InterPARES Trust Project Report

Title and code:		AA03 Dark Repositories as a Service
Document type:		Final report
Status:		Final
Version:		1.0
Research domain:		Infrastructure
Date submitted:		2 July 2018
Last reviewed:		
Author:		InterPARES Trust Project
Writer(s):		Gillian Oliver
Research team:		Gillian Oliver, Aradhana Hamilton

Document Control

Version history			
Version	Date	By	Version notes
1.0	2 July 2018	Gillian Oliver	

Table of Contents

Abstract **4**
Research team **5**
Background **5**
Objectives **6**
Methodology **6**
Findings **6**
Conclusions **7**

Abstract

This research project represents the first step in investigating the feasibility of archival institutions providing 'dark archives' as a service to other agencies. A survey was developed to investigate current experiences and policies of memory institutions internationally with regard to the network-accessibility of their master digital repositories and whether or not a 'dark' repository approach has ever been considered or trialled and the results of those considerations/experiences and the appetite for considering such approaches in future.

Most survey respondents were public sector archival institutions. Respondents reported that policies relating to digital preservation were inadequate. There was very little outsourcing of digital preservation functions. A minority of respondents were using a dark archive service model.

Dark Repositories as a Service

Research team

Gillian Oliver, Victoria University of Wellington
Graduate Research Assistant: Aradhana Hamilton, Victoria University of Wellington

Background

Most institutional repositories of digital preservation masters (or AIPs) are network accessible – albeit with various security safeguards such as firewalls, etc. As long-term storage requirements preclude the option of storing master records in encrypted form because of the problems associated with long-term key management, this leaves the records vulnerable to hacking by unauthorised external parties. Even the most secure repository can be hacked by determined groups or individuals if there is some kind of network connection.

The only way of truly protecting your institutional repository from external cyber attack is to have a fully 'dark' or offline digital repository with 'air gaps' between the external network and the master repository. One archival institution that has done that is the National Archives of Australia (NAA). This approach taken by NAA has come at a real cost in terms of operating efficiency – it means that NAA digital archiving processes are quite manual and cumbersome. The NAA feels that this has been a justifiable cost associated with managing the risks associated with ensuring the integrity and security of the important and often sensitive records in its repository. This has been doable for the NAA, because their digital archive is in-house built, owned and operated. Nevertheless, the NAA is now planning to move away from having a totally 'dark' digital archive in order to achieve operational efficiencies. But would a 'dark archive' approach be achievable in an 'as a service' operating model, where the digital repository might be operated on behalf of the archival institution by a third party?

Current thinking at Queensland State Archives (QSA) is to have a secure dark-fibre, network-accessible digital repository for our AIPs, with a fully 'dark' backup in the form of offline tapes stored in a climate-controlled repository within the walls of QSA's building at Runcorn, Brisbane. The existence of the offline backup tapes would provide confidence that, should the network-accessible repository ever be hacked, QSA could reconstitute it from the authentic, un-hacked offline backup tapes. This addresses the issue of ensuring the authenticity and integrity of the master AIPs. It does not, however, manage the risk associated with closed-access, sensitive records in the repository being revealed to the world as a result of a cyber attack – and the consequential reputational and other risks associated with that. Nevertheless, it may be a reasonable compromise.

Another option is to have a hybrid model whereby highly sensitive AIP records and metadata are managed separately from the bulk of less sensitive content by running a small 'dark' repository in parallel with a larger network-linked repository.

Objectives

The objective of this project was to gain understanding of the attitudes and concerns of archival institutions internationally relating to digital preservation, with a view to establishing whether there was any appetite for the concept of developing dark archives as a service for particular jurisdictions.

Methodology

An online survey was developed using Qualtrics. The survey was sent via email to 237 public and private archival institutions located in the United States of America, Europe, Australasia and Canada. The institutions ranged from local government, national and cultural to corporate (banking and company) archives. Forty-nine surveys were started and 45 complete responses recorded, a response rate of 21%.

The survey consisted of 40 questions, divided into seven sections. Questions asked about digital preservation activities, attitudes to outsourcing, issues relating to cyber security and use of dark archives as a preservation model. The survey was completed anonymously and granted human ethics approval by Victoria University of Wellington, School of Information Management's Human Ethics Committee.

The survey consisted of a mixture of multi-choice, text entry, rank-order and matrix table questions. Some multi-choice questions had free-text entry boxes, which allowed the respondent to further elaborate on their answers.

Findings

The main findings from the survey were as follows:

- The vast majority (86%, 33 out of 38) identified that they were public sector archival institutions.
- The majority of archives that participated in the survey (66%, 25 out of 38) have a digital archive.
    - Out of the 34% (13 out of 38) that said they didn't have a digital component to their archives, 91% (10 out of 11) responded they were interested in developing one.

- The results appear to show a lack of effort in digital preservation, 65% (average, 34%, 12 out of 35 and poor, 31%, 11 out of 35 collectively 23 out of 35 or 65%) rated the policies they had in place as average to poor.

        - There appears to be a risk of losing digital data stored on out of date hardware, 90% of respondents said they had information stored on hardware that was obsolete or at risk of becoming so.

- Seventy-eight percent (78%, 28 out of 36) of respondents do not currently outsource their digital preservation functions, out of those who do not 65% (13 out of 20) said they would consider outsourcing this function to an external third party.
    - Cost was the most widely held concern by respondents (10 out of 19 respondents included it in their answer), followed by security (6 out of 19 respondents mentioned this as a concern).

- On a scale of 1 to 7 (1 being very concerned) 53% of respondents (18 out of 34) picked 1 or 2 on the scale. Indicating that cyber security is a big concern for the majority of respondents. However when asked how much of an issue they thought it was 59% (20 out of 34) said it was only *somewhat* of an issue in their jurisdiction.
    - This shows that security is a large consideration, considering the level of concern compared to the apparent risk of a cyber attack.

- 33% (10 out of 30) of respondents answered they used a dark archive service model to protect their digital material.
    - 67% (20 out of 30) said they didn't use a dark archive model, out of these respondents 63% (10 out of 16) answered that they hadn't thought of using this model to preserve born-digital material. This could potentially signal an opportunity to supply a dark service model to archival institutions globally.

- 60%, 3 out of 5 private sector respondents have a dark archive in place to secure their born-digital material.

Conclusions

The majority of respondents (65%) that did not already outsource their digital preservation needs said they would consider doing it, either now or in the future. If they were to outsource their major concern would be cost followed by security.

Responses gathered from public archival institutions vary from those operating in the private sector. This is logical as the influencing factors affecting these sectors would be different. Public sector respondents were concerned about cost and access; outsourcing may be a cheaper alternative to ensuring data will not become inaccessible. Private sector archives may be less likely to outsource because of their concerns over losing control over the material. In saying this, more definitive research needs to be done on this topic.

The data shows sixty-six percent of archival institutions that participated in this survey have a digital archive. Of the remaining thirty-four percent the vast majority (over 90%) have an interest in developing one either now or in the future. The largest barrier for archival institutions appears to be both a lack of resources, primarily monetary in form as well as a deficiency in the required expertise.

The results concerning the preservation of born-digital data were varied. Over ninety percent of respondents have information stored on hardware that is obsolete or at risk of becoming so in the near future. Sixty percent of respondents have information they can't access. Thirty-one percent regarded their efforts in terms of digital preservation as *poor*. Twenty-nine percent of archival institutions have no off-site master copies of data that can be used in the event of a disaster.

These responses paint a fairly bleak image of the state of digital preservation for the majority of archival institutions. It is not surprising that *access* is the primary issue that concerns archival institutions. In the event of fast-paced technological change the information hasn't been destroyed, it has merely been lost due to lack of access. Private sector organisations primarily consider *losing control over archived material* as their biggest concern.

Archival institutions appear to be aware and proactive about cyber-security. Even though the threat of both an internal and external attack was by the vast majority thought to be *unlikely*, organisations were still concerned to some extent about an attack. This suggests archives take cyber-security very seriously, indicating that they would likely take a strong interest in security if they were to outsource.

Responses about dark archives were positive. Comments from respondents that had implemented a dark archive model for their digital preservation needs describe it as *quick*, *simple* and *secure*. One respondent commented that it was "*cost effective*". This is notable considering that *cost* has arisen as one of the major concerns and is a huge limitation for the majority of archival institutions surveyed.
This research recommends that the archival landscape could benefit from a digital preservation service, especially considering the current efforts. If a service were to be implemented, there may be potential for this to be in the form of a dark archive model. Holistically, the results suggest *cost* and *security* would likely be the primary reasons why archival institutions would not outsource their digital preservation needs.