



11th International Conference on Open Repositories
Trinity College, Dublin, Ireland
13th - 16th June, 2016

Illuminating Otago Heritage

Allison Brown
University of Otago Library
allison.brown@otago.ac.nz

Emmanuel Delaborde
University of Otago Library
emmanuel.delaborde@otago.ac.nz

The University of Otago Library is responsible for the curation of extensive New Zealand and other heritage collections, in particular, those housed within the Hocken Collections.¹ The need to better curate and showcase these collections to the widest possible audience has informed the development of a digital preservation solution for all Library digital assets, including, but extending beyond these valued heritage items.

This presentation will outline the Library strategy for building staff capacity and knowledge around (a) the preservation and curation of our digital assets, (b) the establishment of Curation Framework policies, and (c) the development of a Digital Asset Management System (DAMS). With attention shifting from access (the usual focus of repository requirements) to curation, the University of Otago Library has invested in developing a DAMS to curate digital objects using Fedora Commons / Islandora software, rather than other software currently in use (for example DSpace and Omeka).

Themes: Repositories and Cultural Heritage, Integrating with the Wider Web and External Systems, Managing Rights, Developing and Training Staff

Audience: Repository Managers, Developers, Librarians, Archivists and anyone interested in digital curation.

Keywords: Cultural heritage, DAMS, Digital Curation, Fedora, Islandora

1. Background: Context and history

University of Otago was founded in 1869 and is New Zealand's oldest university. The University of Otago Library includes Hocken Collections and Special Collections. Hocken Collections was established in 1910 when Dr T. M. Hocken gifted his collections to the people of New Zealand, forming its foundation. This collection consists of "...archives, books, drawings, journals, maps, music and film, newspapers, paintings, photographs and posters, dating from the seventeenth century to the present day".^[i] Special Collections² "...contains books and manuscripts which span the earliest examples of European printing (1473) through to modern first editions and publications from the 20th and 21st centuries".^[ii]

There is a strong demand for the Library to digitise items for research purposes. The Library is also increasing its collection of born digital content. As a result, the Library requires a system to provide sustainable management of digital objects; to retrieve, access, and track changes. The Library also needs to centralise the various digital collections in order to provide the core functionality that researchers need, in particular those working within the Digital Humanities.

¹ Hocken Collections www.otago.ac.nz/library/hocken/

² Special Collections <http://www.otago.ac.nz/library/specialcollections/>

2. Preservation and curation of digital assets

“Digital Curation involves maintaining, preserving and adding value to digital assets over its lifecycle”.^[iii] Digital Preservation is the method of keeping digital objects alive so that they remain usable as technological advances render original hardware and software specifications obsolete.^[iii] The difficulty with digital data is that the technology required to read the storage medium may be long gone. It requires continuous active intervention to be preserved.

The Library has developed a Digital Asset Management system (DAMS), along with appropriate policies, guidelines and curation framework to manage current and future digital assets. The Library is working towards establishing a system and workflows to support the whole lifecycle of our digital assets. A good overview of the phases we are striving to cover can be seen in the DCC Curation Lifecycle Model.^[iv] Based on the repository software solutions used previously by the Library, these perform well in areas such as ‘Create’, ‘Ingest’, and ‘Access, Use and Reuse’.^[iv] They also provide some functions for the other phases but not to the standard the Library requires. For example, we will need to provide derivatives and even transform these derivatives into different formats (when required) and still retain a master copy that is preserved, while recording the relationships between each file. At the same time, we want to avoid the need for staff to generate derivatives manually, which is what has been done in the past.

3. Establishing Curation Framework policies

The Library established core protocols and standard formats as a Digital Asset Management / Curation Framework. This is living document that determines how the DAMS will be administered and how digital assets are managed and curated. The protocols determine the administration of the DAMS parameters, procedures and processes. For instance, protocol one requires the Library to include digital assets only if they are covered in the Library Collection Development Policies.

The preferred file format standards must be ratified by an appropriate industry or professional body and be actively maintained or updated. The scope of the first iteration of the Curation Framework and DAMS focuses on the most common formats managed by the Library, where Islandora already provides a Solution Pack, for example large images, audio, video and books. The Library processes then ensure that staff use the preferred formats. The Book module can ingest PDF but the Library preference is to ingest each page as TIFF. This is important for handwritten manuscripts. If any other file formats are required for an item, this is automated within the Solution Pack.

4. Selecting Islandora and Fedora Commons

Library staff have conducted a number of evaluations of different repository software and requirements for the DAMS. In considering the functionality requirements, a comparison was made between the software that had been selected and the workflows applied, and what is covered by the reference model for an Open Archival Information System (OAIS).^[v] This helped us to identify gaps and plan future developments after the DAMS went live. OAIS and the DCC lifecycle helped clarify the gaps in other repository software functionality which the Library is already using (DSpace³ and Omeka⁴) and in part informed the decision to select a different software (Fedora and Islandora) for the DAMS.

A significant advantage of Fedora is its focus on storage and management of digital objects. We can choose to integrate different front-ends via web services. Another advantage is its relationship model that allows the representation of complex relations between objects and their components.

³ OUR Archive: DSpace Institutional Repository <https://ourarchive.otago.ac.nz/>

⁴ OUR Heritage: Omeka Library Digital Collections <http://otago.ourheritage.ac.nz/>

5. Development of a Digital Asset Management System

The Marsden Online Archive⁵, developed with Fedora Commons v3.7.1, Islandora v7x-1.2, Durpal, Apache Solr, was used to pilot the DAMS. This pilot demonstrates the type of collections managed by the Library and what can be built upon the DAMS moving forward. The Mining Marsden project identified 60 requirements, of which 45 could be delivered when the Archive went live. Marsden Online Archive includes:

- Letters and journals which document the first-contact experiences between Māori, Samuel Marsden, and other early missionaries.
- Manuscripts describing the introduction of new plants, animals, technologies, commodities, and ideas to the people of New Zealand, and early attempts to document the Māori language.
- Searchable manuscript transcripts, enabling users to drill down to page level results and display the page image beside the transcript.
- Alternative spelling list added to Solr search index, which directs the user to the page of the manuscript the word or phrase appears in.
- An external application developed to generate the metadata in XML, including the TEI (text encoding initiative) mark-up prior to ingest.⁶

The DAMS, developed with Fedora Commons v3.8.1, Islandora v7x-1.5, Durpal, Apache Solr, highlights include:

- Derivatives and formats generated based on the Library's requirements from the Master file.
- Standardised identifiers providing unique succinct links for every file.
- Staff being assigned roles and provided authenticated access via Otago's Shibboleth system.
- An ingest form developed in Drupal providing a single ingest form for staff, which
 - Pulls metadata harvested via other management systems' OAI-PMH, like the unpublished management system Hākena.⁷
 - Enables files to be uploaded to the Drupal Database before being ingested into the DAMS.
 - Establishes a workflow for the Library's Reprographics to add digitised assets to the DAMS, which other staff have added to the Ingest form.
 - Includes the ability to add item collection metadata or add items to an existing collection.
 - Provides a crosswalk of the OAI-PMH Dublin Core to MODS.

6. Advantages and challenges of the Marsden and DAMS projects

Advantages:

- We can draw from what we have learnt with Marsden and the DAMS.
- We have managed to adapt and extend the software to accommodate specific needs.
- By choosing a solution that is open source and also not monolithic (DSpace) but is composed of tweakable components we gained a lot of flexibility.
- Solr provides indexing of the Fedora objects, which was crucial for the Marsden Online Archive.
- Standardised identifiers providing unique succinct links for every DAMS digital object.
- The Library is closer to being able to say yes to the National Library of New Zealand's question: is our repository trustworthy?

Challenges:

- Complicated set-up, with no truly functional one-step for whole package installation that worked.
- The documentation was poor in relation to the complicated set-up, especially during the Mining Marsden Project. However, this has been improving since we started work on the DAMS project.
- Integration with external management systems in particular syncing metadata, relies heavily on those systems being able to integrate.
- Managing roles and permissions and connecting these to the Otago Shibboleth system.

⁵ Marsden Online Archive: Islandora Digital Humanities Archive <https://marsdenarchive.otago.ac.nz/>

⁶ XML Creator application <https://bitbucket.org/libADS/xml-creator>

⁷ Hākena: Search Hocken Collections archives, pictures and photographs <http://hakena.otago.ac.nz/index.html>

- Long-term use of the DAMS, format conversions, and working on expanding the formats the DAMS manages (for example word processing files).
- Selecting the appropriate metadata schemas and elements required to curate the digital assets and avoiding staff adding information that already exists in other management systems.
- Being more inclusive with the Collection Management staff responsible for curating the collections, while focusing to complete developing the first iteration of the DAMS.
- Establishing a functional Curation Framework that helps the DAMS and processes meet international standards.
- A DAMS should have redundant disk storage in different geographic locations, but currently this is not an option for Otago ITS servers.

7. Future phases for the DAMS

- **Move assets into the DAMS:** Complete migrating existing digital assets covered by Curation Framework from separate servers and potentially other repositories.
- **Integrate with management systems:** Establish the DAMS as a backend for assets managed by other systems.
- **Embed the DAMS as part of staff processes:** Continue training Collection Management staff in the digital curation skills required to take ownership of the DAMS.
- **Fedora 4:** Schedule upgrading from Fedora v3.x to v4.x and work through the complexities this creates because of the huge differences between these versions.
- **Extending the DAMS:** Evolve the DAMS to add more functionality, for example monitoring files so the Library converts them when they are at risk of becoming obsolete. Continue to enhance the DAMS to make sure it meets international standards and enables curating digital assets.
- **Evaluate the repositories:** Compare the different repository software and services the Library manages and decide what should be moved into the DAMS infrastructure.

8. Conclusion

The Library's commitment to curating digital assets has resulted in the development of a DAMS and Curation Framework. Throughout this process the development team has had to address issues and complexities, but this has resulted in building a stronger team who can continue to extend this infrastructure. Integrating the DAMS with other systems enables digital curation processes to be standardised for Collection Management staff. The Library has now built a solid foundation on which to build, with the potential to establish other Online Archives (such as Marsden), drawing from the extensive collections held within and beyond the Hocken and Special Collections.

9. Acknowledgements

The authors thank everyone who participated in the Mining Marsden and DAMS projects.

10. References

- [i] Hocken Collections (n.d). About Hocken Collections. Retrieved 28 Jan, 2016
<http://www.otago.ac.nz/library/hocken/otago038926.html>
- [ii] University of Otago Library Special Collections (n.d.). Special Collections. Retrieved 28 Jan 2016, from
<http://www.otago.ac.nz/library/specialcollections/>
- [iii] DCC (n.d.) [b]. What is digital curation? Retrieved October 30, 2015, from
<http://web.archive.org/web/20151030001736/http://www.dcc.ac.uk/digital-curation/what-digital-curation>
- [iv] DCC (n.d.) [a]. DCC Curation Lifecycle Model. Retrieved October 28, 2015, from
<http://web.archive.org/web/20151028221802/http://www.dcc.ac.uk/resources/curation-lifecycle-model>
- [v] CCSDS (2012). Reference Model for an Open Archival Information System (OAIS): Recommended Practice CCSDS 650.0-M-2 Magenta Book June 2012. Retrieved 14 February, 2013, from
<http://public.ccsds.org/publications/archive/650x0m2.pdf>