

Islandora OCR

Overview

The Islandora OCR module integrates Tesseract into the Islandora Paged Content module. It allows for creation of OCR and HOCR derivatives that can be appended to a page as a datastream. Check the instructions for the OCR-compatible module you wish to use for specifics on how to create OCR derivatives.

Dependencies

- [Islandora Paged Content](#)
- [Tesseract](#) (3.02.02 or later)

Downloads

[Release Notes and Downloads](#)

Configuration

Configuration options for the Islandora OCR module can be found at <http://path.to.your.site/admin/islandora/ocr>, and include the following options:

- **Tesseract:** Islandora OCR requires the path to your Tesseract binary to function correctly. It also requires Tesseract to be version 3.02.02 or higher to function correctly.
- **Languages available for OCR:** Islandora can look for any additional OCR languages you have installed; these are chosen from a drop-down menu at time of ingest or derivative creation.

It is recommended to check the Tesseract page for more information on these options.