islandora™
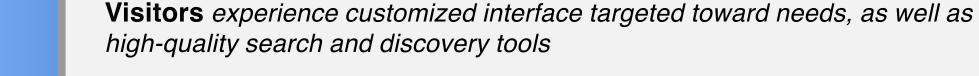
Introducing Islandora Stack and Architecture

Islandora is an open source framework developed by the University of Prince Edward Island's Robertson Library since 2006. It leverages both the expertise of PHP/Java/Python developers and of librarians and other information-professionals. Islandora is committed to utilizing open standards for data description and access, as well as high-standards for data stewardship and security over time. Islandora makes it possible to create, edit, discover, view, and manage repository assets. The system strives to strike a balance between extensibility and usability, by providing out-of-the box support for collections, while maintaining an architecture that lends itself to customization to other software and workflows. The heart of Islandora's data stewardship model is Fedora - Fedora users are capable of accessing and manipulating objects in Islandora's underlying Fedora as in any Fedora installation.

The Islandora project combines and harnesses the power of the Drupal content management system and the FedoraCommons Repository software to create a robust digital asset management system that can be used to meet the short and long-term collaborative requirements of digital data stewardship.

Islandora also represents a community-based approach to integrating robust open source software projects together. The resulting toolkit empowers users to create bespoke solutions and designs that accommodate the integration of pre-existing software solutions. Additional open source applications that provide flexibility and extensibility are added to this core stack to create "Solution Packs."

Visitors *experience customized interface targeted toward needs, as well as high-quality search and discovery tools*

Information Professionals *curate and manage collections that are built on open-standards and are flexible to local preservation and metadata policies, flexibly share collections with client groups, move items through editorial workflow, and place embargoes.*

Administrators and Managers *run reports about the research output of an institution, and view overall statistics for the repository (via Google Analytics integration), configure users, integrate with local identity stores (via Shibboleth or LDAP).*

Scholars and Researchers *access research and collaborate on projects, store data, contribute and edit material, publicize or protect their work.*

Ingest, DIscover, Manage Fedora objects via a Drupal Interface

When a file is ingested into the repository, other software applications are called to extract metadata, transform archival binaries into web-ready derivatives and otherwise prepare the object to be stored and accessible long term.

Document Conversion (Open Office, JODconverter, Ghostscript)

Data Transformation (XSLTs, bibutils)

Image Processing / Transformation (Imagemagick and Djatoka)

Optical Character Recognition (OCR) (Tesseract/Abbyy)

Generate Media Derivatives (LAME, FFMPEG)

Technical Metadata Extraction (file information tool set - fits & Mediainfo)

Data extracted and derived from the original file is then ingested alongside the original file, into the Fedora repository.

Viewers
Internet Archives Viewer
Large Image Viewer (iiv)
JWPlayer
Flexpaper
Flowplayer

Image Annotation (Shared Canvas)

Structural & Semantic Markup (CWRCWriter)

Drupal

"Drupal is a free and open-source content management system (CMS) and content management framework (CMF) written in PHP and distributed under the GNU General Public License. It is used as a back-end system for at least 1.5% of all websites worldwide, ranging from personal blogs to corporate, political, and government sites including whitehouse.gov and data.gov.uk. It is also used for knowledge management and business collaboration. [...] Although Drupal offers a sophisticated programming interface for developers, no programming skills are required for basic website installation and administration" (Drupal. Wikipedia, July 2012).

16,875 Modules
1,418 Themes
431 Distributions
18,923 Developers

Drupal.org, July 2012

In the Drupal 7 version of Islandora, all visual elements are fully integrated into Drupal's theming layer (templated) improving the ease of interface customization.

islandora™ Modules

In the Drupal 7 version of Islandora, a unit-tested API lives & breathes in a continuous integration environment.

Tools

Batch Ingest from .csv, .xml, and .zip
Collection Manager simplifies collection curation
Harvester creates metadata objects from OAI calls and .csv/.tsv files
XML Form Builder enables building forms in any metadata standard
OAI Provider shares your assets with other repositories
Workflow manages editing/publishing of objects
Solr Module configures search options for a repository
XACML Editor empowers security management at the object and datastream level

Solution Packs

Solution Packs are custom packages of content models and forms designed as starting points for users collections.

Basic Image Solution Pack
Large Image Solution Pack
PDF Solution Pack
Audio Solution Pack
Video Solution Pack
Books Solution Pack
Institutional Repository Solution Pack
Digital Humanities Solution Pack

In the Drupal 7 version of Islandora, the 'tuque' API abstracts Islandora's connection to Fedora, providing greater interoperability between Fedora versions, and laying the groundwork for integration into other repository systems.

A Servlet Filter syncs Drupal Roles with Fedora's XACML security layer

Microservices

Islandora Listeners detect messages from Fedora's Java Messaging Service, and trigger transcoding and processing actions based on content models. Utilizing microservices allows for resource-heavy processes to be abstracted to separate processing servers, improves general performance, and improves user experience of the ingest process.

Plugs in to domain-specific services & tools such as Sherpa/Romeo, Zotero...

Java Messaging Service (JMS)

Provides a method for applications to plugin to events triggered through Fedoras external APIs (ingests, edits, reads) by announcing these events.

Fedora Repository

Fedora Commons™

"Fedora (Flexible Extensible Digital Object Repository Architecture) was originally developed by researchers at Cornell University as an architecture for storing, managing, and accessing digital content in the form of digital objects inspired by the Kahn and Wilensky Framework. Fedora defines a set of abstractions for expressing digital objects, asserting relationships among digital objects, and linking "behaviors" (i.e., services) to digital objects.

The Fedora Repository Project (i.e., Fedora) implements the Fedora abstractions in a robust open source software system. Fedora provides a core repository service (exposed as web-based services with well-defined APIs). [...] Fedora helps ensure that digital content is durable by providing features that support digital preservation. " (About. fedora-commons.org/about, July 2012).

Leveraging the Fedora Content Model Architecture. Everything is a object, and everything has datastreams.

islandora™ CM Islandora's extension to the Fedora CMA (via the IslandoraCM stream) Datastreams call functions on ingest into the repository and determine how content will be viewed.

Content Model Objects

Collection Objects

Policy datastreams determine what types of content are permitted in a collection, and what security governs content in a collection

Data Asset Objets

Archival and web-ready derivatives are stored in datastreams alongside descriptive, administrative, and technical metadata streams

Authority records are a type of asset object, and Islandora simplifies the process of relating to authorities while editing asset objects.

Apache Solr

"Solr is the popular, blazing fast open source enterprise search platform from the Apache Lucene project. Its major features include powerful full-text search, hit highlighting, faceted search, dynamic clustering, database integration, rich document (e.g., Word, PDF) handling, and geospatial search. Solr is highly scalable, providing distributed search and index replication, and it powers the search and navigation features of many of the world's largest internet sites" (From:lucene.apache.org/solr/).

Fedora Generic Search Service (Gsearch)

When an item is ingested into the repository, a Gsearch xslt transforms the XML representation of the object into a format that can be read by Solr.

Mulgara Triplestore

open-source, scalable RDF database
Native RDF support
Multiple databases (models) per server
Simple SQL-like query language
Small footprint
Full text search functionality
Datatype support
Supports and tracks W3C Specifications and guidelines
Large storage capacity
Optimized for metadata storage and retrieval
Multi-processor support

Independently tuned for both 64-bit and 32-bit architectures
Low memory requirements
On-disk joins
Streamed query results
Full transaction support
Clustering and store level fail-over
Permanent integrity (docs.mulgara.org/overview/index)

RDF datastreams clarify the relationships between objects. These relationships are retrievable from the triplestore.

DURACLOUD

Coming soon! Integration with Duracloud will enhance the security of assets, and allow for managing data integrity via checksums, reports, and asset syncing.