Adoption of Infrastructure-as-a-Service at the National Library of New Zealand

Cynthia Wu; National Digital Heritage Archive, National Library of New Zealand; Wellington, New Zealand

Abstract

The National Library of New Zealand, along with Archives New Zealand, is part of the wider Department of Internal Affairs (DIA). The DIA has been charged by government with being the lead agency for cloud technology and implementation of Infrastructure-as-a-Service (IaaS). The National Library of New Zealand, in particular its digital preservation programme, the National Digital Heritage Archive (NDHA) is one of the early pilot groups in the IaaS project. This paper will seek to illustrate some of the main drivers and rationale behind the adoption of IaaS and the selection of NDHA as the pilot group. It will examine the concerns expressed by various stakeholders within the National Library, such as infrastructure availability, performance, security, support. Furthermore, it will explore the issue of trust and how this can be managed in order to ensure the continued care of digital objects with significant cultural value. By analysing in detail the scope, processes, and resources required within each distinct phase of the project, this paper will demonstrate both the infrastructure migration method and the consideration put in place to alleviate identified risks and concerns.

Background

In early 2008, the New Zealand economy entered into a period of recession and the economy was further weakened by the global financial crisis in 2009 [1]. The New Zealand government has been, and still is, in a tight fiscal position. Part of the New Zealand Government's response to this crisis was to transform its public sector to realise significant cost savings and economies of scale via shared services. In August 2011, the New Zealand government announced the adoption of cloud computing across all government departments, with the use of Infrastructure-as-a-Service (IaaS) paving the way [2]. It is expected to save the New Zealand government $250 million dollars on Information and Communications Technology (ICT) spending in the next ten years. With IaaS in place, it sets the foundation to implement Storage-as-a-Service, Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS) in the future, and enables agencies to focus more on their core business, thus gaining efficiency in the delivery of ICT services across the government and supporting better services to the public.

The Government Chief Information Officer (GCIO), a role delegated by government to the Chief Executive of the Department of Internal Affairs (DIA), was tasked with leading the cloud technology adoption project. As the lead agency, DIA will be developing a cloud service deployment strategy, thus providing an implementation model for other departments within government [3]. The National Library of New Zealand (NLNZ), in particular its digital preservation programme, the National Digital Heritage Archive (NDHA), was selected as one of the early pilot groups in this IaaS implementation project.

All of Government IaaS is provisioned by Datacom CSG Limited, Revera Limited, or IBM Limited. These three companies were awarded the IaaS syndicated contract from October 2011. DIA Government Technology Services (GTS), who administer the National Library's computing infrastructure selected Revera to supply this initial IaaS implementation for the NDHA.

Since the National Library's current site opened in 1987, NLNZ has been operating and maintaining its own data centre onsite to service the Library's technological needs, which include NDHA servers and storage for the preservation of NLNZ digital collections. The NDHA programme has been in operation since October 2008. As NDHA business processes mature and the National Library's digital collections increases, the infrastructure requirements for the NDHA have changed significantly. To date, the NDHA holds over 8.8 million files spanning over 100 different formats. NDHA permanent repository is also growing rapidly, both in numbers of files and in size of the repository (Figure 1 & 2). The growth of NDHA storage capacity needs has surpassed the capability of NLNZ's own data centre. With the adoption of IaaS the NDHA will be able to leverage current technology and organisation-wide initiatives to transform existing infrastructure and hardware architecture to address the rapidly growing technological demand on NDHA.

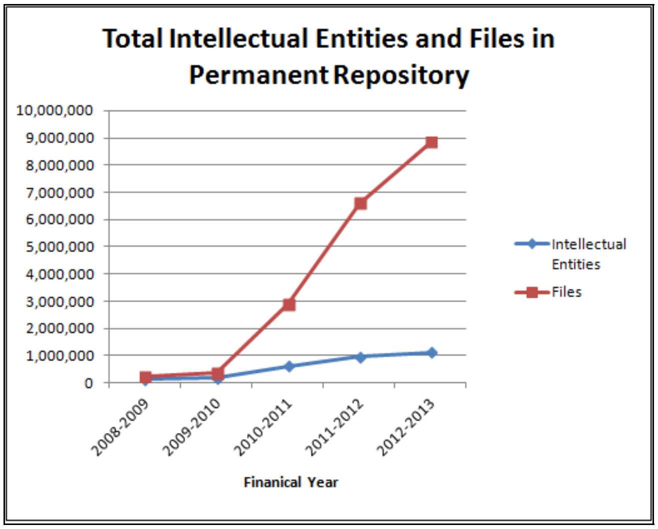

***Figure 1** Increase in number of IEs and files stored within NDHA*

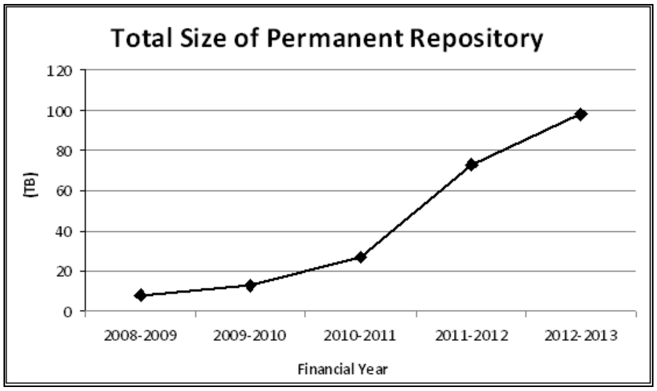

Figure 2 *Growth of NDHA permanent repository*

Risks & Assurance

Under the National Library Act (2003), the Library's legal mandate to collect, preserve, and protect documents, particularly those relating to New Zealand, and making them accessible for all the people of New Zealand, in a manner consistent with their status as documentary heritage and taonga (treasure) was extended to also include digital content [4]. NLNZ collects a wide range of born-digital material, such as websites, manuscripts, photographs, cartoons, annual reports and many digital publications. NLNZ actively collects these materials under the National Library Act, but also relies on public donations of digital material that are not published or available online. NLNZ also has its own in-house image and sound digitisation programmes. Almost all of the collected digital materials are currently preserved in the NDHA. As the IaaS project affects all digital collections stored in the NDHA, assurance must be provided to a wide group of internal stakeholders within the National Library.

Internal Stakeholders

To successfully implement the IaaS project, NDHA must balance the tension between moving towards external service provision, and ensuring that all the digital collections are still appropriately preserved, protected and accessible. Key stakeholders from the National Library were invited to review high-level planning information, and identify any specific risks and concerns relating to IaaS implementation. These stakeholders include:

- National Library Leadership Team – This team include the National Librarian of the National Library, the Chief Librarian of the Alexander Turnbull Library (ATL), and other senior staff members within the National Library [5].
- Guardians Kaitiaki of the Alexander Turnbull Library (ATL) – They provide assurance to the people of New Zealand that the ATL collections are held in perpetuity, under suitable and separate accommodation, preserved, developed and made accessible, thus maintaining the character of the services of ATL as a research library [6].
- Te Komiti Māori – An advisory body set up to provide independent advice and experience to the National Librarian on matters relating to Māori. [7].
- Library and Information Advisory Commission (LIAC) – This group provides advice to the Minister of Internal Affairs on library and information issues in New Zealand, including access to library and information services, and the role of these services in the cultural and economic life of New Zealand [8].

Risks and Concerns Identified

The Library's key stakeholders identified the following concerns to be addressed during the IaaS project either through the development of Service Level Agreements (SLAs) or operating procedures developed during project implementation:

- Data ownership
- Data security and access policies
- Impact on service delivery and performance
- Physical risks (such as earthquake and fire) at new data centre
- Network Dependency between Library and new data centre
- No backup and restore routine for current onsite data centre during project implementation
- Loss of digital collection items during transfer
- Backup and restore routine at new data centre
- Lack of disaster recovery procedures at new data centre
- Service Level Agreement (SLA) does not cover the Library's requirement
- No clear distinction between SLA between the Library and GTS (the DIA branch responsible for ICT support and management), and the SLA between GTS and the hosting company Revera.
- Lack of risk management procedure during IaaS project
- Potential unclear demarcation between faults and issues that should be addressed by GTS or Revera
- Lack of, or insufficient, engagement with the business
- Lack of exit strategy.

It was critical that these risks were addressed as part of the IaaS project implementation. Since the technology is new for the NDHA, the National Library and DIA, assurance to the key stakeholders is paramount to gain trust in the technology and the ongoing preservation, protection, and maintenance of the Library's digital collections. Furthermore, moving the digital collections offsite also created a lessened sense of control over the Library's digital assets. Therefore, it was vital that the Library retain sufficient technical controls and implement business processes to mitigate the, entirely reasonable, concerns of stakeholders and maintain high standards of access. Additionally, thorough backup and recovery procedures, as well as risk management strategies are being put into practice throughout the project implementation.

Service Level Agreements must be granular and well defined, addressing all risks and concerns identified by the key stakeholders, and more importantly they must be clearly actionable with appropriate sanctions. This has proven difficult to implement. Since NDHA is still in the process of fully implementing IaaS, SLAs are not yet fully developed to address all identified risks. However, the following analysis of the implementation process and components will provide a view into how the NDHA and the IaaS project are dealing with some of the identified risks, and the processes put in place to provide assurance to key stakeholders.

Service Level Agreement

The SLA between DIA and Revera covers a wide range of services. At a minimum, all SLA should cover issues such as performance, availability, support, data security, access policies, and redundancy. Revera supplied a comprehensive service catalogue to both DIA and NDHA [9], offering the following services:

- Service establishment
- Data centre services
- Utility compute services
- Storage as a Service
- Backup/restore as a service
- Transition services
- Professional services
- Service reports

As the NDHA is not yet taking advantage of all the services offered by Revera, the current SLA only comprises the following relevant components (Table 1) [10]:

Table 1. NDHA / GTS SLA components

Services	Components
Service establishment	• Network connection establishments • Establishes contact registers, account management meetings, billing formats, governance arrangements. • Service desk accessible time and channels. • Service desk response time • Security Admin Users account creation • Communication protocols with third party vendors • Operational Procedures Manual made available
Data centre services	• Housing co-located equipments at various sites and their availability • Remote Hands and Eyes Service • Availability of all housing services and related network.
Utility compute service	• Provision of VMs, VLAN, RAM and CPU for virtualisation • Various VM service model, such as shared resource pool, dedicated resource pool, baremetal server, pay as you go VMs, • Assisting with licensing impact on VMs • Provision of other utility compute services such as snapshot, cloning, firewall, encryption, switches, internet access etc.
Storage as a service	• Utilisation of disks with varying level of performance dependent of storage tiers • Minimum allocation size • Storage availability and metrics (see Table 2) • Dedicated tape pool and provision of tape media and tape storage services
Transition services	• Services enabling NDHA to commence IaaS, such as: • Installation and co-location of servers and network switch • Provision of power and cabling • Provision of storage-as-a-service • Tape media, management and offsite vaulting service • Assisting infrastructure testing • Assist with establishing a WAN circuit at Revera data centre • Service desk and service management processes provision • Provision of transition and project management to assist with transition activities
Professional services	• Various expertise to implement and / or support infrastructure

For each type of service offering, the SLA may include one or more of the following metrics and parameters to measure service levels:

Table 2. SLA component metrics

Metrics	Description
Service Response Time (SRT)	Time taken to respond to storage incidents and request from time of first notification by NDHA / DIA
Return to Operation (RTO)	Time taken to make storage fully available to DIA / NDHA after an incident
Response time	For example, Input/Output Operations Per Second (IOPS) response time (ms) recorded at the VM interface
Service readiness lead time	Lead time to fulfill new storage requirements
Price	Price for each service components per unit

As the IaaS project is still ongoing, the following risks and requirements are yet to be addressed within the SLA (Table 3). However, it is expected these requirements will be fulfilled by the end of the IaaS project implementation.

Table 3. Risks / Requirements to be addressed

Risks / Requirements	Description
Data security, ownership, and access policies	• Maintaining full ownership of stored data, and full control over where data is stored. • Control over security arrangements relating to data access
Environmental Risks to data centre	• Earthquake and Fire proofing • Network Dependency between Library and IaaS host data centre
Disaster Recovery	• Develop thorough disaster recovery process as part of project planning and implementation
Risk Management	• Formal risk management procedure defined and developed as part of IaaS
Faults resolution	• Clear demarcation between faults and issues responsible by either GTS or Revera
Exit Strategy	• Exit strategy for cloud solution must be clearly defined in SLA
Tape library and media	• Compatibility of tape library and drives with all physical hosts, operating system, adapters, and software. • Access to tape library • Acquisition of tape media • Structure of tape library
Permanent Storage replication	• Snapshot and replication requirement for all storage tiers.

Project Implementation

The process for moving to full IaaS service comprises four specific phases which will provide a pathway forward from the Library's current onsite data centre to the Revera data centre. Specific testing is carried out in each phase to ensure all objectives are met, and that no data integrity or system functionality has been compromised.

Phase 1 & 2

In Phase 1, permanent storage and the file system servers for the NDHA's User Acceptance Testing (UAT) environment was transferred to the Revera data centre (Figure 3). UAT is the Library's specific and isolated environment for testing new software releases, tools, and developments before deployment into Production. The architecture and data within the UAT environment closely mirrors that in the Production environment. Phase 2 of the project involved migration of the Production permanent storage and file system servers to the Revera data centre. This phase commenced upon successful completion of Phase 1.

Planning

In order to successfully implement IaaS, thorough planning was necessary to ensure that all procedures are well thought through and documented. The following plans were produced [11][12][13]:

Table 4. List of plans

Plans	Description
IaaS project plan	Project plan which outlines all required activities to complete Phase 1 and 2 of IaaS, including start and completion dates, and resources required.
SAM-QFS configuration design	Detailed design and description of permanent storage file system servers, network, SAM-QFS software, and backup configuration
Migration plan	Document the agreed method which would be use to migrate data from existing servers to new servers at Revera data centre
Acceptance test plan	Details the tests that will be used to verify the migration was successful, and that the integrity of files had not been compromised during data migration.

These documents, particularly the Migration and Acceptance test plans, directly address some of the risks or concerns raised by key stakeholders, such as:

- Impact on service delivery and performance
- Loss of digital collection items during transfer
- Backup and restore routine at new data centre
- Lack of risk management procedure during IaaS project.

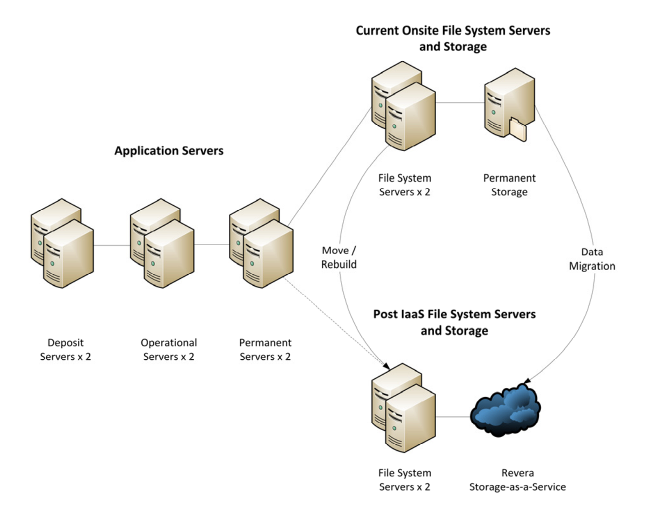

***Figure 3** High level view of migrated components*

Migration Process

This process utilised rsync and functionality of SAM-QFS to migrate data from onsite data centre to new data centre. The permanent storage and its file system were migrated as follows (Figure 4):

1. Two new file system servers were built, installed and tested at Revera data centre utilising IaaS storage tiers.
2. SAM-QFS policy put in place to create copies of files at various IaaS storage tiers including backup tapes.
3. Rsync was then used to transfer files from onsite permanent storage file system servers to those at Revera data centre.
4. As the rsync process transferred the files to new file system, SAM-QFS automatically created copies of files at various storage tiers at new data centre as per the configured policy in step 2.
5. Rsync was then used to keep the two file systems in-sync once the initial transfer completed.
6. Finally, testing was performed to ensure all files have been migrated without any corruption.

If the migration process were proved to be unsuccessful, then any issues would be resolved prior to executing the cutover activity of NFS clients. Once the file system contents were migrated and tested for integrity, NFS clients from other application servers were unmounted from existing file system servers and remounted onto the new file system servers. A reverse rsync process was initiated to ensure new files were written back to old file systems and storages as well. This step is part of the fail back procedure of the migration plan, and is the solution addressing the issue of risk management. If issues were discovered post cutover, then NFS clients would be unmounted from new storages and file systems, and remounted back to onsite file systems and storages. Through the reverse rsync process, there would be no loss of data and NDHA can resume their use of the system until issues are resolved. As soon as these steps were completed, NDHA began cutover and acceptance testing.

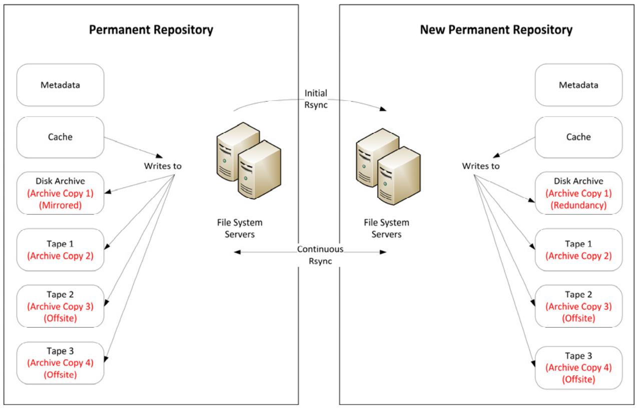

Figure 4 *IaaS data migration plan*

During Phase 2, the Production environment was migrated using similar processes, although with one file system server taken out from onsite data centre, which was then rebuilt as new file system server located at the new data centre.

Throughout the data migration and until NDHA had formally elected to decommission NDHA's use of existing onsite storages and file system servers, all current backup and restore procedure are still in place, and data sychronised at both data centres. This method addressed the risk of unsuccessful data migration and ensured key stakeholders gain confidence in migration outcome.

Testing

Various technical testing and application testing were carried out during Phase 1 and 2 IaaS implementation. The Acceptance test plan detailed all testing criteria and methods in which the tests were executed. These tests aim to:

- Verify and demonstrate the correct operation of the file systems and storages
- Demonstrate system performance
- Ensure file systems and storages are robust
- Prove success of data migration techniques
- Validate integrity of files transferred
- Verify the correct operation of backup and restore procedure

The most crucial of these tests were the data migration and integrity tests, which verified the success of data migration. These tests used two techniques:

1. SAM-QFS maintains an internal checksum for every file in permanent storage. When new copies of the files are created during data migration, SAM-QFS will again create and verify the checksum on every operation. Checksum errors are reported and monitored throughout the migration. Furthermore, SAM-QFS checksums on existing and new files systems were compared for each file, thus verifying successful data migration.
2. Fixity tests also compared checksums stored in the NDHA Rosetta database and IE METs file in the permanent storage, against the signature of the files as calculated in permanent storage. This test also verified the integrity of files migrated.

Other technical tests included:

Table 5. Technical test and outcomes

Tests	Aims	Outcomes
Storage Connectivity	Verify new file system servers could connect to disk storage at Revera data centre, and establish performance base-line.	Pass
	Verify each of the new file system servers can communicate to Revera storage successfully during scenarios of failover / shutdown, with no reduction in I/O rate or response time.	Pass
	Verify new file system servers can connect to IaaS tape storage and establish performance base-line.	Pass with issues
Integration	Verify new file system is archiving and staging content according to configured policy.	Pass

	Verify SAM-QFS cluster failover (managed and unmanaged) maintains continuous availability to both the file system and to NFS clients, without I/O failure	Pass
	Verify I/O issues with QFS volume greater than 4TB (a bug) which was fixed in latest SAM-QFS version being used in new data centre.	Pass
Backup and restore	Verify backup can be successfully performed for new file system servers, and note the throughput of different backup levels.	Pass
	Prove that the entire file-system and a disk archive can be rebuilt and measure time required for operation	Pass
Storage performance	Measure NFS performance between existing onsite application servers and new file system servers located at Revera, comparing that to the performance of all file systems and applications servers being onsite.	Pass
	Measure and compare fixity performance of test completed onsite against test completed at new data centre (NDHA fixity test is the most I/O intensive operation in NDHA)	Pass

In addition to technical testing, application testing was also carried out to ensure no application functionalities had been compromised during and after file system and storage migration to Revera data centre. Normal BAU tasks were performed to ascertain application was behaving as expected before and after cutover to the new environment. Additional testing targeting the application's use of permanent storage was completed to ensure NDHA business unit maintenance activities were not impacted. This testing included:

- Downloading files
- Exporting intellectual entities
- Processing of existing collection
- Delivery of digital collection
- Configuration and use of new storage paths
- Creation of derivative copies
- Adding representation to trigger file upload and addition of metadata in permanent storage

Timing of these activities was also captured prior to, and after cutover to Revera data centre to ensure there is no noticeable performance difference or degradation after switching to the new environment.

Issues Highlighted

In the course of verifying IaaS data migration the technical testing of newly built infrastructure highlighted a number of issues.

The NDHA fixity test that was run as part of the data migration integrity test discovered 415,936 items in which the fixity of the files did not match with the recorded value within the database. On closer examinations, these files were all IE METs created after a particular version upgrade of the Rosetta software. NDHA escalated this to Ex Libris, the vendor of the Rosetta software. It was discovered that in the Rosetta 2.1 release, Ex Libris changed the way the database checksum was generated for IE METs. Pre version 2.1, this internal checksum matched that of the actual METs file. However, post version 2.1, the recorded checksum for METs files within the database is generated from a "stripped" version of the file and no longer matches the METs file on disk. Following discussion with Ex Libris, further development will be carried out to re-store the full IE METs fixity values on this set of files.

SAM-FS checksum verify for certain storage group directories was found to be turned off on the Production environment; hence there was no checksum to compare post data migration. However, all other checksums matched and the NDHA fixity check would have covered the verification of the files within these storage group directories post data migration.

During archiving of files to tapes at the Revera data centre, a performance issue with tape archiving was discovered. Further testing indicated that the generation and verification of checksums during tape archiving was limited by the CPU speed on the file system servers. However, the current speed of tape archiving at the Revera centre is not slower than that of the current infrastructure, and could be remediated by a future server upgrade.

Monitoring

Currently, GTS has custom scripts on the file system servers to receive notifications of issues and potential issues. Monitoring at the Revera data centre is still under negotiation.

Engagement

Resourcing and overall management of this project raised some interesting issues and challenges for the Library.

Before the integration with DIA, the Library had its own technical services teams responsible for managing all aspects of the Library's infrastructure, systems and storage. As part of the integration this resource was relocated into a central technology resource pool within DIA. Whilst a lot of effort has been put into maintaining a close relationship with the now Government Technology Services responsible for managing the NDHA systems, some familiarity with their activities has been lost.

In depth knowledge of the storage requirements, business outcomes and digital preservation aspects of storage resides in the Library's NDHA business unit in particular. Consequently, it was the NDHA team who in the end worked closely with the storage provider and the IaaS team responsible for delivering the SAM-

QFS provision and migration service to define the storage strategy. The NDHA also undertook testing, verification of the integrity of migrated objects and the myriad of tasks required to ensure successful implementation of the NDHA IaaS project.

SLA's and the end to end workflows for managing day to day maintenance activities and resolution of issues have yet to be completed and tested with the business.

Future Project Scope

Subsequent to the successful completion of Phase 2, the Library shall enter into Phase 3 and 4 of the project, with a target completion date in 2013.

Phase 3 & 4

Phase 3 of the project involves physically moving all servers in the onsite data centre to the Revera data centre. This move will affect much of the Library's current applications and technical capability, as well as the remaining NDHA environment. The current onsite data centre carries environmental risks which pose clear danger to the Library's digital collections and technical infrastructure. Although disaster recovery is not part of phase 3, the move will assist the Library in mitigating current environmental risks.

Phase 4 presents the NDHA and DIA an opportunity to transform our infrastructure. Much of the NDHA infrastructure is nearing end-of-life and will soon need to be replaced. As the NDHA has been in operation for close to 5 years, the current infrastructure and architecture are based on older design. When the NDHA project commenced in 2008 there was a lack of genuine benchmarking to model a robust and scalable digital preservation infrastructure. This phase of IaaS implementation will be the opportune time for the NDHA to revisit system topology issues and improve its infrastructure to provide better digital preservation performance.

Beyond IaaS

At present, the NDHA has its own backup mechanism, which is also implemented at the new data centre. However, when the use of IaaS becomes business as usual, NDHA will investigate the possibility of using Backup-as-a-Service, rather than maintaining its own hardware and tapes. Disaster recovery will also be canvassed at this time, including looking at potential for storing one backup at another overseas location to mitigate environmental risks, in particular earthquakes in New Zealand.

Conclusion

Cloud computing such as Infrastructure-as-a-Service offers many benefits, including cost rationalisation, flexibility, and rapid scalability, as well as risks such as network dependency, data security, and reduced level of control. Although the National Library is still working towards full implementation of IaaS, careful requirements gathering and consultation, planning, testing, risk management, and well-defined SLAs should facilitate a smooth transition process, as well as providing key stakeholders with assurance regarding implementation processes and ongoing use of IaaS. Continual improvements should provide NDHA with opportunities to improve its services to the New Zealand public by enhancing our ability to preserve more digital collections. With the adoption of IaaS at the National Library of New Zealand, the NDHA is able to leverage current technology and organisation-wide initiatives to transform existing infrastructure and hardware architecture to address the growing technological demand on the NDHA.

References

[1] The Treasury. (2010). *The Economy of New Zealand: Overview.* Retrieved 2013, February 1 from http://www.treasury.govt.nz/economy/overview/2010/04.htm

[2] Guy, N. (2011, August 18). *Speech – The future of government ICT.* Retrieved 2013, February 1 from http://www.beehive.govt.nz/speech/speech-future-government-ict

[3] Office of the Minister of Internal Affairs. (2012). *Managing the government's adoption of cloud computing*. Retrieved 2013, February 1 from http://ict.govt.nz/library/CabPaper-cloud-computing.pdf

[4] *National Library of New Zealand (Te Puna Matauranga o Aotearoa) Act 2003*. Retrieved 2013, February 1 from http://legislation.govt.nz/act/public/2003/0019/latest/DLM191962.html

[5] National Library of New Zealand. (n.d.) *Structure of the library.* Retrieved 2013, February 1 from http://natlib.govt.nz/about-us/structure-of-the-library

[6] National Library of New Zealand. (n.d.) *Guardians Kaitiaki.* Retrieved 2013, February 1 from http://natlib.govt.nz/about-us/statutory-bodies/guardians-kaitiaki

[7] National Library of New Zealand. (n.d.) *Komiti Māori*. Retrieved 2013, February 1 from http://natlib.govt.nz/about-us/friends-and-advisors/komiti-maori

[8] National Library of New Zealand. (n.d.) *LIAC*. Retrieved 2013, February 1 from http://natlib.govt.nz/about-us/statutory-bodies/liac

[9] Revera Limited. (n.d.) *Schedule 4 – Service Catalogue*.

[10] Revera Limited, Department of Internal Affairs. (2012). *Participating Agency Agreement*.

[11] Gibbs, M. (2012). *IaaS SAM-QFS Configuration Design*.

[12] Gibbs, M. (2012). *IaaS SAM-QFS Migration Plan*.

[13] Gibbs, M. (2012). *IaaS SAM-QFS Acceptance Test Plan*.

Author Biography

Cynthia Wu is the Digital Preservation System Administrator for the National Digital Heritage Archive at the National Library of New Zealand. She received her B.Sc in Computer Science, Information Systems from the University of Auckland, and her Master of Library and Information Studies from the Victoria University of Wellington. She is currently responsible for system configuration, maintenance, and liaison with internal business and external vendor.