

CASE STUDY: ARCHIVES AND RECORDS COUNCIL WALES DIGITAL PRESERVATION WORKING GROUP

January 2015

ABSTRACT

This case study discusses the experience of a cross-sectoral group of Welsh archives as they cooperated to test a range of systems and service deployments in a proof of concept for cloud archiving. The proof of concept has examined the open source Archivemata software with Microsoft's Windows Azure; Archivemata with CloudSigma; Preservica Cloud Edition; and has begun testing Archivemata with Arkivum 100. It explains the organisational context of the Working Group, the varied nature of their digital preservation requirements and approaches, and their experience with selecting, deploying and testing digital preservation in the cloud. It concludes with the key lessons they learned, and discusses current proposals to secure funding in order to move this pilot into operation.

ORGANISATIONAL CONTEXT

The ARCW Digital Preservation Working Group, a consortium within Archives & Records Council Wales (ARCW), was established to ensure long-term access to digitised and born-digital material in Wales. Group 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 Group 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 Group 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 Group members, facilitated specification and evaluation of pilot deployments of the open source Archivematica software with cloud services at Microsoft's Windows Azure, CloudSigma; and Arkivum 100; and finally pilot deployment of Preservica Cloud Edition.

Since the Group 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. The significance of the work has been demonstrated through the short listing of the ARCW Digital Preservation Working Group for an award by the Digital Preservation Coalition.

In 2014 the National Library of Wales applied for a significant grant from the Heritage Lottery Fund (HLF), to support delivery of a conservation and digital archives service for Wales. Although the National Library of Wales was unsuccessful in its application to the HLF, it is drawing up a business case to procure a digital preservation solution for ARCW members. The business case will reflect ARCW requirement and be based upon the evidence regarding functionality, interoperability and suitability that it has gained from testing and pilot projects.

DIGITAL PRESERVATION

Significance

The Group'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 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 Group 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 Group, 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 Virtual library management system and VITAL repository, both from VTLS. The Library is a partner in a consortium of Welsh libraries which have selected Ex Libris as its next Library Management System and it will transition to this over the course of 2015;
- In Conwy, digital preservation is now a formal part of the Authority's information management strategy. The archive has been testing a copy of Archivematica installed locally on the council network, as part of the Group's pilot project in this area (below).

Current approaches within the Group 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 Islandora, DAITSS, Xena, and Archivemataca, the Group concluded that Archivemataca was a best fit and it was thus selected as the preferred open-source option;
- Testing the beta release version of Archivemataca, its hosting in the cloud, and use of Microsoft Azure and then CloudSigma cloud storage;
- Testing of Preservica, the digital preservation solution developed by Tessella, has been undertaken by the Group. This has enabled members to have a hands on experience with the system, from ingest to discovery through its Universal Access Portal; and has enabled members to gain a better practical understanding of how digital preservation theory can be applied for example the OAIS model from SIP to DIP.
- A further trial of Archivemataca will start in January 2015, hosted by Arkivum. This will enable a comparison of the two systems to be undertaken, with regard to functionality, usability and costs;
- A further option, the National Library of Wales hosting a solution, is also being considered.
- Creating some pre-ingest documentation, e.g.:
 - o Hardware recommendations for digital preservation and Archivemataca;
 - 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 Group are presently evaluating the different options against the criteria they have established with the aim of moving to procurement next year.

How they would want this to change over the next three years

The Group feels they have proved that the way ahead is a browser based interface to a shared cloud installation. They would like to go this way in future as it was a successful proof of concept. A business case will be drawn up which will be based upon the findings of the appraisal process. This business case will be used to create a tender document. Potential partners will be identified for collaborative bids. They would like to see organisations working together throughout Wales to provide a national cloud-based infrastructure with browser based access to digital preservation software and linked cloud storage.

Range of content types and volumes of digital material

Group 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 Group 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 Group 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 also evaluated another cloud services provider (CloudSigma) at a later stage.

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 Group 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.

The Proof of Concept work with Arkivum will enable the Group to understand the possible ways that a hosted AtoM/Archivemata/Arkivum solution could work and to assist with costing models.

Archivemata

The version of Archivemata first 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 were addressed in the formal 1.0 software release. As of January 2015, Archivemata is on release 1.3. The Group is sponsoring work looking at the retention of a SIPs hierarchical structure and including levels of description for release 1.4. An Archivemata UK group has been established which will assist with an integrated approach by UK users.

TECHNICAL INFRASTRUCTURE

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

The core systems used today vary across the Group. Local Authority archives tend to use Calm and CalmView, while the National Library has its VTLS system (to be replaced with Ex Libris later this year). Archivemata currently relies upon AtoM, which is also used at Aberystwyth University. Group members are exploring the feasibility and cost of either extending AtoM to meet their needs, or having Archivemata modified to interact more seamlessly with Calm.

Another consideration is understanding how any solution will interact with ERDM systems locally and practical issues such as collaboration within the institutions on transferring electronic records.

The Group 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. The Group is looking at several Cloud based models including: a single shared instance of Archivemata and AtoM; multiple instances of Archivemata pointing to a single instance of AtoM; hybrid instances of cloud and onsite (both/either Archivemata and AtoM) providing flexibility allowing institutions to restrict access of some material in-house; and each institution having their own dedicated Archivemata and AtoM instances. It is also considering Preservica Cloud and Preservica On premise options, as well as a potential service hosted by the National Library of Wales.

BUSINESS CASE AND FUNDING

In 2015, the Group's business case remains broadly in line with the vision laid out back in 2010. Group 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. Group 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 Group's stated requirements, and a pilot project has proven it to be viable.

Funding to date has been in the form of four annual grants from CyMAL, plus significant in-kind contributions from Group 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 Group 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.

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;
- The ARCW Digital Preservation Group is in effect a ‘Community of the Willing’ and this means there is a lot of variability in terms of what voluntary effort people are able to offer. The voluntary effort needed for the Group has to be secondary to their primary functions and everyday work commitments;
- The pilot projects and Proof of Concept proposals have been very useful in enabling hands on experience of systems, but central technical support is required for ARCW members to ensure that the shared systems can work with internal systems;
- 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 Group needed a solution that was independent of the individual members and acceptable to all. A cloud-based solution or a hybrid solution may provide this;
- The proof of concepts did work, so they were pleased, but issues around discoverability, security, and system configuration (if there should be a single instance, multiple instances, cloud or hosted solutions) 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
- Ongoing advocacy is needed to get digital preservation to be 'business as usual' and accepted as part of archivists' jobs.

FUTURE PLANS

The Group hosted an Archivematica Go Live launch event in February 2014 to further publicise the work undertaken so far. This was a very successful. Another event is to be held this year to consider issues relating to system operability. Options for cataloguing, discovery, integration with EDRMS systems will be considered. Any digital preservation solution must be able to work with cataloguing and discovery infrastructures and to integrate with other systems. They will also put a grant application in to CYMAL for next year.

The next stage of the project is to procure a digital preservation software solution which is appropriate for the needs of ARCW members. A lot of consultation with

stakeholders will be necessary to select a system which is suitable for ARCW members' needs. The definition of the requirements will be based upon an evaluation of the various proof of concept including testing of Preservica, the solution

developed by Tessella; and the Proof of Concept in which Archivemata will be hosted by Arkivum, the provider of data archiving solutions. A further option, of a solution hosted by the National Library of Wales, may also be considered. The evaluation of these options will provide evidence regarding the suitability of differing configuration options, such as single or multiple instances, cloud or hosted options, costing implications and integration with other systems. A business case will be constructed using this evidence which will be used as a basis for a tender document.

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/>

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

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