

Web Archive Solution Pack

Overview

The Web Archive Solution Pack adds all required Fedora objects to allow users to ingest and retrieve web archives through the Islandora interface.

Dependencies

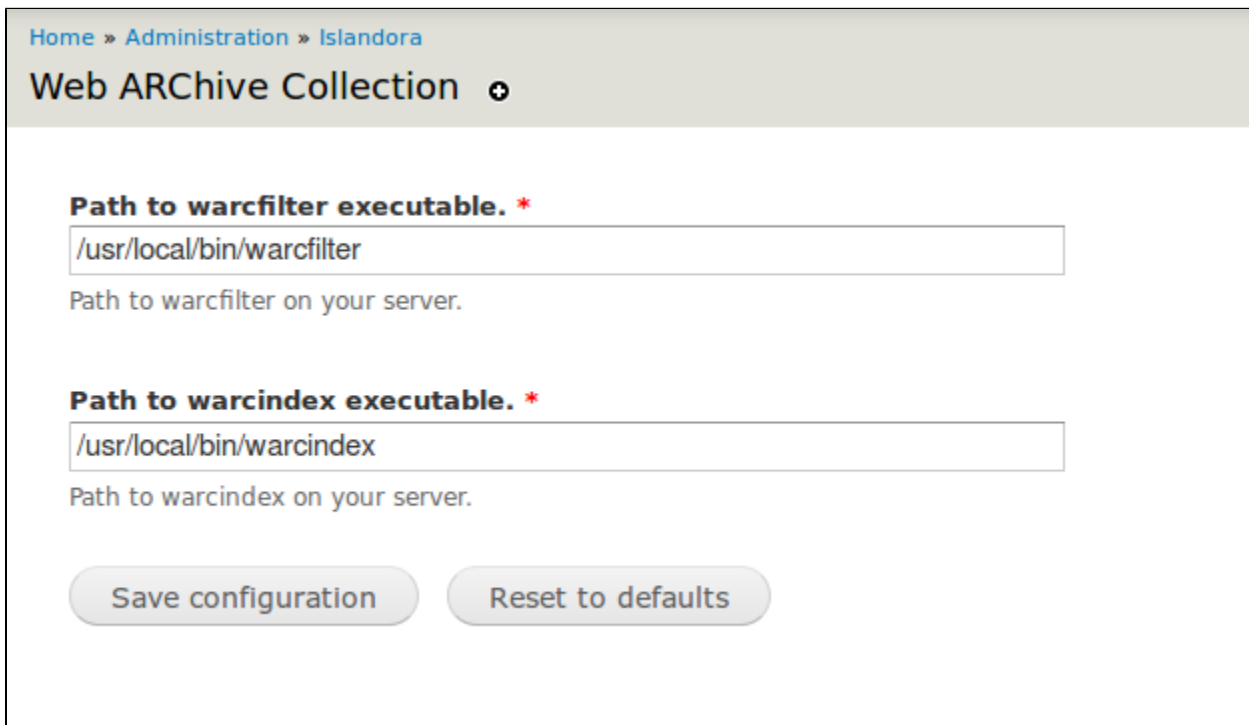
- Islandora
- Tuque
- warctools

Downloads

[Release Notes and Downloads](#)

Configuration

The Web Archive Solution Pack configuration options can be accessed at http://path.to.your.site/admin/islandora/solution_pack_config/web_archive. Set the paths for `warcindex` and `warcfilter` here:



Home » Administration » Islandora

Web ARCHive Collection

Path to warcfilter executable. *

Path to warcfilter on your server.

Path to warcindex executable. *

Path to warcindex on your server.

[Save configuration](#) [Reset to defaults](#)

If you are using Solr 4+, the `WARC_FILTERED` datastream will automatically be indexed via Apache Tika. You will need to add `ds.WARC_FILTERED^1` to the Query fields form in http://path.to.your.site/admin/islandora/search/islandora_solr/settings.

Content Models, Prescribed Datastreams and Forms

The Web Archive Solution Pack comes with the following objects in http://path.to.your.site/admin/islandora/solution_pack_config/solution_packs:

- Islandora Web Archive Content Model (`islandora:sp_web_archive`)

- Web Archive Collection (islandora:sp_web_archive_collection)

A file ingested using the Web Archive Solution Pack's content model will have the following datastreams:

| | |
|---------------|--|
| RELS-EXT | Default Fedora relationship metadata |
| MODS | MODS record filled out during ingest |
| DC | Dublin Core record |
| OBJ | Original WARC file uploaded |
| TN | Thumbnail derivative of SCREENSHOT |
| SCREENSHOT | Optional screenshot to represent the WARC |
| PDF | Optional pdf (screenshot) to store with the WARC |
| JPG | Medium sized JPEG of SCREENSHOT |
| WARC_CSV | Comma-separated index of the .warc file |
| WARC_FILTERED | Full-text filtered WARC for Solr index |

The Web Archive Solution Pack comes with the Web Archive MODS form.