Chapter 4 - Search & Discovery in Islandora

On this page:

- What is Solr? And why does Islandora use it?
- Solr “Out of the Box”
- Solr Configuration
  - Solr Blocks
  - Initial Solr Set Up
    - 1. Navigate to Solr Configuration Panel
    - 2. Verify the Solr URL and Request Handler
    - 3. Set the Default and Secondary Display Profiles
    - 4. Choose Search Terms in Advanced Search Block
    - 5. Choose Facet Fields
    - 6. Choose Sort Fields
    - 7. Set Query Defaults
- Accessing Solr outside of the Islandora User Interface

Islandora uses Solr, in combination with Fedora's GSearch, to provide search functions to users on your site. This chapter assumes either:

a) You are using the **Virtual Machine Image**, or are exploring Islandora via Sandbox (where Solr is already installed and configured)

b) Solr and Gsearch are installed (following the instructions in Chapter 8 - Installing Solr and GSearch), and you have installed and activated the Solr module. The xslt and solrschema.xml documents that come packaged with GSearch should be used in configuration. These files are designed to work with our solution packs. Guidance for creating customizations to Solr are provided in the Customizing GSearch and Solr section.

What is Solr? And why does Islandora use it?

Solr makes it easy to create advanced search features in Islandora, like faceting (arranging search results in columns with numerical counts of key terms). The following comes from the Solr guide; a link to the guide is provided in the Selected Reading section of this guide:

> *Solr builds on another open source search technology - Lucene, a Java library that provides indexing and search technology, as well as spell-checking, hit highlighting and advanced analysis/tokenization capabilities. Both Solr and Lucene are managed by the Apache Software Foundation (www.apache.org).*
>
> *The Lucene search library currently ranks among the top 15 open source projects and is one of the top 5 Apache projects, with installations at over 4,000 companies. Lucene/Solr downloads have grown nearly 10x over the past three years, with a current run-rate of over 6,000 downloads a day.*

Islandora uses Solr to make objects in your Islandora installation discoverable. The Solr search module uses an xslt in Gsearch to index the **FOXML** documents in your repository, and allows you to configure search fields for searching and faceting. Whenever you add a new object in Fedora, the Solr module updates your index, and makes those results available to your users.

To Recap:

- Solr makes digital assets in your Islandora installation discoverable. Solr helps to enable the display/searching of your digital assets' metadata.
- To index Fedora content in Solr, Islandora currently relies on Fedora's Gsearch.
- When going through the Islandora installation process, there is an xslt and solrschema.xml document that come packaged with GSearch that can be used out-of-the-box or customized to suit your repository's needs.
- GSearch uses the xslt to transform a Fedora FOXML document into a Solr xml document.
- solrchema.xml controls how Solr indexes xml files.
- FOXML files are simple xml files that directly expresses the **Fedora Digital Object Model**. FOXML is similar to METS as it is basically an xml container.

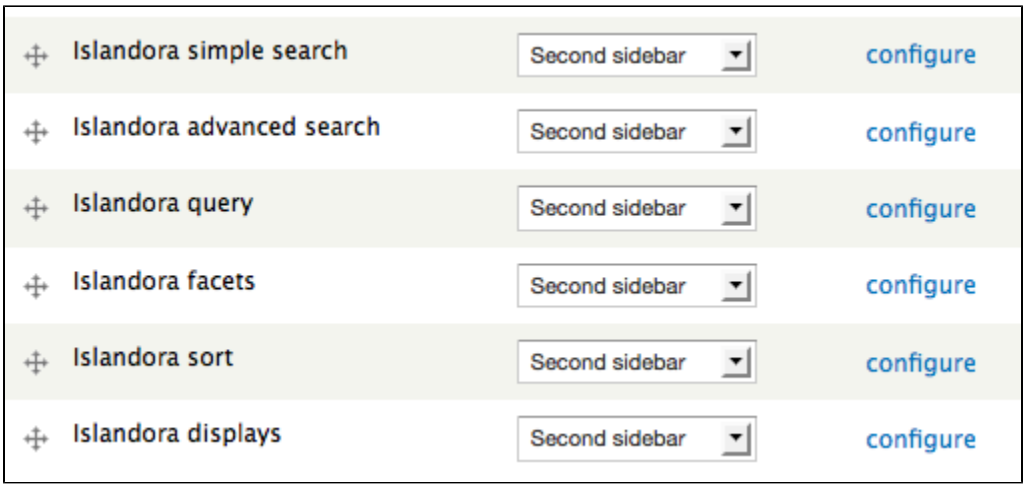
Solr “Out of the Box”

The Solr module comes with support for simple Dublin Core and simple MODS (i.e. no qualified fields!) searching and faceting to support our solution packs.

Solr Configuration

Solr Blocks

When Solr is installed, six additional blocks appear under your site's Structure>Blocks section:



Islandora simple search - provides a simple full-text search of all items in the repository. A term will appear, no matter where it appears in the metadata.

Islandora advanced search - provides a configurable search for users, where specific metadata fields can be searched and combined with boolean operators.

Islandora query - when a user is viewing a set of search results by facet, this block will show the current filters being applied in the search.

Islandora facets - this block will show users the facets they can use to browse and refine search results.

Islandora displays - displays current search query

Islandora sort - displays sorting options for search results.

By placing these blocks in regions of your site and configuring the Solr module under Islandora>Solr client, you can facilitate both full-text searching and faceted searching of items ingested using Solution Packs. The basic configuration of Solr can be modified to change the weight of search fields, and extend the out-of-the-box functionality.

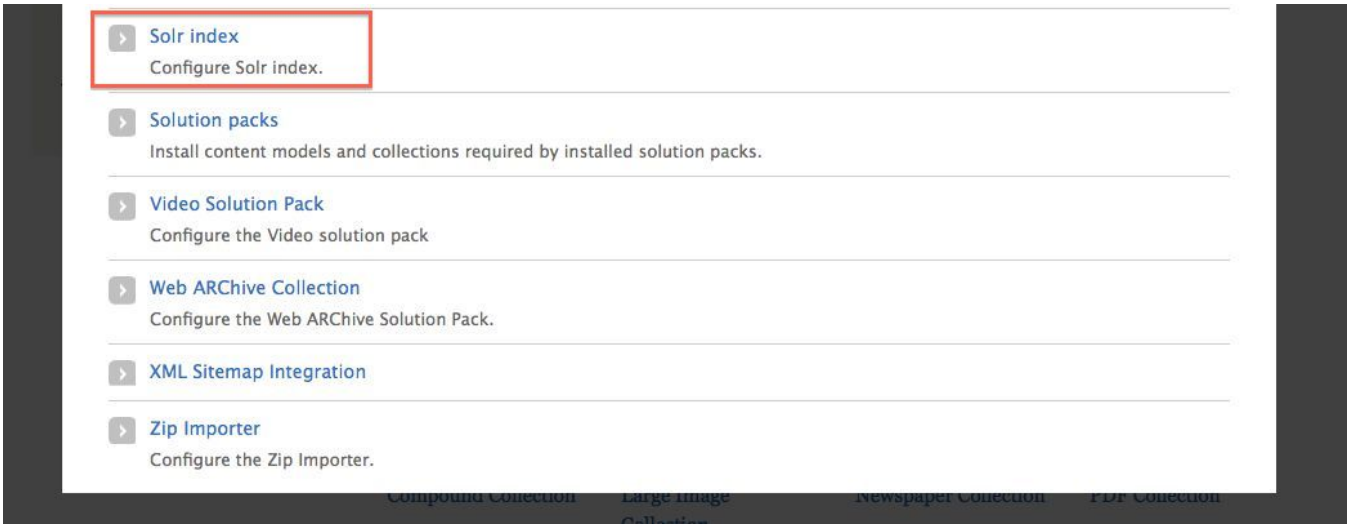
Initial Solr Set Up

In order for Solr to work for your collections, you will have to activate the blocks that you want (see above), and configure them to display your desired results.

The following instructions will show you how to configure the initial Solr set up in your Islandora instance. Additional information about installing Solr and how Islandora uses Solr is provided in Installing Solr and GSearch.

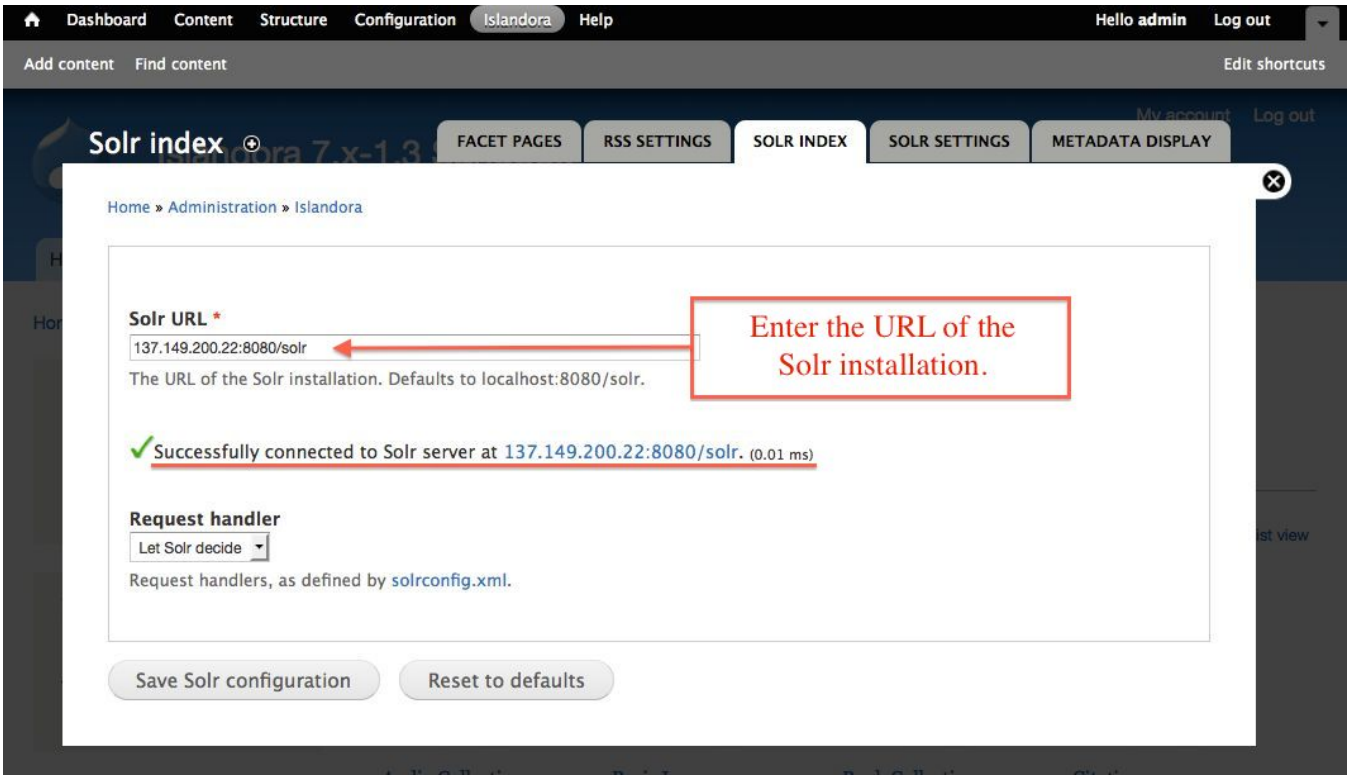
1. Navigate to Solr Configuration Panel

Start by going to the 'Islandora' page in the admin panel and click on the 'Solr Index' link.



2. Verify the Solr URL and Request Handler

The Solr URL should be *ip.address.of.site:port*/solr. So, for example, if you're using the Islandora Virtual Machine Image or another a local installation it should be localhost:8080/solr. If the URL is correct you will see a green check mark.

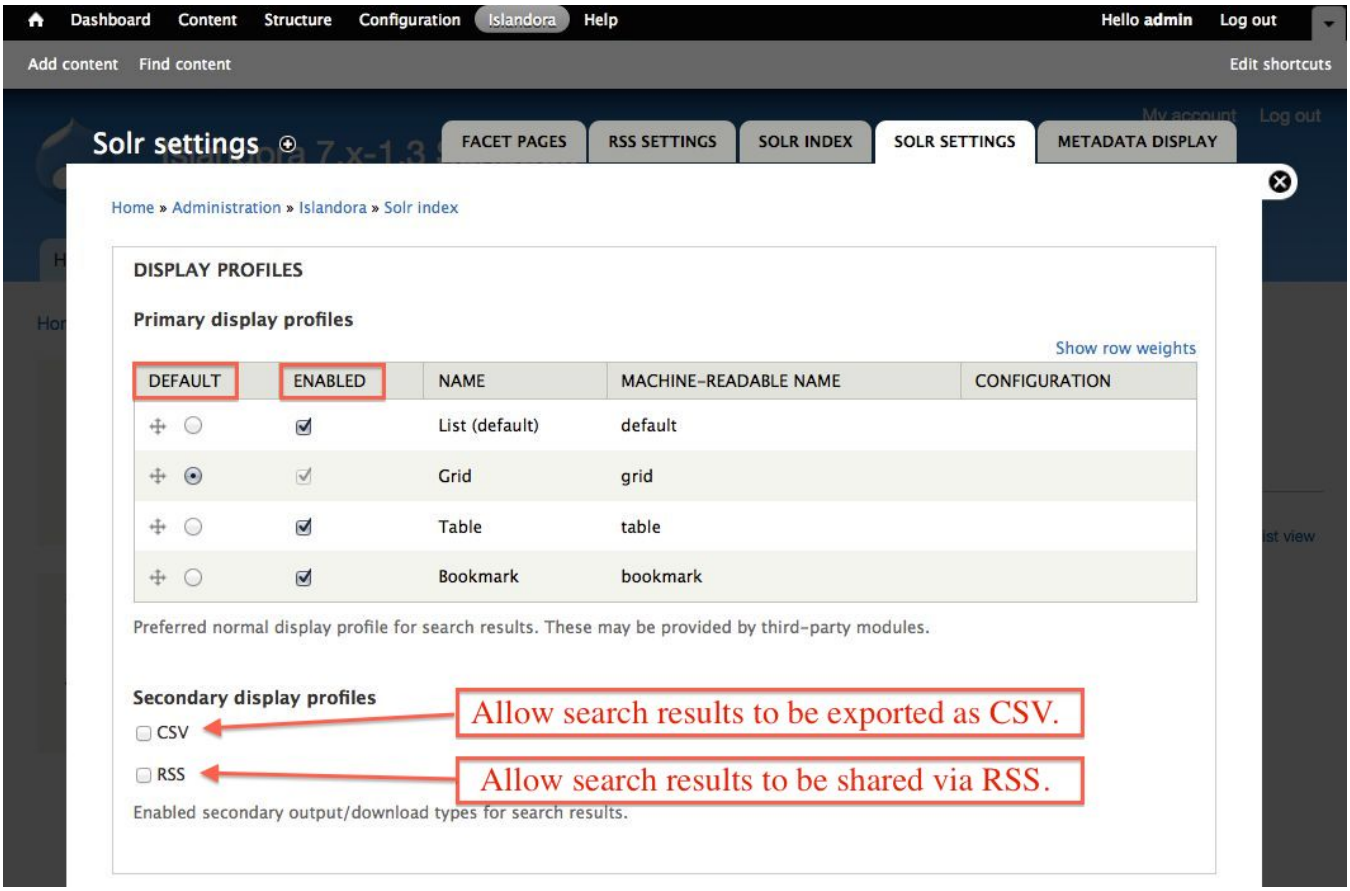


Make sure the request handler is set to ‘standard’. You can customize the request handler by editing *solrconfig.xml* to make other request handlers available. This assumes that you are using the default request handler, which will support all the metadata in our solution packs.

3. Set the Default and Secondary Display Profiles

You can modify the way search results are displayed by configuring the Display Profiles. You can choose from List, Bookmark, Grid (set to Default in the screenshot below), and Table. Secondary display profiles provide optional secondary outputs for search results. Switching back and forth between different display profiles is simple, so feel free to experiment and see which default display profile best suits your site. The display profiles appear on your site through the "Islandora displays" block, mentioned above.

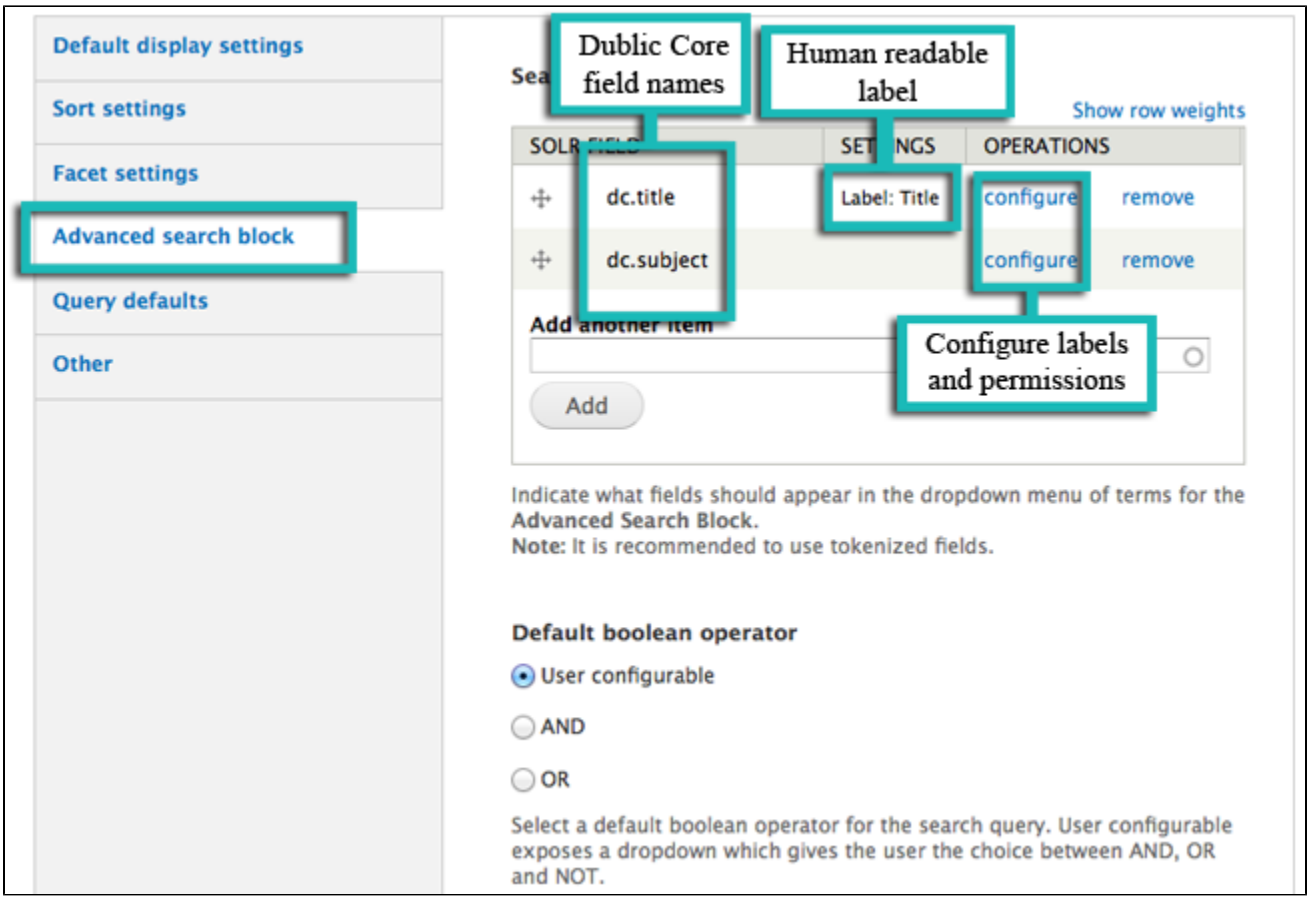
Out-of-the-box, support for RSS and CSV output is also provided. Selecting these options will place an RSS feed and CSV button next to your search results.



4. Choose Search Terms in Advanced Search Block

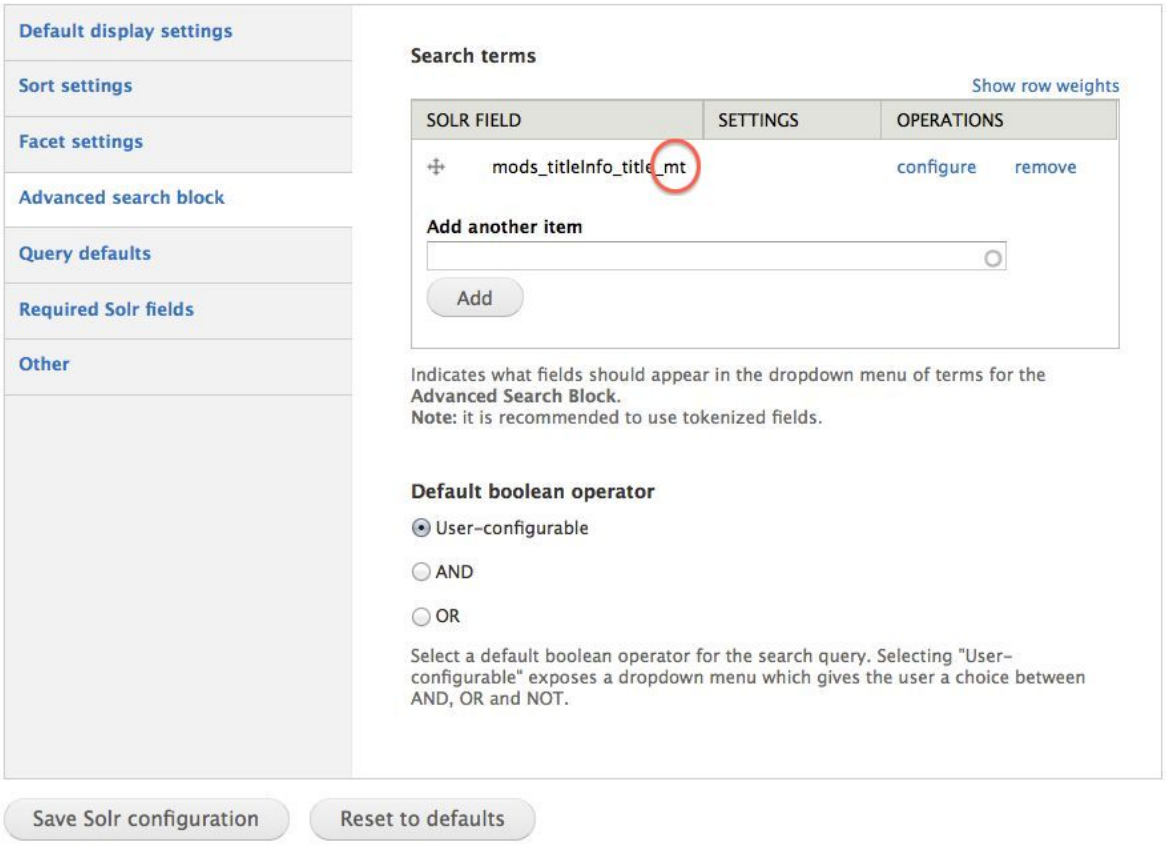
Here you can choose the search terms that will appear in the drop-down menus on the advanced search block. Terms must be entered by their field names (in most cases this will be **Dublin Core**, but you can also use **MODS**), though you can optionally specify a more human-readable label with the **conf igure** option. To determine the appropriate syntax for your search terms, simply edit the metadata on any object in your repository (for instructions, see How to Edit an Object's Metadata in Getting Started with Islandora). Each field label will display the proper syntax for adding it to your Solr search configuration. A full list of the terms made available by the schema provided in the module package is provided in APPENDIX D - SOLR SCHEMA (SEARCH) Term Reference.

Note that you will want to use fields that have been indexed as "text" in the advanced search block. Read APPENDIX D - SOLR SCHEMA (SEARCH) Term Reference for more information.

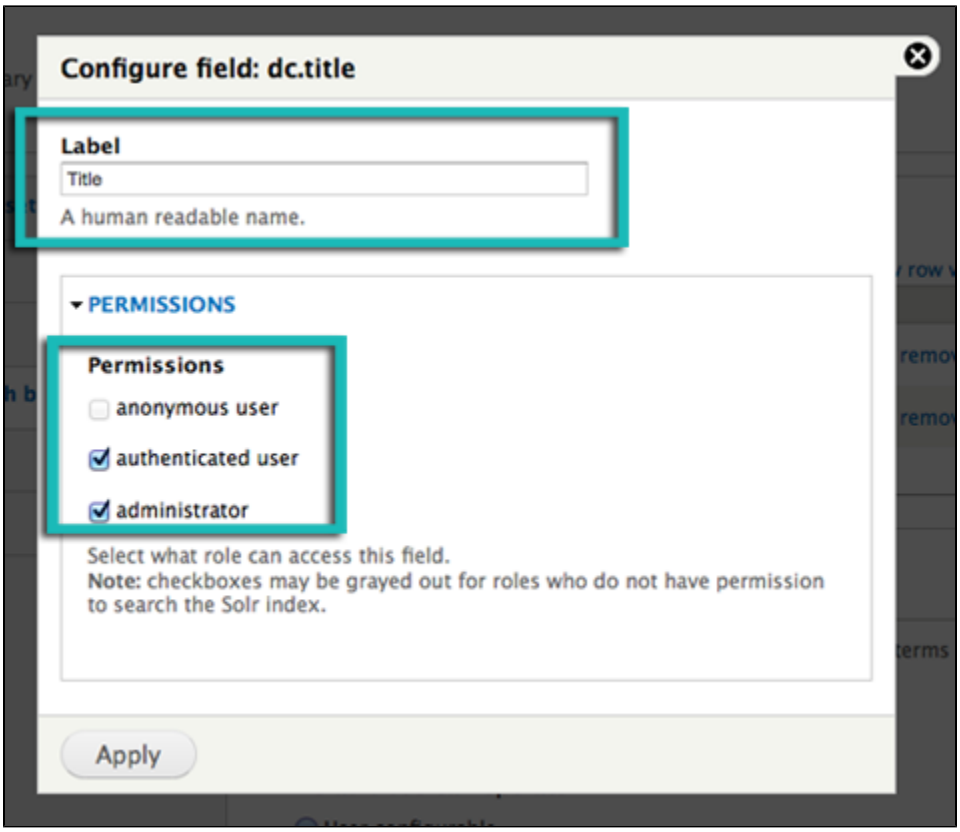

Default display settings
Sort settings
Facet settings
Advanced search block
Query defaults
Other
Dublic Core field names
Human readable label
Show row weights
SOLR FIELD
SETTINGS
OPERATIONS
dc.title
Label: Title
configure
remove
dc.subject
configure
remove
Add another item
Add
Configure labels and permissions
Indicate what fields should appear in the dropdown menu of terms for the Advanced Search Block.
Note: It is recommended to use tokenized fields.
Default boolean operator
User configurable
AND
OR
Select a default boolean operator for the search query. User configurable exposes a dropdown which gives the user the choice between AND, OR and NOT.

Types of Indexing

You may notice when setting up your Solr instance that some fields contain qualifiers like *_t*, *_s*, *_mt*, or *_ms* at the end of the field name. These indicate how the values of these fields are stored in Solr. Read APPENDIX D - SOLR SCHEMA (SEARCH) Term Reference for more information.



You also have the option of setting permissions on a per-field basis, allowing only certain subsets of users access to search across different facets. These permissions, and the human-readable label for each Solr field, can be configured by clicking **configure** to open a new options window:

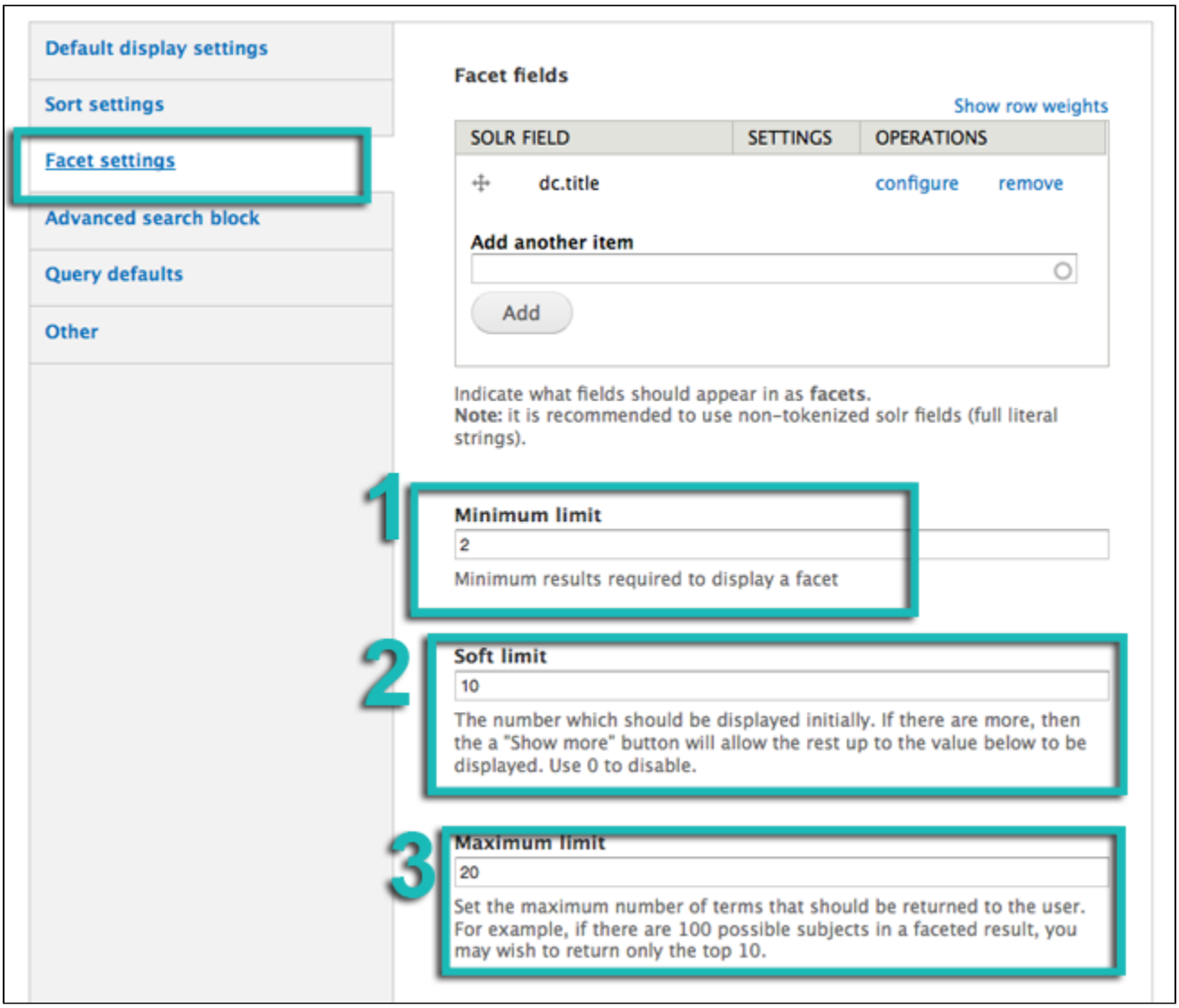


Solr field permissions are dependent on a role having Drupal permissions to search the Solr index. Roles without this permissions may appear in this list, but they will not be selectable.

5. Choose Facet Fields

Solr uses **faceting** to filter search results. Here, you can choose which fields you wish to allow faceting on. The format is the same as the search terms described above. You can also use this screen to configure:

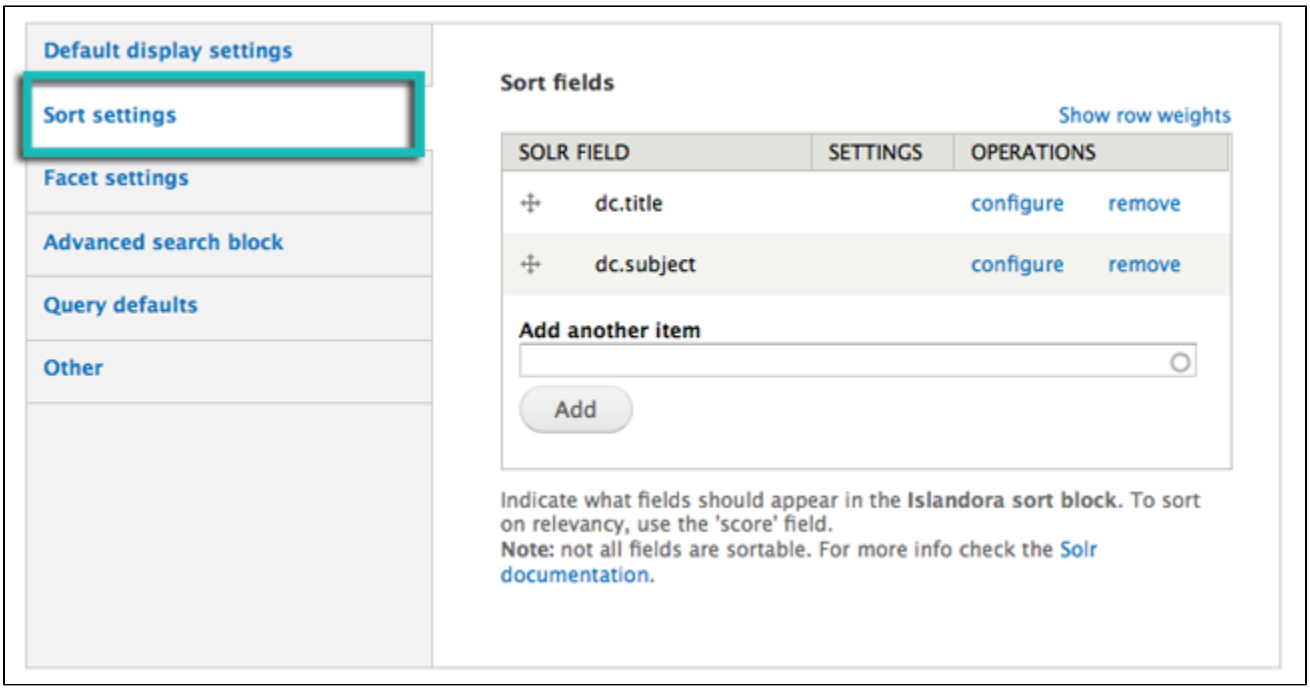
1. **Minimum Limit** - The minimum number of search results returned for a particular facet before that facet will be displayed. For instance, if the limit was set to '3' and a search for "fish" returned only two results for 'Bass', 'Bass' would not be included as a facet.
2. **Soft Limit** - The number of facets to show when a search is first returned. This setting will return the most populous facets first, and include a button to expose more available facets.
3. **Maximum Limit** - Similar to a Soft Limit, but without the option to expand to show more terms beyond the limit set here.



Note that you will want to use terms here that have been indexed as "strings". Read APPENDIX D - SOLR SCHEMA (SEARCH) Term Reference for more information.

6. Choose Sort Fields

Using the same field formatting as Advanced Search and Faceting fields, you can select fields to make available to the user to sort their search results in the Sort block.

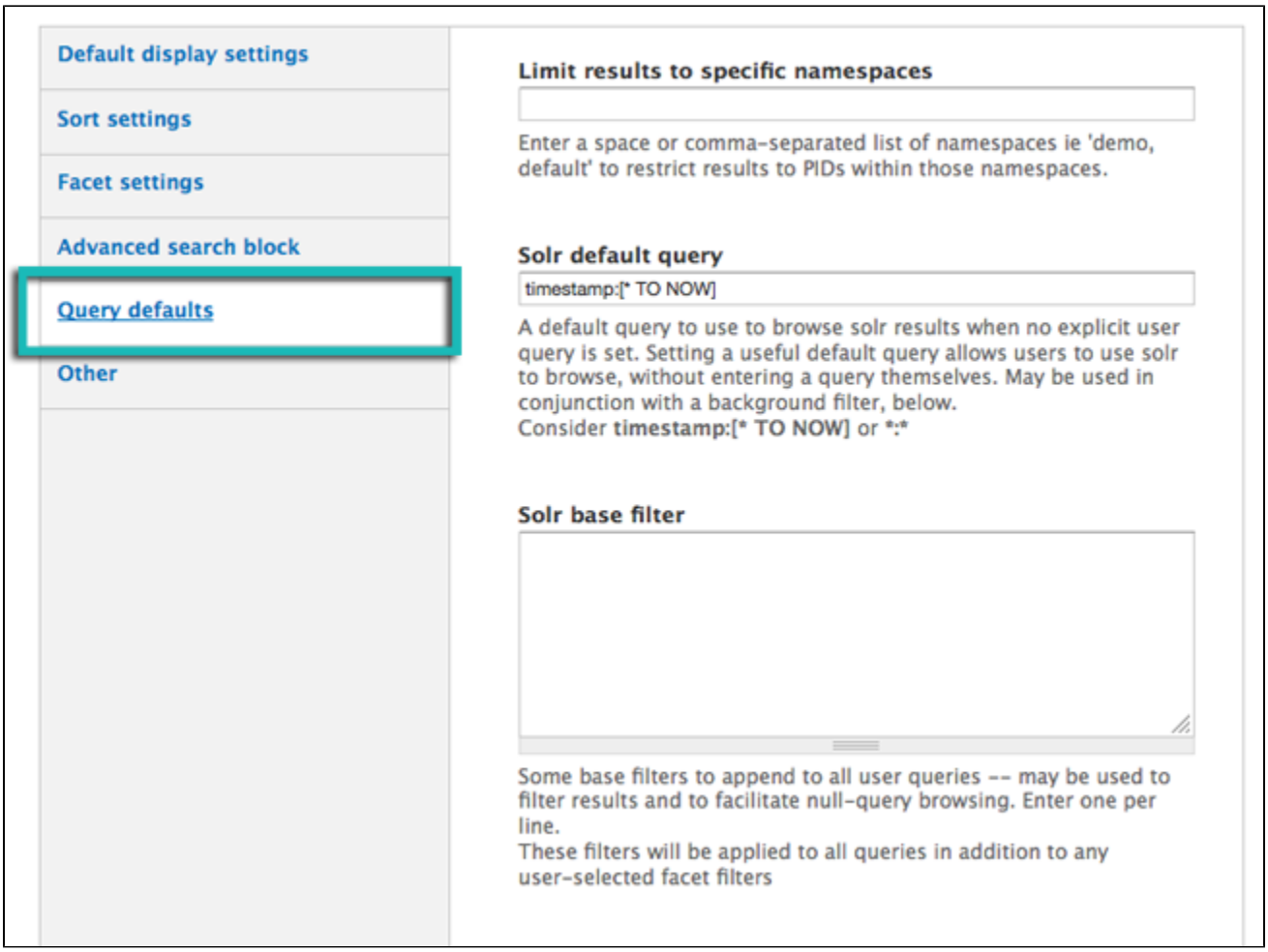


The field for a relevance based sort is "score".

7. Set Query Defaults

This section of the configuration panel provides advanced Solr query customization options:

- **Limit results to specific namespaces-** restrict your search results to a particular namespace. This is useful if there are multiple sites using the same repository and you want to block search results from the other sites. Remember that the namespace is the first half of the PID – everything before the colon.
- **Solr Default Query -** This option allows you to specify a default query used to browse results when no explicit query has been entered. For example, if a user runs a search and then deletes their search term from the breadcrumbs, this default query will be applied in its place.
- **Solr Base Filter -** You can use this option as a blanket way to filter all Solr search queries. For example, you can apply date-based or collection-based restrictions - such as removing page objects from search results by entering *-RELS_EXT_hasModel_uri_ms:"info:fedora/islandora:pageCModel"* in the Solr Base Filter field.



After setting up all of the above sections, you will have successfully configured your Solr client!

You may wish to now use the Islandora Solr Metadata module to configure your site's Metadata Display - find out more in Islandora Solr Metadata Display.

Read more about Solr

Consult the Solr Reference Guide to read more about Solr: https://cwiki.apache.org/confluence/display/solr/Apache+Solr+Reference+Guide

Accessing Solr outside of the Islandora User Interface

You may find it useful to know that you can access Solr, and GSearch in your browser, as well as through your server.

In a browser:

- GSearch - http://url.of.your.site:8080/fedoragsearch/rest
  - Useful for browsing or searching the GSearch index directly
  - Can also be used to update the index
- Solr - http://url.of.your.site:8080/solr/admin
  - From here you can view (but not edit) the schema.xml file and also view all currently indexed fields.

On your server:

- GSearch - /usr/local/fedora/tomcat/webapps/fedoragsearch/WEB-INF/classes/fgsconfigFinal
  - Most interesting files are:
    - fedoragsearch/properties
    - index/FgsIndex/index.properties
    - index/FgsIndex/foxmlToSolr.xslt and it's included files
    - updater/BasicUpdaters/updater.properties
    - repository/BasicRepos/repository.properties
- Solr - usr/local/fedora/solr
  - access schema.xml at usr/local/fedora/solr/conf/schema.xml