Developing a Digital Preservation Programme at a National Library

Steve Knight
Programme Director Preservation Research & Consultancy, National Library of New Zealand
PO Box 1467
Wellington 6140, New Zealand
+64 4 474 3142
Steve.Knight@natlib.govt.nz

ABSTRACT
This paper describes the development of a digital preservation programme at the National Library of New Zealand. It will consider the strategic background, the range of content and users to be catered to, and the application and technology environments providing the core of our digital preservation programme. Issues relating to non-functional requirements (including interoperability) will be discussed along with the nature of the change experienced in an organisation such as a national library, in particular the role of technology staff and the need for the traditional support or service ethic to be overlaid with a more proactive approach to their roles. Finally, the paper will discuss broader implications of digital preservation including the notion that our business is now a digital business and the need for a more focused, collaborative approach to digital preservation research and practice internationally.

1. INTRODUCTION
The holdings of an institution such as a national library are core to our collective sense of identity. The continuing development and exploration of that identity through our present dialogue with the past, and the decisions we make now regarding our behaviours in a digital environment, will determine the resources available for that discourse in the future. To some extent, how we deal with the social, cultural and psychological aspects of a dialogue between the past and the future will determine the success of our digital preservation programmes.

Categories and Subject Descriptors
H3.7 [Information Storage and Retrieval]: Digital Libraries – collection, dissemination, standards, systems issues, user issues.

General Terms

Keywords
Digital preservation, digital curation, digital continuity, interoperability, digital archiving, organisational impact, repositories management, standards.
“A National Library is a place where a nation nourishes its memory and exerts its imagination – where it connects with its past and invents its future.”

As a national library legislation drives our collecting activities, in particular, the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003 which provided for legal deposit of published electronic materials and also reinforced the role of the Alexander Turnbull Library in protecting the documents bequeathed to the Crown in 1918 and in managing ‘bequests, donations, and other additions to, and acquisitions for the purposes of the Alexander Turnbull Library.’

Legal deposit provisions apply to any person, group or organisation that publishes books, magazines, newsletters or any other work, for sale or free of charge, to any section of the public. It is important to note here that this entails collecting the widest range of unpublished and published primary resources in any format with a view to creating a national digital research library. This includes text, images, audio/visual, physical and online media, web pages (ie the .nz domain) and the range of formats and codecs that go to make up each of these varied digital media. Due to our current lack of understanding of the intricacies of the bulk of these formats we have decided not to prescribe specific formats or to normalise formats upon ingest.

Which means that the digital preservation programme of the National Library of New Zealand (NLNZ) needs to encompass any digital material that contributes to the ongoing development of ‘a comprehensive collection of documents relating to New Zealand and the people of New Zealand.’ And clearly the audience for the Library’s digital preservation programme is all New Zealanders - historians, researchers, genealogists, students, browsers – in fulfillment of our primary outcome ‘New Zealanders are connected with information important to all aspects of their lives.’

2. PURPOSE OF THIS PAPER

This paper will provide some insight into the National Library of New Zealand’s digital preservation programme, it’s strategic drivers, content and approach to digital materials.

The primary focus of the paper, however, will be on the idea of non-functional requirements for digital preservation, those requirements which, although not necessarily central to the development of a digital preservation system, are nonetheless central to the successful implementation of an active digital preservation programme.

Digital preservation sits within a broader organisational context and for the purposes of this paper, most importantly in the context of how we manage our technology services.

The central thesis of the paper is that there is currently a sea change going on in organisations such as national libraries and research libraries and, in fact, in any organisation which has the responsibility for managing digital materials for the long-term, whether that be 70 years for compliance with companies legislation or in perpetuity as prescribed for the National Library of New Zealand.

Technology is increasingly at the heart of what we do and how we respond to this will determine the nature and extent of the success of our digital preservation programmes. This paper will argue that it is time for a fundamental shift in how we approach digital preservation and our technology activities which reflects that our business is now a digital business (inherent here is a deliberate distinction from a technology business).

As such we need a different level of engagement between the digital preservation programme and the technology services team. We need a different level of technical infrastructure management and support. And we need a proactive stance within the digital preservation and technology teams that recognises they are equally responsible for the long-term success of the digital preservation programme and the organisation as a digital business.

The paper will note some challenges that have arisen through the work at the National Library of New Zealand and argue for an international response to resolving some of these challenges.

3. CONTENT

The following table shows some of the Library’s digital preservation activity based on its digitisation programmes.

What this does not show is the impact of special programmes, in particular the Library’s Pictures Online project. During the next 2-3 years the Library will be undergoing a substantial building refit which will require that some collections and services are cut back. As a consequence, affected staff have been temporarily repositioned on large scale digitisation projects which will deliver up to 200TB of data by June 2011.

What it also does not show is the range of collecting of primary published and unpublished digital sources (except for the whole-of-domain) web harvesting. There is a very strong argument that it is these primary digital sources that make up the most important component of the Library’s ongoing digital preservation programme. This material will increasingly comprise the core of the National Library’s offering as the national research library.

---


Figure 1 - Digital Library Preservation Activity
4. STANDARDS
To the extent that there are standards in the digital preservation community the NDHA has been designed with regard to those standards and emerging best practice for digital preservation.

Best practice is interpreted in the NDHA context to mean utilising internationally accepted standards and guidelines as they appear or are more broadly adopted.

The Library is aware that standards for many aspects of the NDHA will evolve over time. The availability (or not) of an appropriate standard or guideline was one of the decision criteria used to determine priorities for NDHA software development.

Where there was not a clear direction for standards selection, design choices adhered to the following rules:

- There must be a demonstrable rationale for the choice, the rationale must be aligned to the business objectives
- The rationale must be derived from a research based evaluation
- The broader Library requirements must be taken into account as part of the evaluation
- The evaluation may be subject to international peer review.

The Library adopted the OAIS as a reference model and adopted the following approach to compliance with the reference model:

- OAIS Concepts: The OAIS is accepted as a reference model that provides an internationally accepted terminology and conceptual framework.
- OAIS Responsibilities: The Library’s responsibilities are defined by New Zealand legislation. Within this legislation the Library has responsibility to collect, preserve, and make accessible digital collections in ways that ensure current and future access to New Zealand’s documentary heritage.
- Detailed Models: The OAIS functional model has been adopted as a skeleton for the definition of the Library’s functional requirements. The requirements are documented at a high level, and the rationale for any modification to the detailed OAIS functions has been documented.
- Preservation Perspectives: As noted above, the NDHA should be flexible enough to evolve with the evolution of best practice for digital preservation.
- Archive Interoperability. The design of the NDHA is based upon international standards and best practice for digital preservation. Further work is still required to determine applicable standards for the transmission of data between the NDHA and any other digital preservation systems.

The NDHA data model is designed to support the gathering, cataloguing, versioning, long-term preservation, security, authentication, management, dissemination and effective searching of digital objects.

Acknowledged metadata standards have been used where appropriate to support the creation of digital objects that can be preserved. For digital preservation metadata the PREMIS approach has been followed with some extensions from the Library’s internal preservation metadata activity. Similarly, METS has been used as the primary encoding standard within the Rosetta application.

5. ACCESS
The digital world is increasingly throwing up challenges to traditional models of rights including copyright, indigenous, and privacy rights. While our understanding of the ongoing nature and extent of these rights and how they are expressed socially and legislatively is likely to change in the future, material held in the NDHA is currently subject to the National Library of New Zealand Te Puna Mātauranga o Aotearoa Act 2003 and the Copyright Act 1994 and its amendments, and other restrictions/ usages agreed with donors or at acquisition.

Access provisions to legal deposit material are contained in the National Library of New Zealand Te Puna Mātauranga o Aotearoa Act 2003.

Some material may be donated to the Library under strict terms of confidentiality. Access to such material may be subject to a number of restrictions agreed with the donors. Such restrictions may relate to access methods and user authorisations.

In addition, the NDHA will hold sensitive material embargoed for an extended period (eg 50 to 100 years). Such material cannot be accessed by anyone other than designated staff.

All access or actions relating to material in the NDHA must be recorded and notified:

- Access rights and conditions of use will be held for each digital object and its related metadata
- Access rights and conditions of use will be machine readable and actionable
- New Zealand Copyright Law will be applicable to some digital objects
- International copyright law will be applicable to some digital objects
- Access conditions may be specific to a digital object
- Access conditions include Open Access, where items are freely available via internal or external delivery mechanisms
- Access conditions include access where items are only available via an internal delivery mechanism
- Access conditions include Restricted access, where access requires permission or satisfaction of some criteria; authorised user access is via an internal or a secure delivery mechanism
- Access restrictions may be time based. For example, access may be restricted for 20 years, with restrictions being reviewed after 20 years have elapsed
- Access conditions include software licensing Terms and Conditions
- Access rights may be defined by location. Some digital material may only be accessible at a terminal that is located on Library premises
- In addition to restrictions defined by location, there may be a restriction on the number of concurrent users who may access the material
- Under the conditions of the National Library of New Zealand Te Puna Mātauranga o Aotearoa Act 2003, the number of users who may simultaneously access an object may be restricted.
6. THE NDHA ENVIRONMENT

6.1 Application
National Library of New Zealand has been working with Sun Microsystems and Ex Libris Group for almost three years on the development of a software application to provide core digital preservation services. The result of that partnership, Rosetta, is now available in the market as a commercial product.

The architecture of the Rosetta software application is split into three logical, and functional, units:

- Deposit
- Staging/Operational/Delivery
- Permanent.

All data in the system uses these three logical file system components:

- Deposit Area – A temporary area to store non-submitted and newly deposited materials
- Staging Area – A working area for the management of the objects (eg access derivatives are stored and delivered from staging)
- Permanent Storage – The main storage area for storing the objects after they have been approved for long term preservation.

One of the key operating principles for the digital preservation system software was that it should be a high availability, fault tolerant, and redundant system. The same applies for the hardware configuration and environment within which digital preservation takes place.

This means that any possible single points of failure in the hardware infrastructure level should be eliminated. The system is designed to recover from failure by switching to another computer within the same logical unit.

By supporting redundancy, fault tolerance and high availability within the digital preservation hardware environment the organisation is also supporting the notion of a high quality, enterprise class digital preservation environment which requires not only a digital preservation software system (and its associated preservation strategies etc) but also the appropriately designed, built and managed technical environment within which digital preservation occurs.

We will discuss this impact of digital preservation on the broader operation and management of the technology infrastructure in the discussion of the importance of non-functional requirements later in the paper.

6.2 Storage
Generally, the system differentiates between three levels of storage:

1. High performance storage with high availability and redundancy capabilities, fast recovery, management, and monitoring capabilities. This storage system must be capable of supporting 6000 I/O on average.

It is expected that I/O peaks will be much higher (although that has not been tested to date). Speed of data transfer in Preservation is directly proportional to the I/O speed.

2. Good performance storage with redundancy capability. This storage system must be capable of supporting 2000 I/O, on average.

3. Highly reliable storage that is cost effective for large amounts of data. This storage system must be capable of supporting 2000 I/O, on average.

6.3 Hardware
The Sun StorageTek Content Infrastructure System (CIS/2) is used to implement two discrete environments for the NDHA – one for Production and one for Test, Training and Development. The CIS/2 system is a Sun customer ready solution with pre-integrated Solaris SAM-QFS hosts and storage tiers.

It should be noted in the following that a possible maximum of four archive copies has been allowed for in the configuration of the system, providing a high level of redundancy.

Initial specification of the NDHA Production CIS/2 consisted of:

- Two Sun Fire T2000 Servers
- One FC Storage parcel, customised with an extra Expansion Tray increasing the usable capacity to an estimated 10.6TB
- Two SATA Storage Parcels with an estimated combined usable capacity of 48TB.

Initial specification of the NDHA Test, Training and Development CIS/2 consisted of:

- Two Sun Fire T2000 Servers
- One FC Storage parcel (no customisation)
- One SATA Storage Parcel with an estimated combined usable capacity of 24TB.

The actual SAM configuration consisted of four main components – the archiver, the releaser, the stagere and the recycler.

The Archiver automatically copies online disk cache files to archive media. The archive media can consist of either online disk files or removable media cartridges. By default, the archiver automatically creates one archive copy of all files in a Sun StorageTek SAM file system, while up to four archive copies can be created. The default behaviour of the archiver is to run at a 10 minute interval (configurable) and archive files with an archive age of 4 minutes or greater (modification time) to any available archive media.

The NDHA project also specified data verification for archive copies. This feature checks for data corruption on any data that is copied to secondary and/or tertiary media. The data verification process performs a read-after-write verification test, and records a confirmation of data validity in the metadata properties for that

---


4 See diagram at end of paper for overall NDHA hardware architecture.
file. Data verification forces the generation and use of checksums for archiving and staging, and prevents the release of the file until all archive copies have been created and their checksums verified.

The Releaser automatically maintains the file system’s online disk cache at site specified percentage usage thresholds by freeing disk blocks occupied by eligible archived files. When file system utilization exceeds its configured high-water mark, the file system management software invokes the releaser. The high and low-water marks are set with the high=percent and low=percent file system mount options. The releaser has defined directives that control the release process. After the releaser has determined the priority of the first 10,000 candidates it selects the files with the highest priority for release.

Files have a priority value derived from their age priority and size priority, which determines the order in which they are released. By default, SAM releases the largest, oldest files first, leaving the smallest, newest files on disk. The NDHA project has specified that the most recently accessed time should determine the order of release.

The Stager restores file data to the disk cache. When a user or process requests file data that has been released from disk cache, the stager automatically copies the file data back to the online disk cache.

The Recycler clears archive volumes of expired archive copies and makes volumes available for reuse. The NDHA project has specified that the recycle should run frequently on disk pools, but never on tape pools.

Each of the CIS/2 servers have been added to the Library’s regular backup schedules to ensure the Operating System is recoverable.

7. NON-FUNCTIONAL REQUIREMENTS

As noted above I want to spend some time in this paper discussing a range of non-functional requirements that contributed to the development of the NDHA.

Firstly there are a set of non-functional requirements which we refer to as architectural qualities, ie non-functional requirements that have a fundamental impact on the architecture of the system. In all there are nine architectural qualities defined by the NDHA project.

7.1 Data Assurance

Data assurance is the primary architectural quality for the NDHA system and is expressed in terms of ensuring zero data loss. This is consistent with the Library’s responsibility and goal to preserve digital material in perpetuity.

Data assurance is about being sure of the integrity of the digital material and associated metadata that has been received for ingest and that has been ingested into the NDHA.

The Library must be able to prove any item:

- Is an accurate representation of the material the NDHA has undertaken to preserve

- Has not been modified except by authorised people using accountable, audited processes.

The Library must be able to guarantee the integrity of the data deposited into the archive, including in the event of a disaster within the data centre that results in a loss of archive data from primary storage.

The NDHA system must ensure there is no unplanned loss, nor corruption, of digital objects during or following their registration and storage within the NDHA repository.

7.2 Security

In order to be recognized as a Trusted Digital Repository, the NDHA must provide a high degree of resilience against malicious attacks.

Security is addressed by the physical architecture (eg firewalls), the services provided by the data centre and the processes managed by the Technology Services group, eg threat and event management, intrusion detection and protection, auditing, virus and spy-ware scanning, password maintenance, physical security.

The NDHA system must be capable of using an enterprise-class service for identity and access management to implement role-based security which must be implemented for all processes that access and modify archive content.

7.3 Portability

The system must be implemented in a way that facilitates the migration of software components to other hardware platforms and operating systems as required.

The development language used to build the NDHA system must provide the Library with platform independence. This includes third party applications such as application servers used as part of the solution.

7.4 Flexibility

Flexibility is concerned with the ability to produce new versions of the system in response to new requirements. The intended longevity of the system means that particular attention needs to be paid to the ability of the software to adapt to changing requirements and to evolve.

The software development organisation for the NDHA system must use best practice architectural and software engineering principles including:

- Modularity

Logically related classes must be packaged together in such a way that classes within a package are strongly coupled and packages between classes are loosely coupled, creating a modular architecture with a clear separation of concerns that facilitates reuse.

- Interfaces

Interfaces will be used between the main architectural components to ensure that changing the implementation of a service provided by a client through an interface does not impact the client of that service.
Open interfaces must be used so that the Library continues to have choices regarding selection of components supplied by other vendors.

- **Design patterns**
  Standard design patterns (e.g., MVC) should be used to facilitate reuse. The vendor’s documentation must include a description of the major design patterns that have been used and what their intent is.

- **Database abstraction**
  The software architecture must cleanly encapsulate any SQL dependencies within the code.

A decision to change the underlying storage mechanism must not require code changes outside of the database abstraction module.

- **Layering**
  A standard layered architecture must be used to assist in the clean separation of concerns between components.

### 7.5 Manageability and Maintainability

Manageability is the ability to monitor and alter the system’s operational behaviour, with:

- **Alarms for faults or impending faults** – high temperature, high sustained CPU, low memory;
- **Ability to use parameters to configure the system for improved performance**;
- **Ability to produce reports to facilitate operational troubleshooting and planning**.

Maintainability is the ability to fix problems in the existing system or to upgrade parts of the system independently (e.g., add a new service pack to the operating system).

Configurable aspects of the system should be externalized in a fashion that readily allows an authorised system administrator to modify them – e.g., in an application properties file. The mechanism for configuring applications and the effects of changing configurable parameters must be clearly documented and it must be possible for a trained operator to perform the installation without assistance from the vendor.

The system must provide a mechanism that allows external system processes (e.g., Unix scripts) to be associated with defined events (e.g., change of state of an object) and automatically invoked when the event occurs.

The system must surface interfaces to allow external processes to update the attribute values (and thus state) for instances of the SIP and Entity classes, to trigger a work flow action.

The NDHA application must interact with the Library’s Event Monitoring Service for:

- **Alert monitoring**, including application fault conditions and breaches of resource utilisation thresholds.
- **Performance monitoring** including system availability, system response times, incidents of system failure, processor utilisation and loading, storage media transfer rates, network performance and server processes (e.g., data base, middleware and application).
- **Other system events related to operation monitoring** for the NDHA system.

### 7.6 Scalability

Scalability is the ability of the system to be extended to handle increased load. Common mechanisms that the NDHA system must use include:

- **Vertical scalability** in which performance can be increased by adding CPUs to existing hosts.
- **Horizontal scalability** in which a load distribution mechanism enables additional hosts to be transparently added to the systems to handle additional load and provide redundancy.

Services between major components should be distributable – i.e., it should be possible to invoke a service from a client that is located on different host than the provider of the service.

A mechanism should be provided for distributing stateless services across nodes. The mechanism must provide the ability to add additional nodes to a service pool in a way that is transparent to clients of the service.

### 7.7 Performance

Performance characterises the speed with which a system responds to user request under load.

The system must be designed to perform within parameters that will achieve response times for key transactions, e.g., for the number and size of digital items the Library expects will be submitted to the archive and which will be accessed from the archive.

### 7.8 Availability

The availability of a system is a measure of the proportion of time the system is operational from a user’s perspective. The availability goal for the NDHA system is 99.5% during business hours, 95% outside business hours.

The NDHA software will be designed to take into account the Library’s need for a system that will provide:

- **Availability for access** by Producers, including internal Producers, and other Library roles with responsibilities for registering users, depositing materials and managing the NDHA repository.
- **Availability for read-only access** to objects (i.e., for Consumers, including Library staff as well as researchers, educational users and other designated communities).

There must be no software single-point of failure that cannot be resolved either by distributing the service across multiple tiers or by locating it on a hardware cluster.

### 7.9 Disaster Recovery

Disaster Recovery refers to the process of restoring an operation after an interruption in service, including:

- Equipment repair / replacement
- File recovery / restoration
- Resumption of service to users.
The DR capabilities of the NDHA system depends on systems and processes put in place as part of the Library's business continuity plan. These decisions must take account of the NDHA system application design and infrastructure.

The NDHA software should lend itself to a rapid redeployment on new hardware located in a different data centre.

There must be no hardware dependencies on a specific host that may prevent the software being installed quickly on a different host.

References to IP addresses should be symbolic rather than absolute.

It must be possible to restore multiple nodes in parallel.

8. NON-FUNCTIONAL REQUIREMENTS 2
There are also a number of non-functional requirements relating to system interfaces. The NDHA has implemented a number of interfaces to external systems in order to operate within the Library environment. The range of external systems that the NDHA system may need to provide interfaces for includes:

Collection Management Systems
- TAPUHI (unpublished, archival primary resources)
- Integrated Library System (published primary resources), specifically acquisitions and cataloguing modules
- Meridian (manages subscriptions for electronic publications)

Resource Discovery Systems
- TAPUHI - Online Public Access Catalogue (OPAC)
- Voyager - OPAC
- Timeframes, Discover, Te Ao Hou and other specific, thematic, web based applications

Special systems and tools
- Tools / systems used by internal Producers to prepare material for submission (as a SIP) eg for sound preservation (Quadriga), web harvesting, INDIGO
- Preservation tools and utilities (for preservation actions)
- Consumer and Producer Relationship Management
- Rendering for internal Library users (eg to support Entity Management functionality)
- Other rendering tools

The Library’s Enterprise systems / services for:
- Identity and Access Management
- Security Management
- System Management
- Reporting
- Website (informing / interface)
- Notification service (eg email).

The following examples relate to specific development for collection management, alarms and system monitoring, and ingest and preservation tool interfaces.

8.1 Collection Management Systems
The collection management function is external to the NDHA. Existing collection management systems present links to resources held within the digital archive.

Collection management systems are responsible for maintaining descriptive and management information (eg relationships with publishers and donors) about objects which are held in the NDHA, while the NDHA is responsible for managing the ingest process of those objects and maintaining preservation metadata about the object.

The NDHA is also responsible for long-term preservation of the object including maintaining enough descriptive metadata about the object to allow it to be discovered, even in the absence of a resource discovery system.

Two interfaces have been developed - one that allows the NDHA to notify collection management systems of the presence of a new digital object in the archive together with its URL, and a second that allows the collection management system to notify the NDHA of updates to descriptive Metadata for an object that resides within the NDHA archive.

The latter interface is to allow NDHA’s robust data preservation features to be leveraged for presentation of descriptive data alongside the Entity itself and its Representations, and removes the onus on the collection management system to provide “zero data loss in perpetuity” grade data systems.

Not all collection management systems are capable of supporting modern, open OO interface mechanisms such as Web Services. Therefore two mechanisms are provided:
- A batch-oriented FTP mechanism for legacy systems.
- An event-oriented, standards-based mechanism for non-legacy systems.

Where the collection management system is a legacy system, an FTP based mechanism transfers information between the collection management system and the NDHA.

8.2 Alarms and System Monitoring
The NDHA application must be capable of raising alarms to a Common Services Event Monitoring Service, when it detects application fault conditions. Software components should be decoupled from the specific mechanism required by the Library for raising alarms (eg SNMP) by use of an interface to an adaptor class. Examples of relevant alarms include:
- An NDHA application detects that average user response times have increased beyond a given (configurable) threshold
- An NDHA application detects that the number of active sessions has increased beyond a given (configurable) threshold
- An NDHA application fails in an attempt to commit an item to persistent storage
- A code assertion is violated
- An application detects that a non-recoverable error condition has occurred (ie a fatal exception is thrown)
• An application runs out of memory or some other system resource.

8.3 Ingest and Preservation Tool Interfaces
The NDHA system must be able to invoke various tools and utilities to support the preservation process, including:

• Virus checker
• Fixity
• Metadata extraction
• Format identification
• Format validation.

The system must be able to specify a set of files to be processed by the tool or utility.

It may prove to be sufficient to invoke an external script that passes a directory containing the files to be processed as a parameter, together with an output direct into which the invoked tool should deposit processed files.

9. NON-FUNCTIONAL REQUIREMENTS 3
One of the key issues in deploying a new digital preservation system which is, in effect, a wholly new subsidiary within our normal business streams is integration with and enhancement of existing technology management structures.

Digital preservation creates the need for a significant increase in the capacity and capability of the organisation’s technology support resources. In order to manage this the NDHA project and the Library’s Technology Services team developed a process to cater for NDHA requirements. Key objectives of this process were to:

• Ensure NDHA software, systems and processes integrate with current NLNZ environment and services as appropriate
• Create NDHA environments (production, Development etc)
• Set up storage infrastructure
• Ensure non-functional qualities of the system are progressed including security, integrity, availability, resilience and performance.

There are a number of points where implementation of a digital preservation system butts up to or integrates with the current organisational environment and technology services. In the areas noted below are some that are the usual preserve of our technology teams such as hardware, storage, network management. The rest though are not so clearly demarcated and bring with them conversations relating to boundaries between digital preservation and technical services, the nature of digital preservation as a catalyst for change in our technology management, the change in our business overall as we move from having a technology support business unit to becoming a digital business.

This conversation is critical for national libraries and research institutions dealing with ever increasing quantities of digital materials in their collections whether legislatively mandated or not. Active digital preservation may be the catalyst for a change in how we see our business. It is clear that content management (ie digital preservation) will become a new core activity of our business in years to come.

This conversation needs to take account of the following:

Hardware
• Oversee installation of new equipment
• Deployment of a Production environment
• Deployment of a Development environment
• Deployment of a Training environment
• Deployment of a Test environment
• Allocation of sufficient space for data migration
• Develop appropriate backup/restore procedures
• Provision of audit logging environment.

Storage
• Provision of storage facilities in all environments including hardware, software and appropriate processes
• Implement Storage Archive Manager to more effectively manage data.
• Revision of the NLNZ Capacity Plan.

Network
Implementation of changes due to NDHA including:

• Installing new firewalls
• Ensuring the computer room backbone is appropriately configured
• Ensuring robustness of network to support NDHA operational business, eg deposit.

Software Upgrades
• Virus scanners
• NDHA enhancements to collection management systems
• Reporting services
• Ingest tools
• Collection management systems.

Documentation
• Updated documentation, eg Production Handover documentation and Deployment Diagrams.

Performance Testing
• Performance testing in the NLNZ environment including understanding performance requirements, analysing what performance tests will achieve the project's performance requirements, configuring performance test environment
• Develop the test scripts
• Code the test scripts
• Execute performance tests
• Analyse performance test results
• Tune the system
• Repeat performance tests.

Service Level Agreements
Negotiate SLA’s for NDHA services with Technology Services.
Data Migration
Migration of data from pre-existing systems into the digital preservation system.

Implementation
Of the solution once testing is completed.

Desktop Deployment
Client software is rolled out and training locations are configured and working

Security
Reviewing the Library’s security framework to meet requirements for the NDHA Project and to meet the security needs of the business.

Change Management
New change processes may need to be introduced.

Training
Technology Services staff will require training in operational aspects of the new systems and services.

Aligning with this shift is that the enterprise infrastructure itself needs to be considerably improved to meet the demands of digital preservation. This observation was confirmed when the Library undertook a Cobit analysis of its technological maturity (some years ago now) and compared that with its then stated directions and then again with the requirements of digital preservation as can be seen in the following diagram.5

5 Control Objectives for Information and related Technology (CobiT).
NDHA requirement, Enterprise target, Current Maturity

Figure 2 - Requirements of Digital Preservation
In this context if we start to look at the integrity of digital material and associated metadata that has been received for ingest and that has been ingested into the NDHA we can immediately see that there is a requirement for increased engagement from the organisation’s technical support team. The following gives an idea of the nature of the engagement expected from the organisational technology services team in this one particular area:

- Hardware will be subject to integrity checks prior to installation
- Digital storage media will be integrity checked on a regular basis
- Digital objects will be integrity checked on a regular basis
- Backup media will be subject to regular integrity checks
- Backup and recovery routines must ensure that all components of an AIP can be recovered to the same state
- Backup and recovery routines will include objects being restored to a test server on a periodic basis
- Backup and recovery routines will include the restore of applications to a running state on a periodic basis
- Anti-virus software protection must prevent against infection by malicious or unwanted software
- Anti-virus software protection must have specific focus on the checking of digital material during the ingest process
- Anti-virus software protection must have specific focus on virus checking during user access
- There must be a record kept of which virus patterns were used to check a digital object and its related metadata
- Should a virus be detected as a result of a new virus pattern being received, a process will be required to retrospectively identify and scan digital objects and related metadata which has been scanned against prior virus patterns.
- Where viral vulnerabilities are identified, retrospective validation of objects is required to address any time gap between the point of vulnerability and the point of discovery, and/or the point of vulnerability and the point of a fix being applied. This may apply across process, hardware, software and network components.

What are the economic models for sustaining our digital preservation activities? Recent research notes that ‘digital preservation strategies face the following challenges’:

- uncertainty about selection criteria for assessing long-term value, especially with large-scale data sets, small ‘hand-crafted’ digital collections, and the emerging genres of collective authorship on the web
- misalignment of incentives between those who are in a position to preserve and those who benefit from preservation and access
- lack of clear responsibility for digital preservation, coupled with a prevailing assumption that it is someone else’s problem
- little coordination of preservation activities across diffused stakeholder communities
- difficulty in separating preservation costs from other costs, that is, in distinguishing between the process of making things available now and making things available in the future
- difficulty in valuing or monetizing the costs and benefits of digital preservation, which are necessary to secure funding and investment.

What is it about the current products, tools and services that we use for validating our digital preservation work practices that gives us confidence that they are doing what they should? For example, there are several tools for characterising, validating, extracting data from and managing file formats. These tools are used almost blithely in our digital preservation workflows even though we know that there are problems with them. What does this say about the authenticity and integrity of the objects within our preservation repositories?

Where do we look to for advice on standards and best practice? There is an increasing array of digital preservation projects, models and practices – OAIS, PREMIS, NARA, PLANETS, CASPAR, NDIIPP, SHAMAN, DURASPACE, HathiTrust – but how do we know what to trust?

Similarly where do we look for certification and audit of our systems, repositories, organisational capability, sustainability? While effort has been put into the development of tools such as Drambora8 and TRAC9 it is still not clear yet whether these will be effective mechanisms for monitoring our digital preservation activities. Work undertaken in 2009 and 2010 by the Center for Research Libraries10 in the US is beginning to add to our knowledge in this space.

Issues of understanding (what do we mean when we say digital preservation), economic sustainability, the quality of tools, products and services to support digital preservation and the lack

---


8 http://www.repositoryaudit.eu/

9 http://www.crl.edu/content.asp?l1=13&l2=58&l3=162&l4=91

10 http://www.crl.edu/content.asp?l1=13&l2=58&l3=181
of a cohesive, coherent approach to digital preservation to date are significant challenges yet to be addressed systematically by the digital preservation community. Yet, these are all key components in ensuring that we get as close to success as we are capable of at this moment.

11. CHALLENGES - INTEROPERABILITY

1. Interoperability
   
   a. Describe how repositories interoperability may benefit your organisation
   
   b. Describe how content sensitivity including privacy, confidentiality, security, intellectual property and other rights are affected under interoperability setting

Interoperability is currently a desired state for digital preservation systems but in the long term interoperability will be a fundamental requirement for all digital preservation activities.

What do we mean when we talk about interoperability?

- Interoperability is a property of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, without any restricted access or implementation.
- the ability of two or more systems or components to exchange information and to use the information that has been exchanged.
- Being able to accomplish end-user applications using different types of computer systems, operating systems, and application software, interconnected by different types of local and wide area networks.

Digital preservation is clearly as much a social and cultural set of problems as it is a technological one:

1. the problem of preserving digital information for the future is not only, or even primarily, a problem of fine tuning a narrow set of technical variables. It is not a clearly defined problem … rather, it is a grander problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape. It is a problem of building … the various systematic supports … that will enable us to tame the anxieties and move our cultural records naturally and confidently into the future.’

Similarly, interoperability will need to incorporate social, cultural and technological strands in order to create a genuine interoperability between digital preservation systems. It has recently been suggested that these strands should comprise at least the following:

- Process – what is the boundary between static content, representations, linkages
- Authenticity – how do we (people and machines) know ‘it’ is authentic
- Quality – how do we measure quality and does it change overtime
- Change over time – how do we create ‘dynamic interoperability’ frameworks
- Policy – how do we reconcile policies in a contemporary context and how do we handle policy drift
- Legal – how can we address issues related to legal aspects
- Preservation – how do we preserve ‘interoperability potentiality’ what do we preserve.13

As noted above interoperability of digital preservation programmes is to date still a desired outcome. It is to be hoped that the current NIST work on producing a Digital Preservation Interoperability Framework will rapidly bear fruit.14

12. CONCLUSION

This paper has provided a brief description of the digital preservation programme at the National Library of New Zealand including strategic background, the range of content and users to be catered to, and the application and technological environments providing the core of our digital preservation programme.

It has then moved on to suggest that there is a change occurring in the business of national libraries, research libraries and any institution with responsibility for long –term management of digital materials. This change manifests itself in our becoming digital businesses. Digital, not just technology, is now at the heart of our business not just a support act.

In this context the paper has canvassed a number of issues related to the non-functional requirements (including interoperability) of a digital preservation programme, arguing that these aspects of the wider programme are equally as important to the success of the programme as the digital preservation system that might be used.

In particular the paper argued for a change in how technology services are reflected in these organisations. The traditional support or service ethic needs to be overlaid with a more active approach to their roles. One way that this might be manifested is in the development of a cohort of professional heritage technologists who understand that their roles are intrinsic to the digital preservation process, that proactive storage management, backup and recovery, monitoring, disaster recovery are actually key elements in the digital preservation process.


The paper also looked at some more macro level questions that will need to be addressed at a strategic, global level. Not only do we need a comprehensive internal approach to digital preservation, we also need a more co-ordinated, global approach to digital preservation.

It has been pointed out to me that the embedding of digital preservation into our business could take heed of some of the enduring practices of other more traditional disciplines such as building management and conservation where proactive risk management comprising inventorying, monitoring, assessment, preventative action, treatment and restoration are the norm.

Only if we can achieve this, will we also increase the likelihood of achieving our primary goal, which is to provide safe passage into the future for those digital materials in our care and for which we have taken on the responsibility of long-term preservation.
Figure 3 - National Library of New Zealand - National Digital Heritage Archive Storage Architecture