## Islandora Scholar

•

Introduction Requirements Installation Scholar Core

- Citation & Thesis Content Models
- Sherpa/RoMEO

Citation Collection & Citation Style Management

Generating citations

Google Scholar Integration Scholar Embargoes

- Community Best Practices for Embargoing
- Issues with Object Embargoes
- PDF Configuration

### Introduction

Islandora Scholar is a suite of modules designed to help Islandora function as an Institutional Repository (although some features are helpful in other use cases as well). It is unique compared to other Islandora modules in terms of the number of features it provides as well as the vast amount of submodules that it contains. It is helpful to think of Scholar as being a kind of scholarly content solution pack due to the new content models it provides (citation & thesis), but it also differs from other solution packs in that it provides new functionality that may be used with other cmodels as well.

#### Sample Features

- · New citation & thesis content models
- Sherpa/RoMEO integration
- Creation of new objects from EndNote XML or RIS files exported from other systems
- Creation of new objects from Digital Object Identifiers (DOIs) and PubMed IDs (PMIDs)
- Suppression of objects from display or disable viewing/downloading of particular datastreams
- · Exporting of collections of bookmarks as RIS, RTF or PDF files
- User-selectable dynamic citation styling using Citation Style Language (CSL) files
- · Extra HTML metadata to assist with Google Scholar indexing

# Requirements

This module requires the following external modules/libraries:

- Islandora
- Tuque
- Islandora Solr Search

Islandora Scholar also requires the Citeproc, CSL and Bibutils modules in order to be enabled, but these modules are included in the /modules directory of Islandora Scholar module.

Please note that that enabling Scholar does not automatically enable the various Scholar submodules. They must be enabled separately, and they have their own separate requirements as well.

### Installation

See this for further information about standard installation of Drupal 7 modules.

It is necessary to install the citeproc-php library into the sites/all/libraries directory, such that the mainCiteProc.php file is located at sites/all/libraries/citeproc-php/CiteProc.php.

Also, the included Bibutils submodule only provides PHP interface to the Bibutils tool itself. Bibutils must be installed separately on the underlying operating system. Follow these instructions for installing Bibutils.

## Scholar Core

While "Scholar Core" is not an official community term, it can be a helpful way to keep track of the numerous features Islandora Scholar offers. In this context, "core" features are those that come from the islandora\_scholar.module file itself as opposed to coming from one of the numerous submodules which can be enabled or disabled independently of the main Islandora Scholar module.

#### Citation & Thesis Content Models

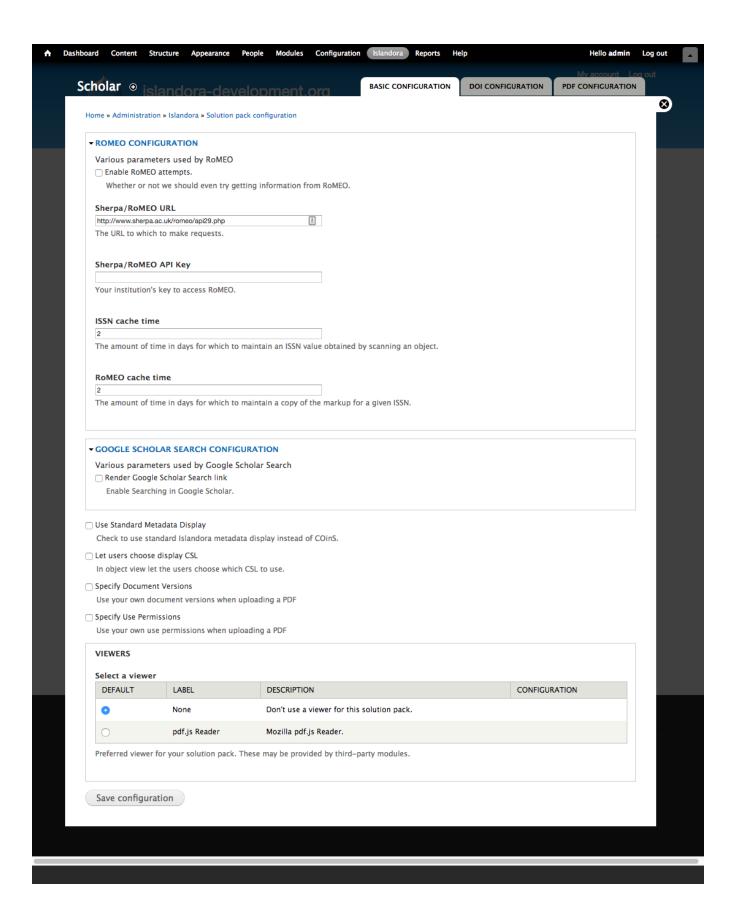
The core Islandora Scholar module provides two new content models: the citationCModel and the thesisCModel. The citation cmodel is intended for general scholarly works, while the thesis cmodel is for handling electronic theses and dissertations (ETDs). While the datastream structure and default display of these two cmodels is nearly identical, they can be tweaked separately to accommodate the different use cases one may have for theses vs. other scholarly works, such as altered displays or separate metadata forms.

Please note that the citationCModel and thesisCModel are in the "ir" namespace, as opposed to the standard "islandora" namespace that other cmodels have had in the past. Due to this nonstandard namespace, Islandora instances that make use of namespace restrictions will need to enable the "ir" namespace in order for the citation & thesis cmodels to work.

### Sherpa/RoMEO

Scholar Core includes Sherpa/RoMEO integration. Sherpa/RoMEO is a service which keeps on file and makes searchable the copyright & self-archiving policies of various academic journals.

Configure this in the Scholar admin menu at 'admin/islandora/solution\_pack\_config/scholar'. Checking "Enable RoMEO attempts" turns on the functionality. When that is checked, then when viewing a Citation Content Model object when the object has a MODS identifier of type "issn", the person viewing the object will see a tab labeled "RoMEO" which shows the journal policies pulled from Sherpa RoMEO.



There is also a place in the admin menu at 'admin/islandora/solution\_pack\_config/scholar' to provide a Sherpa/RoMEO API key. No API key is needed. Instead, there's a cap of 500 requests per day if you don't have an API key, but no cap if you do. API registration is free-of-charge as of summer 2016.

Rights information from Sherpa/RoMEO is not copied into object metadata, nor into any datastream. Instead, this is a quick link to the Sherpa/RoMEO information which can be used in staff workflows.

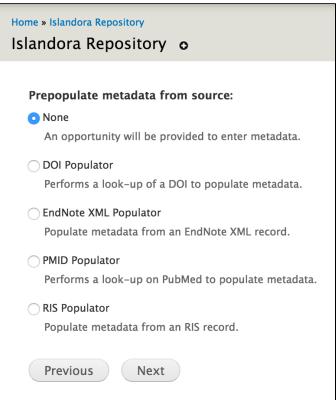
# **Importers**

Scholar provides options for importing objects from various sources.

These are configured by enabling the following modules:

- PMID Populator (imports metadata through PubMed's API)
- RIS Populator (when creating a Citation object, allows upload of a Research Information Systems (RIS) formatted citation to prepopulate metadata)
- DOI Populator (imports metadata through Crossref OpenURL)
- EndNote XML Populator (when creating a Citation object, allows upload of an EndNote XML exported citation to prepopulate metadata) (each module enables import from a different source)

When a module is enabled, and someone clicks to create an object in Islandora, Islandora will display the option to "Prepopulate metadata from source". The person can choose a source, and either upload the file or provide the identifier, and then will be shown a metadata input form prepopulated with metadata from the source.





Here is more detailed info on each source for metadata:

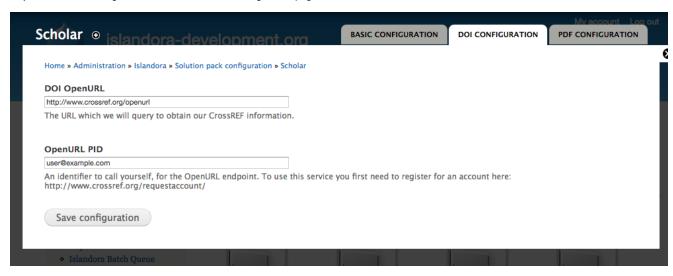
The RIS Importer and EndNoteXML Importer submodules allow users to take citation data files exported from other sources (such as RefWorks, EndNote or Zotero) and turn them into Islandora objects using the standard Islandora importer interface (similar to using the zip importer). The DOI Importer and PMID Importer submodules work in much the same way, but instead of using files exported by other citation managers they use Digital Object Identifier (DOI) or PubMed ID (PMID) strings.

DOI Populator: Registration with Crossref required.

To use this, enable the "DOI Importer" module.

You must register to use Crossref OpenURL. You register with an email address and enter the email address into Islandora in the menu at Home » Administration » Islandora » Solution pack configuration » Scholar . There is no fee to register for this service. (This is an API for inputting a DOI and retrieving metadata in XML. Islandora Scholar Core does not interface with Crossref in any capacity which would allow minting of DOIs.)

Digital Object Identifiers Configuration tab with Cross Ref configuration page.



For more information about Crossref Open URL or to register, click here.

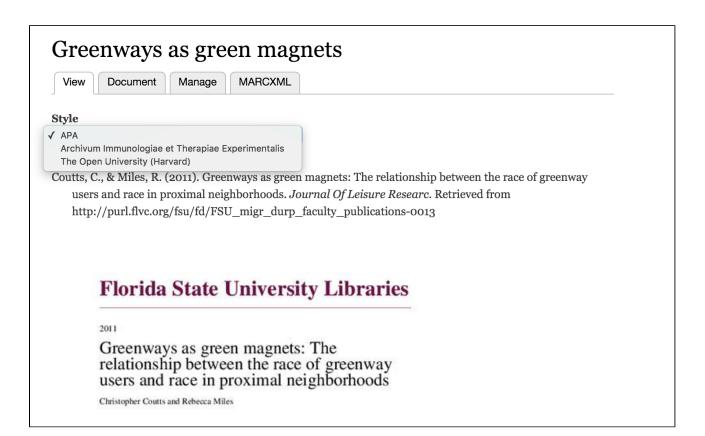
# Citation Collection & Citation Style Management

The Islandora Bibliography submodule extends Islandora Bookmark to allow it to handle collections of scholarly citations, which can be dynamically restyled with the help of the Citeproc and CSL submodules. They can also be exported in RIS, RTF or PDF format with the Exporter submodule.

In order to access new citation styles, go to the Zotero CSL Repository and download whatever CSL files you want to use. You can then upload your CSL files using the CSL admin page at admin/islandora/tools/csl and choose a default style to display on your citation & thesis object pages.



If you want your users to be able to dynamically choose any available citation style from the citation/thesis object display interface, make sure to check the "Let users choose display CSL" checkbox at the bottom of the Islandora Scholar admin page at admin/islandora/solution\_pack\_config/scholar.



## Generating citations

The appropriate type of citation will be displayed according to the content of the <genre> field in the MODS record for each Citation or Thesis object. In the default form, the "Publication Type" entry box near the top of the form is used to enter the MODS <genre>. A case-sensitive entry is required in order to generate the correct formatting:

MODS <genre> term</genre>	Format of the citation generated by Islandora Scholar
blank/default/ any-term-not- in-a-controlled -vocab	Citation will be formatted like a journal article.
<genre>journal article</genre>	Citation will be formatted like a journal article.
<genre>book chapter</genre>	Citation will be formatted like a book chapter.
<genre>book section</genre>	Citation will be formatted like a book chapter.
<genre>book</genre>	Citation will be formatted like a book.

# Google Scholar Integration

Getting indexed by Google Scholar is a primary concern for any institutional repository due to its overwhelming popularity with researchers. A properly configured IR should be seeing most of its visitors coming from Google Scholar, but Google Scholar has many requirements for repositories who want to be indexed. The Islandora Google Scholar submodule pulls relevant information from the MODS record of citation & thesis cmodels and turns it into meta tags and embeds it into the HTML of the display page which is the primary way that Google Scholar learns about your repository and its contents. See the meta tags with the "citation\_" prefix in the screenshot below:

# **Scholar Embargoes**

The **Scholar Embargo** submodule is very similar to (and can even be used in conjunction with) the Islandora IP Embargo module. Scholar Embargo allows you to suppress the display of objects in collections entirely, or allow the objects to display but disable a specified datastream on an object. Embargoes may be indefinite (meaning that they will persist until manually lifted) or temporary (meaning that they will persist until the specified expiration date chosen by an administrator, with email notifications 10 days before and on the day of an embargo's expiration).

The two primary types of embargoes are object embargoes and datastream embargoes. Objects with an "object embargo" will behave as if they don't exist for public users; they won't display in collections or search results, and will yield an access denied page when public users attempt to access them directly by PID. Only privileged users will be able to see them (more on that below). Datastream embargoes allow the object to display normally in collections and search results, but a particular datastream (specified when the embargo is placed) will be unavailable to public users to view or download. For citation and thesis objects this will be the 'PDF' datastream, but Scholar Embargoes can be applied to any content model, extending the list of possible datastreams. In order to put a Scholar Embargo on content models other than citations and theses, you must enable them from the "Embargo Settings" tab of the Islandora Scholar Embargoes admin menu at admin/islandora/solution\_pack\_config/embargo (example below).

PID	CONTENT MODEL
ir:citationCModel	Citation Content Model
islandora:collectionCModel	Islandora Collection Content Model
islandora:sp-audioCModel	Islandora Audio Content Model
islandora:sp_basic_image	Islandora Basic Image Content Model
islandora:pageCModel	Islandora Page Content Model
islandora:bookCModel	Islandora Internet Archive Book Content Model
islandora:compoundCModel	Islandora Compound Object Content Model
islandora:sp_large_image_cmodel	Islandora Large Image Content Model
islandora:sp_web_archive	Islandora Web ARChive Content Model
islandora:sp_videoCModel	Islandora Video Content Model
ir:thesisCModel	Thesis Content Model
islandora:sp_pdf	Islandora PDF Content Model
islandora:newspaperCModel	Islandora Newspaper Content Model
islandora:newspaperIssueCModel	Islandora Newspaper Issue Content Model
islandora:newspaperPageCModel	Islandora Newspaper Page Content Model
islandora:OACCModel	Islandora Annotation CModel
islandora:oralhistoriesCModel	Islandora Oral Histories Content Model

You may also use the "Manage Embargoed Items" tab in the Islandora Scholar Embargo admin menu to see all objects that are currently under embargo.

LABEL	PID	TYPE	EMBARGOED UNTIL	MANAGE
Citation – Temporary PDF Embargo	fsu:6965	PDF	Apr 08, 2036	Manage
Thesis – Indefinite Object Embargo	fsu:7270	Object	Indefinite	Manage
Thesis – Temporary PDF Embargo	fsu:7273	PDF	Apr 08, 2034	Manage
Thesis – Indefinite PDF Embargo	fsu:7274	PDF	Indefinite	Manage
Thesis – Temporary Object Embargo	fsu:7278	Object	Apr 08, 2036	Manage
Citation – Temporary Object Embargo	fsu:7969	Object	Apr 08, 2036	Manage
Citation – Indefinite Object Embargo	fsu:7982	Object	Indefinite	Manage
Citation – Indefinite PDF Embargo	fsu:8006	PDF	Indefinite	Manage

### Community Best Practices for Embargoing

It is worth noting that the current best practice in the world of institutional repositories (citation needed) seems to be exposing the object display and metadata, but blocking the viewing or downloading of the full text record (typically a PDF file). This is so users can know what information is in the system through collection displays or search results, even if the system blocks them from fully accessing it. Many faculty and students like to check to make sure their records are in the repository and still embargoed, and when they can't find their record they may misunderstand and think it hasn't been loaded.

### Issues with Object Embargoes

In order to place embargoes on objects, you must grant the "Manage embargo on any objects" permission to the appropriate user roles under the Islandora Scholar Embargo section of admin/people/permissions. This will allow a user to place object and datastream embargoes on objects, as well as remove or update datastream embargoes. Due to issues with the connection between Drupal permissions and the way Scholar Embargo implements its embargo policies through XACML, <u>currently only the object's owner will be able to remove or update object embargoes</u>. This is because the XACML policy works independently of Drupal permissions, and the embargo XACML policy will block everyone except the object's owner from managing the object at all once embargoed regardless of what Drupal permissions they have. A fix for this is in the works and should be part of the Islandora 7.x-1.8 release.

## **PDF** Configuration

Text extraction relies on the text stream being embedded into the pdf. This isn't going to generate OCR for the book if it wasn't already done. Consider converting text-filled images with no text streams to TIFFs and using the Book Solution Pack with OCR enabled.

