PDF Solution Pack

Overview

The PDF Solution Pack module adds functionality to Islandora for ingesting and viewing PDF files. It uses the ImageMagick library and module to create derivative thumbnail and preview images. Because of the text-based nature of PDF files, it can also be used to create or append easily searchable text datastreams to the object, which can later be configured through Solr to appear in searches.

Dependencies

- Islandora
- Tuque
- ImageMagick is required to create derivatives. Debian/Ubuntu `sudo apt-get install imagemagick`
- pdftotext is required to automatically create a FULL_TEXT data stream. Debian /Ubuntu `sudo apt-get install pdftotext.`

Downloads

Release Notes and Downloads

Configuration

The configuration options for the PDF Solution Pack module can be found at http://**path.to.your.site**/admin/islandora/pdf, and include the following:

Text

Users can either upload a text file of their own, or allow Islandora to extract one from the PDF. Text accompanying the PDF is stored as the FULL_TEXT datastream. If both options are checked under the **Text** configuration section, and a valid path to pdftotext is entered, preference will be given to a supplied text file on ingest.

- **Allow users to upload text file with PDF?**: This file must be plain text stored in .txt format.
- **Store extracted (or supplied) text?**: Checking this box, will display an option to enter a path to the pdftotext executable. This package is not native to most server setups and will need to be installed manually for this option to be functional. Check the pdftotext dependency page for more information.

Thumbnail and Preview

These options set the parameters that will be sent to ImageMagick when ingesting a PDF. ImageMagick will attempt to create these using the first page of the document. Changing these will simply change the size of the derivatives being created.

Content Models, Prescribed Datastreams and Forms

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

- Islandora PDF Content Model (islandora:sp_pdf)
- PDF Collection (islandora:sp_pdf_collection)

A collection created using the PDF Solution Pack's content model will have the following datastreams:

RELS-EXT	Default Fedora relationship metadata
MODS	MODS metadata record created during ingest
DC	Dublin Core record
OBJ	Original PDF file uploaded
TN	Thumbnail image created by ImageMagick during ingest
PREVIEW	Preview image created by ImageMagick during ingest

FULL_TEXT	Optional datastream either uploaded during ingest, or created by the pdftotext executable

The PDF Solution Pack comes with the PDF MODS Form.