and include the following options:

Configuration

- **Repository Name** - The name that harvesters will attach to metadata pulled from your repository.
- **The path of the Repository** - The URL that harvesters will make requests at.
- **Repository unique identifier** - The middle section of the identifier used when metadata harvesters pass the **identifier=** key. With this in place, an identifier for each of your objects' metadata will be generated as **oai:unique_identifier:namespace_pid**.
- **Admin Email** - An optional email address to be attached to harvested metadata
- **Maximum Response Size** - The maximum number of records that will be issued per response. If the number of records requested exceeds this number, Islandora OAI will also issue a 'resumption token', which the harvester can use to issue another request from the point they stopped at. This method is used to control flow and prevent servers from diverting too much resources to metadata harvesters.
- **Expiration Time** - The amount of time, in seconds, before a resumption token should expire.
- **OAI Request Handlers** - Handlers available to be configured

CONFIGURATION

Repository Name *

Repository

Path to the Repository *

http://localhost:8000/ oai2

The path where the OAI-PMH service will respond, e.g. http://localhost:8000/oai2

Repository unique identifier *

drupal-site.org

The identifier for this repository, e.g. oai:**drupal-site.org**:123.

Administrator Email

N/A

Maximum Response Size

20

The maximum number of records to issue per response. If the result set is larger than this number, a resumption token will be issued

Expiration Time

86400

The amount of time a resumption token will remain valid, in seconds. Defaults to one day (86400s).

Query Backend

◉ SPARQL

○ Solr

For larger repositories, OAI may perform poorly when attempting to perform the SPARQL queries it requires. In these cases, using the Solr backend may provide better results.

Select an OAI request handler

DEFAULT	LABEL	DESCRIPTION	CONFIGURATION
◉	Islandora OAI	Provides a standard OAI implementation for Islandora.	configure

Preferred OAI request handler for Islandora. These may be provided by third-party modules.

Save configuration

If you select "configure" you see the following screens.

- **Solr date field** - A datestamp to be appended to the metadata via the Solr index.
- **Solr RELS-EXT collection field** - Fields entered here establish the object relationship of metadata to be passed on to the harvester.
- **Solr XACML role field** - The site's Solr fields defining viewing permissions.
- **Solr hasModel field** - The site's Solr field defining an object's content model.
- **Exclude Content Models** - A list of content models, defined by their PID, to exclude from harvests.

CONFIGURATION

Solr Date Field

`fgs_lastModifiedDate_dt`

The Solr field containing the date to be used.

Solr RELS-EXT Collection Field

```
RELS_EXT_isMemberOfCollection_uri_ms
RELS_EXT_isMemberOf_uri_ms
```

The Solr fields used to determine what collection, if any, an object is a member of.

Solr Content Model Field

`RELS_EXT_hasModel_uri_ms`

Field which RELS-EXT datastreams use to define an object's content model.

Excluded Content Models

By default, all objects are visible to OAI metadata harvesters. This field allows you to exclude all objects with a certain content model, e.g "islandora:collectionCModel" to exclude all objects with the Islandora Core Collection content model. Content models are separated by line. NOTE: If islandora:collectionCModel is added, it will break the ListSets verb.

☐ Exclude objects within the "islandora" namespace?

Excluding the Islandora namespace is only possible when namespace restrictions are enabled within the Islandora module.

☑ Append on dc.identifier.thumbnail to OAI_DC requests?

If this option is selected, a link to an object's thumbnail will be added to OAI_DC responses.

Below this is the "Metadata Format" section:

Metadata Format

This section allows you to configure the settings for the OAI-PMH'smetadata_prefix verb; Islandora uses XSL files to define the method for transforming your site's metadata datastreams into a format compatible with the OAI-PMH. Islandora OAI comes with two XSL files; they convert the MODS datastream of an object to either Electronic Thesis and Dissertation Metadata Standard format or Dublin Core format, which then can be served up to a harvester.

- **Metadata Format** - The metadata format you would like to use. This will change the next three fields.
- **Metadata Prefix** - The default variable for the **metadata_prefix** verb.
- **Metadata Namespace** - The URL that contains XSD files defining the Metadata Format
- Schema Location - The actual XSD file in the Metadata Namespace that defines the Metadata Format.

Transformations - This section allows you to configure the way Islandora converts your metadata datastreams into a format compatible with the OAI-PMH.

- **Metadata Datastream ID** - The datastream ID where object metadata is stored (MODS by default).
- **File to use for transforming** _____ - The XSL file used to convert that datastream into a metadata format OAI will recognize and use.
- **Upload a file** - If you want to run custom conversions from a different datastream or to a different Metadata Format, you can upload these here.

METADATA FORMAT

Metadata Format

| MODS | ⬍ |

Metadata Prefix
mods
Default metadata prefix for the selected format.

Metadata Namespace
http://www.loc.gov/mods/v3
Default metadata namespace for the selected format.

Schema Location
http://www.loc.gov/standards/mods/v3/mods-3-4.xsd
Default URI for the selected metadata format's schema.

☐ Force include a link to the object within Islandora?
This is used in cases where metadata may not have links or Handles that point back to the object in the repository. Services like WorldCat expect a linkback to the object. This functionality can be achieved using XSLTs as well.

TRANSFORMATIONS

Metadata Datastream ID

| DC |

(Note that this is case-sensitive)

File to use for transforming mods

| No transformation selected | ⬍ |

XSL or XSLT file used to translate existing metadata to an appropriate OAI-PMH format.

File to use for self transforming mods

| No transformation selected | ⬍ |

XSL or XSLT file used to transform xml prior transforming to an appropriate OAI-PMH format.

[Save configuration]

After you have exposed content types and some fields, your repository is available at /oai2

Some example requests are as follows:

- `*/oai2?verb=Identify`
- `*/oai2?verb=ListMetadataFormats`
- `*/oai2?verb=ListIdentifiers&metadataPrefix=oai_dc`
- `*/oai2?verb=ListRecords&metadataPrefix=oai_dc`

Services like WorldCat expect links back to the object such as a Handle URL. If your metadata doesn't have this there are two approaches that can be used. Self transforming XSLTs can be used to add specific elements tailored to individual needs. However, there is options in configuration to append on URL values to the XML output of OAI. Each metadata prefix has an individual set of configuration. If selected, a user can define where the object URL will get appended in the output returned.

Similarily, OCLC's Digital Collection Gateway can take advatange of thumbnail URLs for rendering. This option is only currently available for oai_dc requests. If selected, a URL to the object's thumbnail will be added as a dc:identifier.thumbnail if the object has a thumbnail.

If existing content has already been harvested and/or the url and thumbnail are not mapping in Digital Collection Gateway, you will need to map those manually in the 'Metadata Map' for a given collection/set.

If the XACML module is present you will need to configure the `rels.isViewableBy` fields in the admin page such that the OAI requests respect these object restrictions.

The responses generated by this module have been validated against Open Archives' Validation.

Customization

By default the vanilla islandora_oai module provides a very basic output. It is possible to add additional content to the description field of the repository. This includes pointing at other harvesters and repositories, branding information etc. An example of how to implement these can be referenced in the 6.x version of the module.

Notes

The original 6.x version of this module was based off of the OAI2ForCCK module located here.