

## **CASE STUDY: ARCHIVES & RECORDS COUNCIL WALES DIGITAL PRESERVATION CONSORTIUM**

**January 2014**

### ABSTRACT

This case study discusses the experience of a cross-sectoral consortium of Welsh archives as they cooperated to pilot deployment of the open source Archivematica software with Microsoft's Windows Azure public cloud service. It explains the organisational context of the consortium, the varied nature of their digital preservation requirements and approaches, and their experience with selecting, deploying and testing Archivematica in the cloud. It concludes with the key lessons they learned, and discusses current proposals to secure grant funding in order to move this pilot into operation.

### ORGANISATIONAL CONTEXT

The Consortium is a project sub-group of the Digital Preservation Group within Archives & Records Council Wales (ARCW), established to ensure long-term access to digitised and born-digital material in Wales. Consortium members are drawn from many different sectors within the Welsh archives community, including national institutions, Higher Education, and Local Authorities.

A 'State of the Nation' survey, carried out in 2009 by Sally McInnes and Vicky Phillips, explored the quantity and format of digital objects being deposited with Welsh archives. Evidence from the survey informed the drafting of a business case and roadmap for phased implementation of a Welsh digital preservation initiative, submitted to ARCW in 2010.

Given the economic situation at the time, the business case concluded that the only realistic way forward was for the archive sector in Wales to work together as a distributed consortium. It was felt that this approach would enable action to be taken to address some of the urgent needs for digital preservation for digital objects being deposited in archive services across Wales. Given that no additional finance resource was available at the time, it was also felt that the consortium would need to

use free open-source software rather than commercial software needing regular licence fees, although it was recognised that even free software was likely to involve cost in the future for development effort, user support and training, etc.

The Consortium received over £35,000 in grant funding from Museums Archives & Libraries Wales (CyMAL), which enabled recruitment of a technical officer based within the National Library. This post, plus significant in-kind contribution from Consortium members, facilitated specification and evaluation of a pilot deployment of the open source Archivematica software with Microsoft's Windows Azure cloud service.

Since the Consortium was formed, its membership has grown to include representation from every Local Authority archives service in Wales, the National Library, and the archives of Welsh universities. A number of the regional Archaeological Trusts in Wales are also considering participation.

At the time of writing the National Library of Wales has applied for a significant grant from the Heritage Lottery Fund, to support delivery of a conservation and digital archives service for Wales. If the application is successful, the money will go a long way towards supporting the aims of the Consortium and its members.

## DIGITAL PRESERVATION

### **Significance**

The Consortium's membership is drawn from across a set of very different sectors, with discrete requirements that need to be balanced.

In terms of what is being deposited in local authority record offices from the public side, digital collections (particularly hybrid) are on the increase but they are still quite small, few and far between. Archivists need to act now to prepare for the future. Local authority archivists have to be proactive in setting up a suitable digital preservation repository, engaging with the public to dispel traditional views of archives being predominantly analogue and to raise awareness of the issues that come with digitally born objects and their continued survival. Otherwise this material won't be available in the future. From the internal local authority view it is also of major importance as all departments are shifting to a paperless environment. It is their duty as part of the organisation's Records Management System to preserve records of cultural importance. However in most cases there is less demand

pressure on the access side currently: most users are not asking for digitally born material from local archives yet as they are not aware of it.

Immediate access to digital material is more pressing for the National Library of Wales. Born digital is coming in all the time, so they have set up systems and the workflows to deal with it. The National Library has been digitising for 20 years now so there is also significant digitised as well as born digital material to deal with and make available.

The Consortium expects the volume of deposited material to increase dramatically over the next few years. The public will become more aware over the next few years that they need to be proactive if their material is going to remain accessible.

## **Current approaches**

Current approaches to digital archiving vary widely across organisations in the Consortium, and one of its aims is to share best practice and to harmonise approaches where feasible:

- The National Library of Wales has the most developed internal systems, based upon the Virtua library management system and VITAL repository, both from VTLS;
- In Conwy, digital preservation is now a formal part of the Authority's information management strategy. The archive has been testing a copy of Archivemata installed locally on the council network, as part of the Consortium's pilot project in this area (below).

Current approaches within the consortium have focussed on the following areas:

- Specifying the functionality required for the shared digital preservation software. Having considered the merits of the available open-source archiving software options including Isladora, DAITS, Xena, and Archivemata, the consortium concluded that Archivemata was a best fit and it was thus selected as the preferred option;
- Testing the beta release version of Archivemata, its hosting in the cloud, and use of cloud storage;
- Creating some pre-ingest documentation, e.g.:
  - o Hardware recommendations for digital preservation and Archivemata;

- o ITC business case – something to be given to IT departments;
- o Depositors form – to enable services to capture all of the necessary information;
- o Guidance sheet for depositors.
- Organising a number of training days in both the North and South of Wales. The aim of which has been to build up skills and raise awareness in the region.

In terms of current technical activities the consortium are presently looking at Archivemata, ICA Atom (cataloguing software bundled with Archivemata), and Cloud Storage. They have been working with an external consultant to pilot and test a cloud based infrastructure initially using a cloud hosted instance of Archivemata and Microsoft Azure storage. Locally installed copies were problematic, as Archivemata is still in beta release, so some local authority IT departments were unhappy about installing and maintaining it on their infrastructure. The cloud pilot was of short duration as it cost £200 per month for the cloud deployment. The feasibility of a cloud solution for archives in Wales has been successfully tested as a proof of concept. However, there is much still to do for a full implementation.

### **How they would want this to change over the next 3 years**

The consortium feels they have proved that the way ahead is a browser based interface to a shared cloud installation of Archivemata. They would like to go this way in future as it was a successful proof of concept. The National Library has applied for a grant from HLF for digitisation and conservation. If successful, as part of this they would like to see the Library working together with organisations throughout Wales to provide a national cloud-based infrastructure with browser based access to Archivemata software and linked cloud storage.

### **Range of content types and volumes of digital material**

Consortium members accession a range of data types, and cope with data at a range of scales from the relatively small to multiple terabytes.

Conwy County Borough Council, for example, currently holds under a terabyte of digital material deposited by members of the public. This collection is dominated by digital photographs, but also includes oral histories, moving images, and word-processed documents and spreadsheets. Most deposits use current file formats and

storage media, but some are also made in obsolete formats and on older media such as floppy disks.

The Consortium has undertaken some work to agree normalisations for storage formats, initially concentrating this activity on the most straightforward – and common – Office file formats.

## CLOUD STORAGE FOR DIGITAL PRESERVATION

### **Microsoft Windows Azure**

Given the mixed nature of members' IT infrastructure, and some discomfort with installing beta release versions of Archivemata software on institutional infrastructure, the Consortium was keen to explore mechanisms for hosting the pilot and its data with an external cloud service provider. Under guidance from an external consultant who was most comfortable working in a Microsoft environment, they initially selected Microsoft Azure as a cloud platform for the pilot. As Archivemata is designed to run on Linux, the team needed to add a virtualisation layer between the cloud and the software, adding an additional element to the cost and complexity of their implementation. Although the pilot was successful, they would wish to evaluate other cloud services providers such as CloudSigma at a later stage. There is no money to pursue this further at the moment and they want to wait for the final release of Archivemata before exploring Cloud service options further.

Cloud services such as Windows Azure (and its competitors) tend to bill each month for capacity that has actually been consumed. As a result, the consortium found it can be difficult to budget ahead, or to accurately predict the amount of data likely to be uploaded, stored, or downloaded; each of which can incur a separate cost. This lack of predictability is a particular problem in the early stages of a project where usage patterns have not stabilised. Applications often need to run for several months before you can gain a realistic view over cost, then using this knowledge to predict future spend.

### **Archivemata**

The version of Archivemata used during the pilot was a beta or pre-release version of the software. The pilot team encountered a number of issues that they reported and expect to see addressed in the formal 1.0 software release this year.

## TECHNICAL INFRASTRUCTURE

### **Main software systems used for electronic content management, preservation and access services**

The core systems used today vary across the Consortium. Local Authority archives tend to use Calm and CalmView, while the National Library has its VTLS system. Archivematica currently relies upon AtoM, which is also used at Aberystwyth University. Consortium members are exploring the feasibility and cost of either extending AtoM to meet their needs, or having Archivematica modified to interact more seamlessly with Calm. Completely new alternatives such as ArchivesSpace are also being considered.

The Consortium continues to work through the practicalities of offering both shared and private storage within the system, enabling partners to collaborate and pool resources when appropriate whilst retaining secure areas for their own material.

## BUSINESS CASE AND FUNDING

In 2014, the Consortium's business case remains broadly in line with the vision laid out back in 2010. Consortium members – and Wales – need new solutions to meet the growing demand for preservation of and access to digital material. Given the disparate nature of archival bodies across Wales, and the realities of funding for the sector, a lightweight and collaborative solution remains desirable. Consortium members agree that an externally hosted application could ultimately be simpler and more cost-effective to deploy than multiple local installations of software. A cloud-based solution appears to meet the Consortium's stated requirements, and a pilot project has proven it to be viable.

Funding to date has been in the form of three annual innovation grants from CyMAL (£35,000 in total), plus significant in-kind contributions from Consortium members. EU funding has been explored as a supplement to the CyMAL grants, but the process has so far proved too complex to justify the likely return. The Consortium will continue to apply to CyMAL for new grants, and these can support the development of the system. Members recognise the need to find a sustainable way of funding ongoing delivery of services once they are operational, and various forms of usage-based charging have not been ruled out. A National Library of Wales submission to the Heritage Lottery Fund will, if successful, provide some funding to accelerate development of the system.

## KEY LESSONS THEY HAVE LEARNT

- The original business case in 2010 was very useful in terms of providing a roadmap for what they needed to do. They have aligned efforts with it and it has positioned them well;
- Communication is important. Getting people involved and maintaining the level of interest needs a dedicated resource. It helps that Wales is a small country;
- A 'Community of the Willing' means there is a lot of variability in terms of what voluntary effort people are able to offer. The voluntary effort needed for the consortium has to be secondary to their primary functions and everyday work commitments;
- People will eventually realise that in some ways digital preservation is not new – it's just an expansion of what they are already doing for physical materials in new ways;
- Make a concerted effort to get organisation IT technical specialists on board and involved as they can help clear a path through organisational red tape. They are an asset when anything moves beyond the technical scope of archivists' professional training;
- The consortium needed a solution that was independent of the individual members and acceptable to all. A cloud-based solution provided this;
- The pilot did work, so they were pleased, but issues around discoverability, security, and providing separate cloud areas for different consortium members still need to be worked out for a full implementation;
- Integration between systems can be a significant issue;
- There were some challenges around being the early adopters of pre-release software.

## FUTURE PLANS

The Consortium is hosting an Archivemata Go Live launch event in February 2014 to further publicise the work undertaken so far. They will also put a grant application in to CYMAL for next year.

Over the next year the aim is to roll it out more widely in the community (they are waiting for full release of Archivemata in February 2014). Following this, they need to concentrate on integration with CALM and with the other tools in use by

stakeholders. The main development areas are likely to be around ATOM, training, the cloud storage and hosting, plus extra development work. Alongside Microsoft Windows Azure, the team will evaluate alternative cloud platforms. A limited evaluation of CloudSigma has already begun.

If the phase 1 application of the HLF grant application is successful, they will move to a phase 2 application. A lot of consultation with stakeholders will be necessary to define what the priorities should be.

The recent Williams Commission Report on reorganisation of Welsh local authorities raises a further set of real-world challenges with which any production system would have to cope; if geographic boundaries shift, then part or all of an archive may need to be transferred from one authority to another. A shared system might make this transfer of ownership more straightforward to achieve than was the case in earlier reorganisations.

#### FURTHER INFORMATION

Access to Memory (AtoM): <https://www.accesstomemory.org/>

Archivemata: <https://www.archivemata.org/>

Archives & Records Council Wales (ARCW): <http://archiveswales.org.uk/>

ArchivesSpace: <http://www.archivespace.org/>

Calm: <http://www.axiell.co.uk/calm>

CloudSigma: <http://www.cloudsigma.com/>

Islandora: <http://islandora.ca/>

Microsoft Windows Azure: <http://www.windowsazure.com/>

*Survey of Digital Preservation in Wales: Summary of Findings* (Sally McInnes & Vicky Phillips, November 2009)

[http://www.llgc.org.uk/fileadmin/documents/pdf/DP\\_Archives-Wales\\_Report-final.pdf](http://www.llgc.org.uk/fileadmin/documents/pdf/DP_Archives-Wales_Report-final.pdf)

Virtua: <http://www.vtls.com/products/virtua>

VITAL: <http://www.vtls.com/products/vital>

Xena: <http://xena.sourceforge.net/>