



## Early learnings from the national library of New Zealand's National Digital Heritage Archive project

**Steve Knight**

Programme Director Preservation Research and Consultancy  
National Library of New Zealand  
Wellington, New Zealand

**Meeting: 146. Asia and Oceania**

---

*WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL*  
23-27 August 2009, Milan, Italy  
<http://www.ifla.org/annual-conference/ifla75/index.htm>

---



### **Abstract:**

*This paper follows on from the author's paper for IFLA 74 held in Quebec in August 2008 which described the strategic and technical environment in which the National Library of New Zealand (NLNZ) developed its National Digital Heritage Archive (NDHA) project and detailed some of the key components of that project.<sup>1</sup>*

*The current paper will address the launch of Phase 1 of the NDHA, the functionality that has been delivered, some organisational lessons learnt and some issues of concern regarding the overall maturity of digital preservation as a discipline.*

*In particular the paper will address:*

- *The current state of digital preservation at the National Library of New Zealand*
- *Staff engagement in the development of a digital preservation programme*
- *Some thoughts on requirements for getting started in digital preservation*
- *Some of the issues facing the broader digital preservation community.*

---

<sup>1</sup> Knight, S. 2008. From theory to practice: digital preservation at the National Library of New Zealand Te Puna Mātauranga o Aotearoa. IFLA, 2008. <http://www.ifla.org.sg/IV/ifla74/papers/084-Knight-en.pdf>. Accessed 30 May 2009.

## Digital Preservation at the National Library of New Zealand

NLNZ formally launched Phase 1 of the NDHA project in February 2009. This provides us with an end-to-end digital preservation system starting with producer management, workflow automation and delivery, but also encompassing core back end systems relating to logging, audit trails and reporting. The Library has worked for the last two years with Ex Libris Group to develop a digital preservation system (now available in the marketplace as Rosetta). Rosetta is embedded in the business and comprises the following key features.

- Producers
- Deposit 1
- Deposit 2
- Validation Stack
- IE Data Model
- SIP Submission
- SIP Processing
- Registration
- Technical Analyst
- Assessor
- Approver
- Arranger
- Set Management
- Directory Structure Transformer
- GO API
- Audit Trail
- Workflow / Process Automation
- Staff Management
- User Management
- Permanent Repository
- Delivery
- Meditor
- Reports

The following table briefly describes each of the above functions.

Function	Covers
Producers	Establish and maintain accounts for NDHA producers, with information that is pertinent for preservation of and reporting about digital content from each producer.
Deposit 1	Registration and management of individual depositor accounts and associations linking each with a Producer account.
Deposit 2	Establish and make use of (by depositors) deposit flows and forms that govern the deposit of digital content, covering web deposit, staff-mediated deposits, internal bulk deposits and external bulk deposits.
Validation Stack	Automated processes applied at time of ingest. Includes format identification, extraction of technical metadata, fixity and virus check services (integrity checking).
IE Data Model	The agreed digital domain and data models that provide the essential foundation for the digital repository and preservation system.

<b>Function</b>	<b>Covers</b>
<b>SIP Submission</b> SIP Processing Registration Technical Analyst Assessor Approver Arranger	Collectively, these functions cover the system processes and workflows that are required for ingesting new content. Covers automated ingest, 'enrichment' routines and staff intervention for addressing technical issues affecting deposited content, selection / arrangement (approval) needed prior to storage in the archive's permanent repository.
Set Management	Staff intervention to carry out maintenance activities on discrete instances of content held in the archive. Set management also supports preservation planning and actions that will be delivered in the Phase 2 system.
Directory Structure Transformer	Includes pre-transformers for converting non-standard content structures into standard content structures and transformers for converting standard content structures into SIP METS format.
GO API	Enables bypassing of the Deposit UI/Deposit Client for submission of batches of METS in a predefined format <u>directly</u> into the Deposit Server.
Audit Trail	Full audit records and attachment of provenance events affecting each digital object taken into the archive's permanent repository.
Workflow / Process Automation	Management and configuration of system workflows and automated processes.
Staff Management	Definition of roles available within the NDHA system and assignment of roles to Library staff who need to access and use the NDHA system for permitted actions.
User Management	Definition of roles available within the NDHA system and assignment of roles to external Users who need to access and use the NDHA system for permitted actions.
Permanent Repository	Final storage location of objects in the system. Provides services such as Delivery and Publishing and informs preservation risk analysis and preservation actions.
Delivery	Determining the exact/proximate line between the DPS and institutional delivery/presentation softwares and the level of integration expected.
Meditor	The suite of object and metadata editing capabilities (interface screens and flows) and the point of intersection with other DPS tasks, eg Set Management which identifies where Meditor functions are invoked.
Reports	Management, operational and statistical reporting, using views of the object data model.

## Digital Preservation and business change

With emerging technologies and user demand for direct access to material there is an echoing change in how we need to view our core operational systems. At NLNZ we are seeing a move from the hegemony of the bibliographic record to an hegemony of the digital object. Over time this will mean a shift to a three tier business model of resource discovery/delivery, collection management and content management, with the latter becoming increasingly more important.

If this trend continues this has implications for how digital preservation (ie content management) is embedded in the business. There are clear issues regarding organisational capability and capacity and, in particular, staff competencies in the digital world.

With this in mind and in order to achieve as high a level of staff buy-in as possible into the NDHA project, a business change workstream was incorporated into the project at its inception. The business change team worked with staff across the Library but also with a particular focus on those staff who would need to use the system as part of their daily work.

Engagement in the programme was varied:

- Specialist 'subject matter experts' were formally seconded to the project for up to 100% of their time for the duration of the project
- Business representatives were engaged with the project in order to leverage their specialist knowledge, eg mapping of current and design of future business processes
- Managers, Curators and Team Leaders were engaged both as managers of seconded staff but also as key operational managers able to assess the level of buy-in to the project's objectives
- A Library Review Group was also set up to ensure an overall management perspective of the business impacts of the project.

This engagement of the business throughout the project was invaluable and using a policy of 'no surprises' the business was kept apprised throughout the four years of Phase 1 of the project and were able to successfully manage their business unit representation in the project.

These groups provided input into requirements gathering/documentation and then into the design and developments phases of the project. This has led to a level of ownership and buy-in into the project that might not otherwise have eventuated.

Major business change deliverables were:

- Current and future process design
- Training and documentation design and delivery (using a 'Train-the-Trainers methodology)
- Producer management including training for some producers and usability review
- Identification, trial and implementation of performance measures
- Identification of capacity and capability issues and suggestions for their resolution.

In parallel with this activity there has been some element of structural change within the organisation to support a digital preservation programme. From an organisational

structure perspective management of the system and activation of digital preservation processes happens within the NDHA business unit. Processing of digital material, eg ingest, processing, loading to the permanent repository etc occur in the business units that have responsibility for the relevant material.

This decision to implement a digital preservation programme as an integral component of the Library's operations as opposed to setting up digital preservation as a parallel but separate activity within the Library has contributed greatly towards the organisational capability issue noted above. Other benefits include:

- Staff involvement in testing increased their expertise and their ability to champion the system back in the business
- These business experts were able to ease the broader roll-out of the system and early implementation was able to concentrate on issues such as bug identification, workflow inconsistencies, format problems, types of content we need to address rather than training, getting to grips with a new system etc.

In order to provide some indication of the impact of the business change workstream and its benefits for this paper I undertook a very brief, very arbitrary, very un-scientific and probably biased survey of the impressions of staff engagement in the NDHA project. Some representative comments from that survey follow.

Question	Answer
How did you find your participation in the NDHA project professionally?	<p>Very rewarding ... this is the most rewarding piece of work I've been involved in ... the challenge of the work and the ability to address these challenges in an environment that allows solutions to be created is most satisfying.</p> <p>I am able to learn, understand and appreciate quite a lot about various aspects of digital preservation ... which I might not have even thought about 2 years ago ... why provenance events, why the use of descriptive data to "decorate" the entities being preserved ... why the "bit stream" analysis...</p> <p>Very much enjoyed the opportunity to work with a different group of people, especially the contractors who brought a new dimension to NLNZ in terms of urgency and focus.</p>
How has it impacted on your understanding of things digital?	<p>Deepened it ... clarified my previous thinking, either in terms of reaffirming certain assumptions ... or in terms of challenging previous assumptions.</p> <p>Participation in the NDHA project has caused me to re-evaluate previously held notions and interpretations of preservation concepts. As one of the few working preservation repositories we are in the position of using community tools "in anger" and so are exposed to their deficiencies in a way that is not obvious when playing with them in an ad-hoc manner.</p>

Question	Answer
<p>How successful or otherwise has actual implementation been?</p>	<p>I must comment on my delight at the number of people ... that have fully engaged with this entire “thing” and have an understanding of the issues and care about getting it right. It’s very exciting.</p> <p>The actual implementation successfully delivered a repository application capable of providing bit preservation for deposited items. It brought to light the deficiencies that exist in community format registries, identification, and validation utilities.</p>
<p>Has the project been beneficial in articulating or resolving issues regarding how the Library deals with digital stuff?</p>	<p>It has articulated the need for a coherent and defined world view of digital content in the Library. I’m thinking here in terms of a collections policy overhaul. What stuff do we collect, why, when, etc.</p>
<p>How do you feel the project covered the end-to-end processes for digital preservation in the context of your own job?</p> <p>Was it just a matter of ‘digital preservation-ing’ current process/procedures?</p>	<p>The greatest strength of the project is that it provided a clear objective ... along with a definition of what preservation means.</p> <p>It wasn’t until we had access to the websites in the NDHA that we realised we had to refine our quality checks ... we discovered some content we thought was archived wasn’t necessarily archived after all ... so the actual end-to-end process in production has improved the quality of our archived websites.</p> <p>In some cases it wasn’t that a new process was required in addition to the old but rather that the NDHA uncovered basic flaws in operating procedures.</p> <p>Differences were handled differently – in ATL<sup>2</sup>, additional staff to cope with digital material ... extra processes ... whereas all Content Services staff added the new work to their existing analogue work, an enrichment of all the roles.</p> <p>Indirect benefits were that staff working in different units who had never met others in the same building worked together and understood the other’s way of working.</p>
<p>Has it impacted your thinking on how your career might progress?</p>	<p>It has taught me that Testers sent in ... to do a job at a client site ... have few tools available to them to do their job. I have a number of ideas as to how we could be better prepared in the future. Tomorrow’s testers will have a “toolbox” containing what they need to do their job more effectively (scripting tools, tracking tools, test management tools etc). Functional testing, as a discipline, lags far behind development in terms of tools / resources available.</p>

<sup>2</sup> Alexander Turnbull Library. At the same time Arrangement and Description staff and Curatorial staff have had aspects of digital preservation embedded in their day-to-day operational activities.

Question	Answer
Were there many changes to current processes/procedures and what was the nature of the changes?	<p>We did look at existing workflows and procedures but the basic methodologies have been extended to include digital material rather than completely new procedures developed. To a very large extent ... digital material is “just another format”.</p> <p>We have noticed that ... e-material is resource hungry; we are not “acquiring” this material as much as “harvesting” material. There is also a need for a generally higher level of skill in dealing with this material as opposed to analogue material.</p>
How did you find the training processes?	<p>Train the trainers was successful as a method. However, several of the trainers have had precious little to do as most users were trained “up front”.</p> <p>I did find that total immersion in the project did help develop expertise in the business.</p>
Any other comments?	<p>Really useful for the business to be involved in product development. I think the one thing all SMEs<sup>3</sup> added was the knowledge of the need to build a collection as opposed to acquiring material. The knowledge of the need to interact with the CMSs<sup>4</sup> and the knowledge of producer behaviour.</p> <p>Overall the NDHA project changed my life significantly – there were exciting bits, boring times and worrying moments – but it was a worthwhile and broadening experience and I’m proud of what we achieved.</p>

## Beginning a digital preservation programme

NLNZ formally launched Phase 1 of the NDHA in February 2009, with Phase 2 due in December 2009, so it is still very early to make any serious evidence-based or metrics-based claims regarding its impact or success as the cornerstone of our digital preservation programme.

However, given that qualification and that the thinking around the following notes is still far from complete, there are some comments that we can make that might be useful in guiding other institutions currently thinking of embarking on their own digital preservation programmes.

### ***Strategic drivers***

The first caveat must be, of course, that any digital preservation implementation must be predicated on the requirements of the particular institution.

This should include a clear discussion of the strategic drivers for digital preservation including:

---

<sup>3</sup> Subject Matter Experts

<sup>4</sup> Collection Management Systems

- does your organisation have a long term preservation mandate, eg Legal Deposit?
- what is the nature of your digital collections?
- what is the extent/size of your digital collections now and in the future?
- what are your institutional policy requirements for digital preservation, eg Collections Policy?
- what is the status of digital preservation within your institution?
- What is your available resourcing/staffing to implement/support digital preservation?
- what is your funding environment for digital preservation?

### ***Business models***

Business models/costs may vary from institution to institution and may significantly influence the nature of a digital preservation programme:

- does the institution have a national/regional mandate?
- is there potential for a consortial arrangement?
- is there potential for revenue generation, eg for 3rd party hosting?

### ***Defining and deciding***

When deciding to embark on a digital preservation programme it is important to understand what you want to achieve:

- it is important to get started
- it is important not to allow what you don't know to dictate your approach to digital preservation
- it is not necessary to feel that you have to do everything at once, in fact it is not possible to do everything at once.

The work at NLNZ can be arbitrarily represented as a four tiered approach to digital preservation comprising

- storage
- provenance, context, authenticity and integrity
- risk management, planning, migration, emulation etc
- futures.



Each of these is a substantial and necessary aspect of the overall digital preservation puzzle. However, Phase 1 of the NDHA project concentrated primarily on issues related to provenance, context, authenticity and integrity. Phase 2 will develop and implement our thinking regarding risk management and preservation planning.

The point here is that we have taken components of the digital preservation continuum and not attempted to implement a big bang system, which has allowed us to progress at a pace that suits our overall capability and capacity.

A corollary of this is that we need to be mindful that what we do now may not necessarily have any longevity in the context of a sustainable digital preservation programme, ie today's solutions for digital preservation undoubtedly will not be tomorrow's solutions.

### ***Deployment and implementation***

Deployment and implementation of a digital preservation programme needs to be undertaken with a view to the available funding and staffing resources. Critical project staffing includes:

- Project Manager (preferably high quality to manage overall implementation)
- Technical Lead (preferably internal resource with good knowledge of the infrastructure)
- Business Lead (preferably a champion from the business)
- Each of these should be supported by an appropriately resourced and sized team.

Deployment and implementation of a digital preservation programme also needs to take into account the materials to be preserved. Determining which materials to be preserved in the first instance should begin with:

- A resource type where the parameters of the objects are well understood, eg the results of an internal digitisation program where all the specifications have been set by the institution.
- New resource types being added depending on need, learning complexity, internal capability/capacity etc.

Note that there are other aspects of deployment and implementation that will need to be addressed as the programme is progressed including development, migration, integration etc.

### ***Ongoing staffing***

It is not yet clear what ongoing resourcing will be required for a sustainable digital preservation programme although it is likely to vary from institution to institution. NLNZ has created a new NDHA business unit comprising:

- Manager NDHA
- Preservation Policy Analyst
- Preservation Technical Analyst
- Rosetta System Administrator
- Preservation Ingest Analyst

Because of the current emphasis on Phase 2 planning and execution **we do not yet know whether this will be sufficient staffing for a digital preservation programme**. However, what has become clear is the need for a deeply technical resource to undertake research into formats, develop format management tools and integration tools with current format tools, eg JHOVE, DROID, MET.

### ***NDHA and Rosetta***

The NDHA programme has been running since 2004. For the first two years the project concentrated on developing business and functional requirements which were used to define what we thought digital preservation was. These specifications are

freely available to organisations interested in developing such documentation for their own purposes.<sup>5</sup>

Rosetta is the central software component of the NDHA, the engine of our digital preservation programme, which we have been working with Ex Libris to develop since 2007.

This equates to over four years of requirements specification, technical specification and development behind the current status of Rosetta This is a substantial amount of sunk cost which does not need to be undertaken again.

### ***Getting started***

Getting started is the key. Given the newness of digital preservation as a discipline a combination of the above approaches would allow an institution to implement at their own speed and according to the funding and human resources available.

And it would provide the time window to undertake the strategic and policy planning to support the funding and resourcing of a sustainable digital preservation programme.

### **Digital Preservation as a Dialogue with the Future<sup>6</sup>**

When we talk about digital preservation and Trusted Digital Repositories we are by default attempting to define the parameters of a conversation with the future. What are those unknown future users going to want to see from our digital preservation programmes that will give them confidence that the materials they want to use are what they purport to be or, if they are not, why they are not?

This notion that any 'preservation environment manages communication from the past while communicating with the future' has been recently canvassed by Reagan Moore in the context of a potential 'theory of digital preservation' and an articulation of the necessary requirements for sustaining this 'validation of communication from the past.'<sup>7</sup>

In order to pursue this dialogue with the future there are some things that we, as a community, need to address. These include language, products/tools and services, quality assurance, standards and best practice, and a cohesive, managed approach to the challenges ahead, both research and practical.

- 1 What do we mean when we talk about digital preservation? Where is the agreed definition of what digital preservation comprises at a granular level, ie what are the business and functional requirements for digital preservation that will provide us as practitioners and vendors as suppliers with the knowledge we need to do digital preservation?

---

<sup>5</sup> NDHA Business Requirements Specification and NDHA Functional Requirements Specification. 2005/2006. <http://ndha-wiki.natlib.govt.nz/ndha/pages/BackgroundInformation>. Accessed 26 May 2009.

<sup>6</sup> Much of the following is taken from a paper presented at Archiving 2009 and published in *Archiving 2009. Preservation strategies and imaging technologies for cultural heritage institutions and memory organizations: Final program and proceedings*.

<sup>7</sup> Moore, Reagan. 2008. *Towards a theory of digital preservation*. The International Journal of Digital Curation, V3 (1) 2008. <http://www.ijdc.net/index.php/ijdc/article/viewFile/63/42>. Accessed 31 May 2009.

- 2 What are the economic models for sustaining our digital preservation activities? Recent research notes that 'in many institutions and enterprises systemic challenges create barriers for sustainable digital access and preservation' including:
  - inadequacy of funding models to address long-term access and preservation needs
  - confusion and/or lack of alignment between stakeholders, roles, and responsibilities with respect to digital access and preservation
  - inadequate institutional, enterprise, and/or community incentives to support the collaboration needed to reinforce sustainable economic models
  - complacency that current practices are good enough
  - fear that digital access and preservation is too big to take on.<sup>8</sup>
- 3 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?
- 4 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 Drambora<sup>9</sup> and TRAC<sup>10</sup> it is still not clear yet whether these will be effective mechanisms for monitoring our digital preservation activities. Work to be undertaken in 2009 by the Center for Research Libraries<sup>11</sup> in the US should 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 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 our dialogue with the future is as well-formed as it can be.

## Conclusion

This paper has attempted to provide a brief description of the digital preservation programme at the National Library of New Zealand including where we are up to at this moment, the benefits of broad based participation within an organisation, some

---

<sup>8</sup> Blue Ribbon Task Force on Sustainable Digital preservation and Access. 2008. *Sustaining the digital investment: Issues and challenges of economically sustainable digital preservation*. [http://brtf.sdsc.edu/biblio/BRTF\\_Interim\\_Report.pdf](http://brtf.sdsc.edu/biblio/BRTF_Interim_Report.pdf).

<sup>9</sup> <http://www.repositoryaudit.eu/>

<sup>10</sup> <http://www.crl.edu/content.asp?11=13&12=58&13=162&14=91>

<sup>11</sup> <http://www.crl.edu/content.asp?11=13&12=58&13=181>

ideas on getting started and some more macro level questions that may need to be addressed at a strategic, global level.

As I noted last year 'the relevance and viability of national libraries may be determined by their ability to respond to these changing expectations of our customers. It is imperative that going forward we are clear about what services we deliver, how we deliver them, and that we resource them appropriately. Otherwise, there is a very real risk that national libraries will cease to be relevant now and into the future and that one of the key pieces of a nation's information infrastructure will not have a part to play in an increasingly globalised information market.'

There is no other issue facing our institutions at the moment which will have as deep an impact on our ongoing activities.

We need a comprehensive management approach to digital preservation which will allow us to develop the strategies to identify the risks to our digital content and the strategies to mitigate those risks. We need a more co-ordinated, global approach to digital preservation, eg centralised registries to support format management and an agreed set of risk grading criteria for formats to aid in preservation risk management and planning.

The quality of our response to the demands of digital preservation will determine to a significant extent the quality of the digital legacy that we leave for future generations of researchers, students, scholars, genealogists and all those unknown others who will be relying on our having provided safe passage for our digital materials through the years.