

Upgration Pilot - NLW

Project Overview

We are looking to migrate one Newspaper title and one Modern Welsh Journal. The newspaper we are looking to migrate is called the Evening Express and is available from our Fedora 3 repository below:

<http://welshnewspapers.llgc.org.uk/en/title/view/3143635>

The modern Welsh Journals are still under copyright so have complex rights. We hope to investigate how rights can be managed and inherited from a Journal Title and issue down to the page. The Welsh Journals can be accessed at:

<http://welshjournals.llgc.org.uk/>

We will try and point a development Newspaper site to Fedora 4 to see what changes are required to make the website work.

-

Project Goals

Collection Description(s)

- Welsh Newspapers - <http://welshnewspapers.llgc.org.uk>
- Modern Welsh Journals - <http://welshjournals.llgc.org.uk/>

Object Models

- Newspapers
 - Newspaper Title - model:scif-newspapers-title
 - Newspaper Issue - model:scif-newspapers-issue
 - Newspaper Page - model:scif-newspapers-page

Fedora 3 Details

- Storage: Legacy storage (or Akubra)
- XML metadata : datastreams
- XML metadata : inline
- Content models
- Datastream types (inline, managed, redirect, and external)
- Identifiers
- Indexing strategies (GSearch, RI-Search vs. F4 approaches)
- Replication/Journaling
- Security policies: XACML
- OAI-PMH
- Versions
- Disseminators
- Audit history

Fedora 4 Details

Project Goals

With the Evening Express Newspaper title we are hoping to find out how Fedora 4 will deal with 187,331 objects (149, 869 pages and 37,461 issues) in a IIIF environment. The Website we have developed relies on SOLR and IIIF images so we will look at what changes are required to change the custom php indexing script to work with Fedora 4.

For the Modern Welsh Journal we will investigate the upgrading code to see if we need to customise it to ensure we can manage the rights in a sustainable way.

Collection Description(s)

Welsh Newspapers - <http://welshnewspapers.llgc.org.uk>

[Welsh Newspapers Online Beta](#) is a free online resource from the National Library of Wales where you can discover millions of articles from the Library's rich collection of historical newspapers.

Welsh Newspapers Online Beta currently lets you search and access over 725,000 pages from over 100 newspaper publications generally up to 1910 and will grow to over 1 million pages as more publications are added during 2014. This resource also includes newspaper content that has been digitised by [The Welsh Experience of World War One](#) project.

Welsh Newspapers Online Beta is part-funded by the Strategic Capital Investment Fund and the European Regional Development Fund through the Welsh Government.

Modern Welsh Journals - <http://welshjournals.llgc.org.uk/>

Welsh Journals Online provides students, teachers and researchers with free online, searchable, access to a selection of 19th-, 20th- and 21st-century Welsh and Wales-related journals held at The National Library of Wales and partner institutions. These materials cover a very wide range of subject areas, including humanities, social sciences, science and technology.

An innovative and dynamic web delivery interface allows the user to both browse and keyword-search a selection of the most significant journal and periodical titles – a resource estimated to contain 400,000 pages of text in both Welsh and English.

The website is maintained by the National Library of Wales and contains material licensed by the publishers of the journals. It excludes some material for copyright reasons. Use of the website is governed by terms and conditions.

Object Models

Newspapers

Newspaper Title - model:scif-newspapers-title

- METS
 - Type: Managed
 - Mime Type: text/xml
 - Versionable
 - Description: a large XML document which contains:
 - Descriptive information in MODS
 - Links to all issues
 - Links to all pages
- RELS-EXT
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: contains:
 - References to collections
 - Original Workflow ID
 - Catalogue Bibliographic ID
 - handle
 - workflow process ids
- DC
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: very basic Admin data
- POLICY
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: XACML permissions for this title
- RIGHTS
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: METS rights describing public access to content
- ISSUE_POLICY
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: XACML policy for Issues
- PAGE_POLICY
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: XACML policy for Pages

Newspaper Issue - model:scif-newspapers-issue

- METS
 - Type: Managed
 - Mime Type: text/xml
 - Versionable
 - Description: XML document which contains:
 - Descriptive information in MODS (only issue date at this level)
 - Article information in MODS for articles in this issue
 - Link between ALTO for each page and article in case of articles that cross pages.
 - Links to all pages in this issue
- RELS-EXT
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: contains:

- References to collections
 - link to Newspaper title (rel:isPartOf)
 - Original Workflow ID
 - handle
- DC
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: very basic Admin data
- POLICY
 - Type: Externally Referenced
 - Mime Type: text/xml
 - Versionable
 - Description: XACML permissions for this issue, this is linked to the Newspaper Title ISSUE_POLICY datastream. This allows one datastream to be changed in the Title and all issues will have the new statement.
- RIGHTS
 - Type: Externally Referenced
 - Mime Type: text/xml
 - Versionable
 - Description: METS rights describing public access to content, this links to the title RIGHTS datastream.
- PAGE_POLICY
 - Type: Externally Referenced
 - Mime Type: text/xml
 - Versionable
 - Description: XACML policy for Pages. This links to the Newspaper Title PAGE_POLICY datastream

Newspaper Page - model:scif-newspapers-page

- METS
 - Type: Managed
 - Mime Type: text/xml
 - Versionable
 - Description: XML document which contains:
 - Descriptive information in MODS (only page label at this level)
 - Technical information in MIX
 - Fixity and original file name information in PREMIS
 - Audit history of file pre-ingest in PREMIS
 - Links to Archive, Access and Coordinate datastreams
- RELS-EXT
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: contains:
 - References to collections
 - link to Newspaper Issue (rel:isPartOf)
 - Original filename pre-ingest
 - handle
- DC
 - Type: Inline XML
 - Mime Type: text/xml
 - Versionable
 - Description: very basic Admin data
- POLICY
 - Type: Externally Referenced
 - Mime Type: text/xml
 - Versionable
 - Description: XACML permissions for this page, this is linked to the Newspaper Issue PAGE_POLICY datastream.
- RIGHTS
 - Type: Externally Referenced
 - Mime Type: text/xml
 - Versionable
 - Description: METS rights describing public access to content, this links to the title RIGHTS datastream.
- ARCHIVE
 - Type: Externally Referenced
 - Mime Type: image/tiff
 - Versionable
 - Description: a http link to the archival tiff held in a near line tape library.

Fedora 3 Details

Reference: [Fedora 3 to 4 Upgration Checklist](#)

Storage: Legacy storage (or Akubra)

...

XML metadata : datastreams

See object models above.

XML metadata : inline

See object models above.

Content models

See object models above.

Datastream types (inline, managed, redirect, and external)

...

Identifiers

...

Indexing strategies (GSearch, RI-Search vs. F4 approaches)

...

Replication/Journaling

...

Security policies: XACML

...

OAI-PMH

...

Versions

...

Disseminators

...

Audit history

...

Fedora 4 Details

TODO: Outline Fedora 4 details!

Storage: Legacy storage (or Akubra)

XML metadata : datastreams

XML metadata : inline

Content models

Datastream types (inline, managed, redirect, and external)

Identifiers

Indexing strategies (GSearch, RI-Search vs. F4 approaches)

Replication/Journaling

Security policies: XACML

OAI-PMH

Versions

Disseminators

Audit history