Welsh Journals Online

The National Library of Wales | Llyfrgell Genedlaethol Cymru

This is one of four case studies created to illustrate digital preservation practices within digitization projects, sharing honestly the challenges of establishing digital continuity from a project of limited lifespan. It is selected from 16 projects which were funded by the JISC Digitisation Programme between 2004 and 2009. This Programme has provided digital access to collections of international significance that would otherwise be inaccessible. In doing so it has also generated an opportunity to learn about the critical success factors in digitisation, such as ensuring long term access from short term projects. Each case study is based on an interview carried out on behalf of the JISC by the Digitisation Preservation Project.

Introduction

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.

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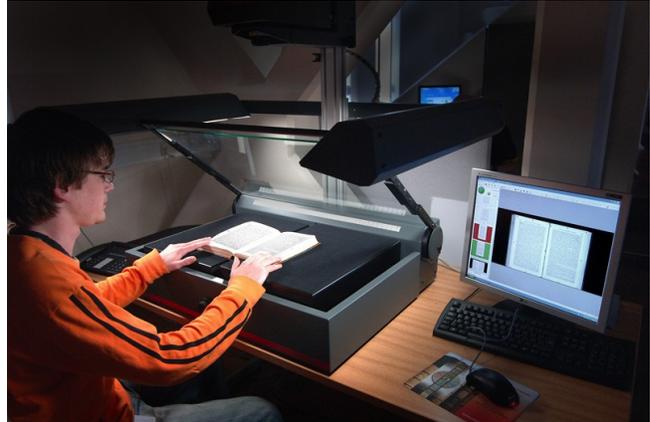
The website is maintained by the National Library of Wales and contains material licensed by the publishers of the journals. It excludes some material for copyright reasons. Use of the website is governed by terms and conditions.

General

The project is a successful one as it demonstrates a cohesive and well managed approach towards managing the project in relation to continuing access. There are many approaches to be recommended.

Staff and skills: The project had 3 core staff as well as 11-12 funded posts with other staff allocated for ongoing maintenance during the post funding period. Project staff identified it as crucial to have good project

Digitisation technician at work, NLW

management skills as well as knowledge of data standards, system development requirements, cataloguing rules. This was useful in understanding the issues and knowing what corners can be cut when needing to do things quickly. As with many of the projects, it helps to consult an expert community with regard to which standards to follow – although there was a lot of discussion amongst the project heads on the existing suite of standards adopted by NLW. In the end the team moved outside this remit, rather than create sub-standard catalogues – however the team underestimated the metadata required to represent the project material in terms of resource discovery and rights management.

NLW had the skills for digitising and preserving digital surrogates but optical character recognition (OCR) was outsourced. The team had not undertaken large scale OCR before and were aware of the technical challenges

---

Key Message

Long-term access often requires co-operation from many staff. There is a risk that responsibilities are unclear. Consequently it is important that a senior member of staff is charged with delivering an organization's digital preservation strategy.

---

posed by Welsh language diacritics. Off-the-shelf optical character recognition software has a poor record for processing accents over letters. Although NLW has now acquired these skills, it would probably still outsource OCR for future projects.

The team had initially estimated that 30% of the pages scanned would create potential copyright challenges so, to avoid inadvertent infringements, they deployed a co-ordinates-based system to redact elements of the scanned text. On reflections, this was unnecessarily complicated and would have been easier to do manually: in fact there were fewer than 40 pages that needed to be redacted in this way.

Content
The collection comprises a selection of 50 journals from 19th century onwards. 40% of the collection is in Welsh, comprising the largest collection of Welsh language resources on-line. Journals include scholarly, literary, cultural, popular and educational titles.

Content management
The NLW used a Digital Asset Management System, VITAL by VTLS, to manage the content. This was built using Fedora and SOLR software and is also the delivery system. VITAL is a proprietary modification of an open source repository platform.

Description and metadata:
The team used the *METS*, *MIX*, e-Government Metadata Standard and Dublin Core metadata standards to describe and manage the collection.

Digital Preservation:
The NLW Digital Preservation Strategy was produced in 2008. Staff at the library had been aware for some time of the need for such a strategy, but it has only recently become a formal function. A specific digital preservation plan for the project was written. Einion Gruffudd (ICT Operations Manager) and Owain Pritchard (Senior System Administrator) are responsible for operational preservation of digital files in the digital archive, including masters, access copies and metadata. Digital preservation in general, including the NLW main Digital Preservation Strategy is the responsibility of Sally McInnes. Development of the DAMS is the responsibility of Illtud Daniel (Chief Technical Officer), and the Metadata, including METS, is the responsibility of Lyn Lewis Dafis. The preservation system is AMASS by Quantum.

Interface to Welsh Journals Online

NLW has chosen to preserve its master images in their robust digital archive system. However, the *METS* files which contain descriptive and preservation metadata and the structural information about how to construct archival information packages and the full-text *TEI* files created by merging of the OCR text and descriptive metadata are not preserved in the digital archive. The *METS* and *TEI* files exist only in *VITAL*, the content management system that manages the delivery site and provides an entry point into the digital archive.

The *METS* files are an intrinsic element of the entire archival information package. Without the *METS* files, the preserved image files become a data store without an entry point. The *TEI* files could be recreated by merging together a re-OCR of the image files and the metadata from the *METS*, and as such are not as important to long-term preservation as the *METS,* although the costs of having to redo that OCR work and recreate the *TEI* files is not inconsiderable.

The Digitisation Programme recommended that NLW place the *METS* files in the Digital Archive alongside their image counterparts and asked them to consider the costs of preserving the *TEI* files in the Digital Archive against the costs of having to recreate them. Such recommendations notwithstanding, the fact that the review team could have a detailed discussion of specific issues was because the project had already developed a detailed digital preservation plan and was clear about which member of staff was responsible for which tasks.
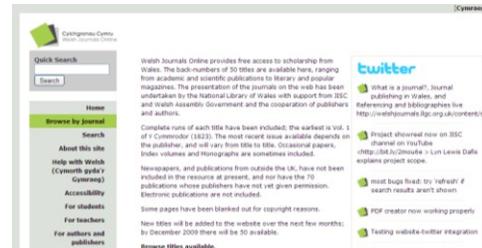
Discussion
The Welsh Journals Online project is technically complicated and although it is based in only one institution it draws on many staff with different types of

expertise. Small amounts of staff time are required over an extended period meaning that long term preservation may have to compete with other, short term priorities. But the National Library of Wales has avoided this danger by being clear about responsibilities. An organization which is clear about responsibilities and which appoints a manager at the right level is much more likely to succeed in its preservation actions than when responsibilities are ad hoc and lines of communication are unclear.

Welsh Journals Online has one of the stronger preservation plans of the whole JISC Digitisation Programme projects. They are one of the few projects that were able to supply the review team with a preservation plan: other projects referred to such plans, but did not provide the plans to the review team. NLW maintains tight control over their production workflow, with the process managed through a bespoke workflow system. This helps insure that content is not misplaced and metadata is not mis-assigned. In addition, NLW has committed to ongoing staff support of the system

beyond the length of funding. This was reflected in the preservation plan, and, in the survey of NLW staff, several commented that through the process of creating this plan they realized they must commit 25% of the time of several staff people to support the ongoing access and preservation of this content. Such planning and appreciation of the on-going nature of digital preservation is exemplary model for the management of such digitisation projects.

Welsh Journals Online can be found at: http://welshjournals.llgc.org.uk/