**Document Distribution Note**
Release to Members: Immediate
Release to Public: 09/05/2014

# Islandora Camp, UK

London, 7-9 May 2014

## About the event

From 7 to 9 May 2014 Islandora Camp UK took place at Kings College London. You can find the programme at http://islandora.ca/camps/uk2014.

Angela Dappert represented the DPC on the first day only. The following DPC members were present in their own right: The University of Limerick as contributor to DRI, as well as Kings College. These notes are intended to provide an informal briefing for members of the DPC not able to attend the event. For an authoritative and comprehensive report readers are encouraged to contact the organisers of the event and the speakers directly. Kirsta Stapelfeldt, Alan Stanley, Nick Ruest and Donald Moses presented the event.

## Presentations

**Introduction**

The first session was an introduction to Islandora. Islandora is an open source asset management tool. A Drupal front-end integrates basic functionality, such as authentication, into Islandora. Solr is used for access and Drupal for workflow management. A Fedora back-end provides the storage repository. This XML data store will remain stable even if the top layers are changed. New versions - Islandora 7 and Fedora 3 - were just released last week.

Islandora employs a community driven release model. The University of Prince Edward Island (UPEI) has a grant that funded the initial development; they are in the process of developing this into a membership supported foundation. They pay for releases but need community driven development to make it sustainable. Discovery Gardens Inc. has developed a large portion of the code, offer cloud hosting, and provide installation, configuration, customization and other integration services for Islandora users. But they want to get away from the notion of being the code owners and have managed to develop a strong sense of community ownership, participation and code and thought exchange.There are also other providers, such as LYRASIS. The Chair of the foundation's Board, Mark Leggott may be contacted for information about service providers.

There are many ways of adding new modules for added functionality. Islandora supports best practice for digital preservation by packaging all functionality, such as creating derivative formats and metadata for a given content-type together. More about these solution packs later.

Code Management: There are about 40 modules that are currently supported. Modules that no longer have support in the user community will go away. They still work, but stop being supported. Discovery Gardens sets up tests and integrates with Travis. Jenkins and Travis CI are used for continuous development and JIRA for ticket management. DuraSpace, who hosts both DSPACE and Fedora Commons, provides the JIRA account. A Confluence wiki is used for documentation. Islandora uses LSAP (licensed software acceptance procedure) and CLA/CCLA (corporate license acceptance

procedure) models to ensure the quality of the code and the contributing provider's right to offer it. Islandora has the history of all code contributors and will retroactive licensing all parts of the code so that users may use the code with confidence.

Community: There are 150+ sites worldwide, with 500+ members on Google groups. They are mostly GLAM, government, and some commercial organisations.

Training: There are now 5+ training events per year and they are in the process of developing online training materials. They are asking for input of what people find difficult to understand about Islandora modules, so that they can prioritise their training material development.

Foundation: The foundation helps to maintain the code, provides training, supports community participation, etc.. 90% funding is based on the membership model with various contribution levels, the rest comes from camps and other sources. In the future they are looking for crowd funding and grants as income source. Because of the Drupal framework one can leverage Drupal community modules such as the commerce module for credit cards. Crowd sourcing project are also considered. For example, individual people sponsor the digitisation of local newspapers for digitalisation (see library.upei.ventures).
There is a board that is drawn from the top-membership level, with Mark Leggott as the Board chair; a weekly roadmap committee (a steering committee); a bi-weekly committers group (meetings for coordination of the developers and group that accepts tickets from the community), interest groups structured along the lines of various preservation topics. For example, there are interest groups for documentation (to ensure processes for uniform documentation) or for planning for the future as Drupal changes.

Consortial Implementations: There is a number of consortia with differing solution approaches. DGI is working on tools to help consortium managers manage their implementations.

Community contributions:  Examples of community contributions are WARC solution packs, checksums, Windows patches, a relationship editor, ontology management, a PREMIS module, a BagIt module, a checksum checker, a sync module for creating exhibits, a Solr metadata module, and a one-click Chef install.

At the camp all participants received a virtual Islandora machine that contains all the newest contributions.

**Attendants**

Attendants came from the Zagreb repository for HE institution, a Swiss library for research institutes, the Fundación Juan March, the Italian National Research Council, an Italian Museum of Tibetan Culture, a commercial Korean Drupal service provider, the University of Limerick, Kings College, the UK Freshwater Association, the German Federal Archives, and the Zuse Institute Berlin.

**Solution packs**

Solution packs provide the framework for a given content-type and contain all aspects of handling a digital asset. They are hidden from by the repository user. They define the behaviours for ingestion, derivation, metadata creation, persisting, etc..  Solution packs have dependencies but stand alone.

For example, a solution pack may manage audio material in a number of formats. Everything is plug-and-play so the components can be swapped out without affecting already ingested material.

Ingestion: The most common form of ingestion is upload through the UI or batch upload, but it is also possible to create a digital object on the fly (e.g. done in digital humanities).

Derivative creation: TIFF, for example, is not friendly for display. The image pack, therefore, also automatically creates J2k derivatives at several resolutions at ingest.

Organisation into collections: This solution pack supports data modelling in a facetted rather than a tree formation. Collections are just labels that are applied to the objects; objects can exist without collections.

Persisting the metadata: Preferred schemas are MODS, METS, PREMIS, etc.

Persisting the content and metadata: in Fedora

Display: Make the content available in a number of pluggable ways. Every content type offers a selection of viewers, such as OpenSeadragon, for originals and derivatives. They can be replaced at any time. Viewers can be selected in each solution pack if there are several viewer choices. If you have a proprietary viewer that is not of general interest you can also drop it into the pack for your own use.

Solution packs exist for audio, basic image, book (which has a dependency on the large image pack because they tend to ingest TIFFs), collection, compound, large image, newspaper, pdf, video, web archive. They all have default derivative formats.

Islandora generalise some of their solutions so that they can be pushed back to the community for general use. Rather than having the local custom solution, it is kept general and can be customised by ticking boxes.

- Newspaper

Islandora created workflows and metrics for out-of-the-box newspaper management and access that users can customise. It uses SolrView for access with OpenSeadragon as a viewer. Using facets, thumbnails, different views, metadata (enriched about editors of the newspaper at the time, etc.). To create a browsable view of a newspaper with a very large number of runs they came up with a calendar view from which the users can pick the desired issues. This view helps anomalies to show up. If, for example the paper is not published on Sunday then an entry on the calendar views for Sundays shows cataloguing errors. It also shows missing issues, and encourages the public to provide them, if they should have copies. You can create a deep link directly to the year, so that you do not have to traverse the hierarchy. For batch ingest, the pack uses PHP listeners that can be extended for different workflow systems.

- Book

Books are in general zip containers that contain sequences of images. They also contain thumbnails and OCR. Tesseract OCR open source can be used on TIFF, jpg and pngs for OCR creation. You can search inside a book with Solr. It uses the Internet Archive book viewer.

- Collections

Collections are a good way to show metadata. You can place restrictions on content types, individuals or objects with great granularity. You can migrate or delete objects. If you delete, it checks which collections the deleted item is a member of and only severs the membership to the one in which you delete. Language around deletion is confusing: deletion, which only renders the item invisible, is different from purging, which is irreversible. Metadata for deleted or purged objects can show up in searches if that is wanted.

- Compound object

This is a pack for objects consisting of sub-objects that you don't necessarily want to see individually. Parents are not aware of the children, but children are aware of the parents. Children have sequence numbers so that they can be ordered.

- Large Images

The image pack, by default, has TIFF masters and creates j2k derivatives, web derivatives and thumbnails and associates them with the SeaDragon viewer.

- PDF

The pdf pack, by default, has pdf masters and creates text and thumbnail derivatives.

- Video

The video pack uses ffmpeg for default. Islandora users use Video.js player, or JW Player or Media Elements, which can be used for audio and video. Different browsers have different default streams. Islandora have put in a sniffer which identifies the right stream automatically. In order to maintain acceptable service times for users they do the processing on the back server, rather than on web server which just pushes the streams out.

- WARC

This pack uses filtered WARC as derivative. The Wayback machine viewer will be incorporated as through descriptive metadata at DGI.

**Tools**

Islandora offers a variety of configurable tools to its users.

- BagIt

Container to share content in a folder (bag) for all the content in a data directory combined with its metadata. It gets developed from a generic Drupal bag in Islandora and is used for download, for containerisation of a whole directory, or for migration. It can also be used for ingest including checksums.

- Checksum

They are created when you import an object;

- ChecksumChecker

This works in conjunction with the Checksummer. New content goes to the bottom of the list; the user can configure how many objects should be checked every cron period so that they can adjust the load on the system.

- FITS

A file information tool set for characterisation of file formats. It is a wrapper for many other characterisation tools. Solr can index the FITS output and extract information from it, such as the resolution, the scanner used for digitisation, the date of scanning, etc.

- PREMIS

PREMIS metadata is created on demand from event logs and technical metadata which is stored in various Islandora data locations. Events, such as derivation events or checksumming events are logged. Technical metadata is extracted through the FITS module and from the Fedora audit trail, etc. All of them contribute to the PREMIS description of the objects. For the object location the 'Display' command does not display the actual location since it exposes information about the setup of the server. Entitled users can see the location.

- Annotation

Annotations are used to annotate a region in an image by drawing a rectangle, oval or polygon around the area and adding descriptive information. Each annotation is a Fedora object. It can be opened up as a social object, by community users or owners or experts. Community user annotations can be authorised by the owner. This also can be used as a teaching device. Student annotations can be authorised by the professor. Annotations are implemented by mark-up without touching the original and are done in a different tool altogether.

- Internet Archiving Book Reader

An open Source book reader which exposes serialised jpegs in a page-oriented (rather than an image oriented) way. It is the only viewer choice for books at the moment. Another choice might be OpenSeaDragon, which looks more like a collection of photographs than like a book-flipper. For example, for Tibetan manuscripts you need to flip up pages, so you should use OpenSeaDragon. It gives you OCR text, but has no text image linking that would identify line-by-line mark-up. It is designed for well digitised printed books. There is also H-OCR to pull back lines and apply CSS. But

since most OCR does not work terrifically well having to clean up H-OCR is daunting. In version 6 they had a viewer to deal with different page sized, fold-outs etc. which requires software additions.

- Importer

This is a helper module to help with batching. You have to prepare your data first outside Islandora. You give it a zip file that contains the content model, collection, the file (e.g. named foo.tiff), descriptive MODS metadata (e.g. named Foo.mods), technical metadata (foo.techmd). For metadata schemas other than MODS you would have to extend the importer.

- Book Batch

There are two types of book batching. In the first a book is just considered a collection of ordered pages. Page ingest is simple. The metadata is usually at book level, but is allowed on page level. The second does not require data preparation. It zips the entire directory that is structured in the metadata rather than through the GUI.

- MARC XML

This provides a MARC XML cross-walk to MODS so that you can reuse already catalogued metadata.

- Form Builder

There are built-in forms that you can clone and from which you can create derivative forms. Features of the tool include autocompletes and extension with Google taxonomies. There are existing forms from DC, METS, MODS, PBCORE that you can reuse. The forms match the xml in the background. You can pull any of them into Oxygen. You have to know xpath to use the tool.

- XACML

XACML is used to enable restrictions on viewing (copyrights, unfinished versions, reference copies) and for editing. You have roles and users and can assign the one to the other. One goal is that they want to use search on full-text but not allow users to view the whole text, only a text snippet, in order to protect viewing rights.

- Sitemaps

The Drupal sitemaps module can be extended to be used in Islandora. It grabs the information from Solr to create the sitemap from it. You can then submit it to the search engines.

- Simple Workflow

The simple workflow is bypassed by admin status, but others can use it. It sets the object to an inactive state. It can then be activated. In an annotation content model you might add a workflow that lets you look at comments before they are made public. Most Fedoras are shared, but you would like a separate instance for anonymous uploads. Inactive objects are not searchable and viewable through Solr. There are 3 states: active, inactive, deleted (and purged). Some projects have many more states.

It shows versions for all different aspects: DC update, derivatives. You can make any data streams versionable or not versionable. Duplicate the workflow so that you choose the right one only when you want it. You could return the object to any state it had at a given time.

- Display Profiles through XQuery

The module has just been released. It is a search and replace tool that is potentially dangerous if not used with care. But they put in some DRUPAL hooks to address this. The example showed was just for demo purposes and was a function that turns the default display results to upper case. There is one default display per site at the moment. With XQuery you get the different search results. Uperize turns all search results found into upper case. A preview function lets you see what you did and lets you cancel out of it.
Tokens make the XQuery reusable and parameterizable. The tool can be used for metadata clean-up based on patterns. You could do this with any Solr query. This is exclusively for XML streams in Fedora. Data streams are not deleted and are versioned so you could roll-back if you messed this up. There also is object locking where you identify content-types and data streams and when you open them they are locked to avoid parallel changes through different editors. The lock expires after 30 minutes. If you don't safe the data stream in this time you lose your changes. Before you run an update process the whole batch is locked. If someone else modifies an overlapping subset the locked data streams are excluded from the update and the system gets out of sync.

This concludes the brief overview over the first camp day. We finished the day by installing a virtual machine on our laptops. The second day was a full day of workshops/to get your hands on Islandora and gain some practical experience, either via the user interface (admins) or by digging into the underlying code (devs). Friday featured a close look at many of the tools available for Islandora and sessions from the Islandora community.

## About this document

| | | | |
|---|---|---|---|
| Version 1 | Written on the day | 7 May 2014 | AD |
| Version 2 | Distributed | 9 May 2014 | DPC members |