Practical preservation with Arkivum & Eprints

An EPIG talk proposal by Rory McNicholl & Jose Martin Cuesta, Academic & Research Technologies, ULCC

Background

Academic & Research Technologies team at ULCC hosts and supports around 20 EPrints repositories for a variety of institutions and a broad range of content. Repositories range from traditional publication repositories to image archives and research data repositories.

The ART team also has a depth of experience consulting on preservation issues, running the Digital Preservation Training Programme and managing digital preservation projects for Jisc and the Linnean society.

Arkivum Ltd provides a large scale digital storage service that is flexible and cost effective. ULCC has recently partnered with Arkivum to installed a gateway appliance (“A-Stor”) at the ULCC data centre, and begun installing the Arkivum plug-in to hosted repositories.

So how have we used our experience and this new infrastructure to find practical solutions for our clients? We propose to present two case studies detailing our real experiences with using these two systems in tandem.

A Repository for preservation management - Jisc Collections

Jisc Collections have a lot of digital content and are mandated to preserve it. Existing content had undergone “bit-level preservation” at several locations and institutions, however it had become necessary to consolidate content in a single location, conducting a thorough audit, to ensure their continuing preservation of existing and future content.

ULCC is working with Jisc Collections to design appropriate preservation policy and a set of practices to manage the ingest and storage of the digital content in the long term. Data ingest would be ongoing and it was not immediately clear how much existing data there was. The necessary technical infrastructure to achieve good quality preservation would need to include a flexible, cost effective large scale digital storage solution.

Arkivum provide such a storage solution with added benefits brought by implementation of preservation best practice. Taking in to account issues beyond bit-preservation such as the legal jurisdiction under which the data is held and the possibility that the data held may need to out-survive the organisation holding it.
It became clear that organisation and metadata management would be needed. To achieve this we decided to use an EPrints repository. All content would pass through the repository on its way to the archive storage. This added a layer of metadata management to the content, but also to the preservation processing. Events relating to the digital objects in both EPrints and Arkivum are recorded and can be collected and stored with the data as part of the Archival Information package.

Technologically, the pairing of EPrints and Arkivum proved simple with the use of the Arkivum bazaar package. The plugin was used in its default mode, which begins the arkivum archiving procedure once an item is in the “live” repository.

The repository provides the ability to create different views of the content based on metadata, as well as metadata searching of the archived material. Both being useful administrative functions.

EPrints also allows for more nuanced access management simply by using the default user access options. There was no requirement to provide general access to the content so in this case the whole repository is run as a “dark archive”. The next case study explores further the notion of actively managing the preservation workflow using EPrints and Arkivum.

Preservation and dissemination for an Image archive - The Linnean Collection

Since 2007 ULCC has used an EPrints repository to provide access to the Linnean Collection. This is the user-end of a number of digitisation projects undertaken by the Linnean Society with a view to increase access to their historical collections of specimens, documents and artwork.

In parallel with this work ULCC has also undertaken to preserve the original tiff images that form the raw output of the digitisation projects. In 2014 we have been able to combine and streamline these two activities. The preserved images have been moved onto the arkivum service and the repository has been linked to this using the Arkivum plugin.

The preservation copies are not made available to the public, but there is now a real link between the repository record and the preserved digital object. They are managed by the same repository system and described by the same metadata fields. They are however not stored in the same place. The dissemination copies (lower resolution pyramidal tifs) are stored locally and can be accessed quickly by all via a tiling image server. The preservation copies are archived and are accessible only by Linnean Society staff. To achieve this we have introduced some simple logic into the EPrints storage management layer. Full size originals go one way, lower resolution copies go another.

For new collections we have also embedded the creation of the dissemination copy into the EPrints import procedure. So when a full size tiff arrives a dissemination copy is created, stored locally and then the original tiff is sent off to the arkivum service. This simple preservation policy
(ie securely store original tiffs, create dissemination copies for access) has been designed into the repository so that it all happens automatically.

The outcome of these changes, apart from a more competitive storage price, is an automatic and tangible connection between the item metadata and the dissemination and archival data. This connection is something that, in the past, would have required some “joining of the dots” by ULCC staff.