

## CASE STUDY: THE PARLIAMENTARY ARCHIVES

January 2014

### ABSTRACT

This case study covers the Parliamentary Archives and their experience of procuring via the G-Cloud framework and running public cloud storage as part of their digital preservation infrastructure. For extra resilience/an exit strategy they have selected two cloud service providers with different underlying storage infrastructures. The archive is not storing sensitive material in the cloud and is using local storage systems for that material. It has a locally installed preservation system (Preservica Enterprise Edition) which is integrated with cloud and local storage. As such it is an example of an archive using a hybrid set of storage solutions part-public cloud and part-locally installed for digital preservation.

### ORGANISATIONAL CONTEXT

The Parliamentary Archives manages, preserves, and provides historic access to the archives of the House of Lords, the House of Commons and to other records relating to Parliament. They also provide a records management service for both Houses. Its remit is to preserve and make available its collections to the public, and they operate a full archive service and public search room. Digital is part of this equation, and they have an established digital repository and digital preservation policy and strategy. The Parliamentary Archives is a shared service of both Houses of Parliament and is not subject to the Public Records Act.

In 2012 Parliament developed a new ICT policy, specifying Cloud First. In future, when procuring new or existing services, Parliament will consider and fully evaluate potential cloud solutions first – before they consider any other option. The Parliamentary Archives became early implementers of the new policy when reviewing their digital preservation storage requirements.

They were the first department in Parliament to procure via the G-Cloud framework. The parliamentary procurement office managed this process for them.

## DIGITAL PRESERVATION

### **Significance**

Digital material is of major significance in their collections and they have support and buy-in for their work in digital preservation from management boards of both Houses. There is a lot of digitisation by Parliament as well as born digital material. The digitisation programmes include original archives and historic editions of publications such as Hansard and other parliamentary printed material. The significance, range and volume of digital material for preservation can only increase.

### **Current approaches**

They have established Record Disposal Practices which define retention policies of Parliamentary records. The records management team is based within the Archives, so they can coordinate well and they have a joined up approach to selection policy and archiving.

They have a locally installed preservation system, Preservica Enterprise Edition, which is integrated with cloud and local storage (see below).

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

They are now entering the last phase of the digital preservation project, which is delivering enhancements to SDB and will end on 31 March 2014. Following that, digital preservation will move fully on to a “business as usual” footing. Up to now they have been concentrating on ingest, so over the next three years they will be developing a process for preservation planning and will be thinking more about how to promote the use of the material and broaden the range of information systems from which they can ingest.

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

They currently have about 50 Tb of priority material to ingest. Over the next three years the quantity will only increase, and they will have more complex formats to handle. Digital material currently includes Hansard, web archives, EDRMS records, and Standing Committee papers. Formats are mostly standard office types, as well

as PDF, JPEG, TIFF, Audio Visual (AV) and CAD material, plus web archive files, and XML structured data exported from internal systems. The amount of potential AV material is huge, depending on future decisions about selection. There is also a separate analogue AV archive that could be digitised.

## CLOUD STORAGE FOR DIGITAL PRESERVATION

The experience of procuring and managing cloud storage for preservation has been informative. They identified their requirements and reviewed cloud storage options in light of them. The main issues were maintaining their ability to fulfil legal obligations such as Freedom of Information, sovereignty, managing data integrity, information security, and getting the data back in the event of business failure or a decision to change provider, etc.

There was some initial concerns about data security in the cloud but they took the decision only to use the cloud for storing open data, which is already in the public domain. Remaining material is stored locally. They may review this decision in future with a view to using relevant cloud service providers which are accredited for storing higher impact level material.

There were also concerns over dependencies on a single service provider. For example, in the event of business failure, so for risk management purposes they have chosen to use two cloud providers in parallel with different underlying technologies. One is using Amazon S3, and the other is based on EMC Atmos.

They found the process of procurement through G-Cloud itself to be very straightforward. What was more complex first time around was agreeing the contract as the standard terms and conditions available at the time did not have the safeguards they desired on getting the data back in a timely fashion on exit. The G-Cloud framework has since been updated and the procurement for the second cloud service was quicker. They have also realised that while there are a lot of suppliers on G-Cloud, many are re-sellers of the same underlying cloud service e.g. Amazon S3, so there is not as much choice of underlying infrastructure as first appears. They started to use the system operationally with the first cloud provider in August 2013. The second provider is due to come on stream in 2014.

## TECHNICAL INFRASTRUCTURE

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

For most preservation functions they are using Preservica EE out of the box, but have needed some configuration and enhancements for ingest from specific local systems. Preservica EE integrates with CALM, their archive cataloguing system. As noted above local archival storage is supplemented with that from two cloud service providers. Their Cloud storage is predominantly a deep archive and is not used by end users as they access separate copies. There is a bespoke online delivery system to provide public access to repository content, which integrates with the archive catalogue, Portcullis.

They are taking in material from a variety of in house systems but like most public sector bodies Parliament tends to standardise on Microsoft for most office functions so these predominate.

## BUSINESS CASE AND FUNDING

### **Main issues in their business case for cloud storage for digital preservation**

The wider business case for preservation was already in place before the Cloud First policy was instituted, but had to be revised to reflect the different cost model of less upfront capital investment, but more ongoing revenue expenditure. This wasn't a difficult case to make as Cloud First was a strategic decision Parliament had already adopted. Their initial budget profile needed reworking however in light of experience of use. They needed to predict volumes and usage and found it is important to get good figures here, as typically you pay most for your highest volume direction (in or out). They are mostly ingesting as the Cloud storage is not directly accessed by end users. They have a digital asset register so they can predict what will be coming in, storage demands, and future costs, but a lot depends on how quickly you can ingest.

## KEY LESSONS THEY HAVE LEARNT

1. Look carefully at the contractual arrangements for your exit strategy. Note however within framework agreements, you are quite limited in what changes you

can make to the terms and conditions as these have already been defined in the pre-selection phase.

2. Spend a lot of time getting your requirements right before you start.
3. Suppliers may cite excellent durability figures, but their claims are not always scientifically based. It can be difficult to define your durability requirements in a way that allows you to assess suppliers against them.
4. The quality of your information about likely usage is fundamental for budgeting for your use of cloud service providers. Try to establish accurate figures for your future storage and activity levels. A digital asset register can help here in assessing future ingest requirements and likely costs if you are primarily using cloud for deep storage.
5. The ongoing revenue commitment for the cloud on the basis of what you use, as opposed to a big upfront capital investment for local IT infrastructure, has pros and cons. It is important for management to understand and endorse the different cost model.
6. Despite early concerns, their experience of the cloud has been very positive. Other archives can take confidence from their success in working through practical approaches to using cloud for digital preservation and to addressing the most common issues raised.

#### FUTURE PLANS

1. They will be implementing the second cloud storage provider.
2. They are only storing impact level 0 public data in the cloud at the moment, but might consider storing closed data in the cloud when appropriate supplier certification is in place.

#### FURTHER INFORMATION

Parliamentary Archives

<http://www.parliament.uk/business/publications/parliamentary-archives/>

Digital Preservation in Parliament

<http://www.parliament.uk/business/publications/parliamentary-archives/digitalpreservation/>