NATIONAL LIBRARY OF NEW ZEALAND

Te Puna Mātauranga o Aotearoa

DIGITAL LIBRARY DEVELOPMENT REVIEW

DR SEAMUS ROSS
ERPANET & HATII

FINAL REPORT
July 2003

ISBN Number: 0-477-02797-0

Table of Contents

Executive Summary

So far the National Library of New Zealand has made substantial headway in laying the intellectual foundations for its digital library services. The Library has not rushed headlong into digital library service development, but has done the fundamental intellectual work to make appropriately scaled, timely, and effective development possible. The National Library of New Zealand Act 2003 has mandated the Library to collect, preserve, and make available New Zealand's electronic documents. The Act responds to the changes in how documentary heritage is created, disseminated, and used. The diversity and complex nature of electronic objects whether websites, digital manuscripts, packaged publications (such as those on CDs), or products of in-house digitisation can only be handled by new ways of working and enhanced curatorial and technical services.

The primary focus of this review was to assess the work of the Digital Library Transition Team and to identify new lines of development (see Appendix 5 for Terms of Reference). This report concentrates on a digital library service rather than a digital library. This choice of terminology reflects the need to ensure that digital library activities are perceived as part of, and incorporated into, the context of the larger library environment. The current level of commitment to digital capacity development and resource access needs to be enhanced if the Library is to fulfil its duties under the Act. All these developments build on the leadership and success that the Library has had in this emerging area, where its team is on the forefront of digital library thinking and activity at international level. Developments are needed in five core areas: (a) creation of a dedicated digital services team, (b) formalising policies and procedures for handling digital objects, (c) establishing of long term preservation services including a digital repository, (d) mechanisms to enable the delivery and use of digital objects acquired by the Library, (e) strategies to enable professional staff in the Library to gain new and enrich existing capabilities in the digital library domain, and (f) improved communication among curators, technologists, creators of digital objects, and users to promote preservation and access.

Although few institutions have produced adequate costing models and there is no agreed set of metrics for calculating the costs of constructing a digital library it is feasible to estimate the amount of staffing effort and the scale of the technical infrastructure that is necessary initially and to estimate costs associated with progressive improvements. Whereas current developments have been funded within the National Library's baselines, the scale of services that are now needed means that new investment will be necessary if the Library is to make a suitable response to its extended mandate and to ensure its collection reflects the emerging information landscape.

The legislative mandate gives the Library the opportunity to provide the leadership in the New Zealand library community that is essential if the educational, economic, and enjoyment potential of the digital objects and the opportunities created by the National Library of New Zealand Act 2003 are to be earned.

Eight Key Recommendations

(1) The Library needs to complete and adopt its digitisation policy within this financial year.

(2) A digital library policy needs to be established by the end of the first quarter of the next financial year. This must include a sketch of the shape and character of the digital library that it is trying to develop.

(3) The Library should consolidate its digitisation activities within Corporate Services and manage these through a cross-disciplinary oversight committee to ensure that resources are effectively allocated and there is consistency to its digitisation work whether this is conducted in-house or outsourced.

(4) The Library should ensure that in developing its digital library activities it exercises its leadership role and supports public and academic libraries in the area of digital preservation.

(5) The Library should complete as a matter of urgency its Survey Objects Project and publicise the results.

(6) The Library should specify, design, implement and deploy a digital repository, if possible in conjunction with other national memory institutions, but if necessary independently of them.

(7) The Library should establish a time-constrained Digital Library Delivery Service with responsibility for strategic developments in the areas of selection, acquisition, cataloguing, providing access to and preserving digital materials.

(8) The Library improve its economic modelling of the costs associated with collecting and maintaining digital objects whether coming from digital objects on fixed media, digitisation, or web harvesting. This should include consideration of the economic impact of these resources on building the creative economy of New Zealand, encouraging tourism and services to industry.

Introduction & Background[1]

> Digital libraries promise new societal benefits, starting with the elimination of the time and space constraints of traditional bricks-and-mortar libraries. Unlike libraries that occupy buildings accessible only to those who walk through their doors, digital libraries reside on inter-networked data storage and computing systems that can be accessed by people located anywhere. At their full potential digital libraries will enable any citizen to access a considerable proportion of all human knowledge from any location. From an access vantage the Internet provides a preview of the possibilities. (Thanos, *et. al., Delos2 Proposal*, 2003).

The National Library of New Zealand Act 2003 mandates the National Library to collect digital entities.[2] If the Library is to fulfil this mandate it needs to craft adequate procedural and policy guidelines, as well as robust technical and staffing infrastructures that will enable it to discharge its new responsibilities. Access to and preservation of digital resources, whether the product of digitisation or born digital, have become crucial activities of memory institutions worldwide.[3] In developing its electronic holdings the National Library of New Zealand (NLNZ) will continue to emphasise its collection of material produced within and related to New Zealand and the Pacific. It is unlikely the Library would store large volumes of non-New Zealand material, although it does collect overseas material in certain areas, some of which may more commonly arrive in the future in digital form (e.g., on CD-ROMs). Within its remit the Library will collect websites, unpublished digital products (such as the emails of key figures in NZ life and letters and other electronically stored papers), packaged digital publications, and other types of digital objects as they emerge.

A digital library is the infrastructure, policies and procedures, and organisational, political and economic mechanisms necessary to enable access to and preservation of digital content. In some instances a digital library may be a new entity, but in most cases it will be the electronic or digital face of a traditional library and its activities will be embedded within current and evolving service structures.[4] In this sense the digital library is not an entity that can be owned by any sub-part of the Library, although at times and in response to different emerging needs some parts of the Library may have greater responsibility for it than others. There are numerous digital library experiments both within the academic sphere and within

---

[1] All URLs were verified on 29 June 2003.

[2] The National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003, http://www.natlib.govt.nz/files/Act03-19.pdf came into force in May 2003.

[3] The Spanish Resolution on Digital Preservation, July 6 2002, Document 2002/C162/02, http://www.ibeurope.com/Newslink/311.htm#5850

[4] It will be some time before we reach the stage where the infrastructure for the digital activities of a library are handled and considered in the same way we think of shelving, but this will be the ideal goal.

national, regional, and university libraries.[5] Some are services provided through many libraries, others subscription services[6], and still others are the digital resource face of traditional libraries. They all vary in character and type of content, with some being homogeneous collections on particular topics or media to others being heterogeneous entities. What is lacking though is general agreement as to what a digital library is. However, a rigid definition might restrict the imagination of developers and thereby limit the information landscapes and mechanism for accessing them that the NLNZ might want to lay down for its current and future user communities.

The Library is not the only institution in New Zealand with a legislative mandate to collect digital materials. Archives New Zealand, Te Papa, the Ministry of Culture and Heritage, and universities all have a mandate for some material. Nearly all these organisations need access to the technical, organisational, and economic expertise to fulfil their respective obligations. Its successful digital library development initiatives give the National Library the capabilities to provide leadership for New Zealand in this area and to promote widespread adoption of standards and coherence of practice. Given the complexities of developing and designing the necessary infrastructure, the similarity in the kinds of technical support required for digital library initiatives, the limited resources available in New Zealand, and the need to respond quickly to change it would seem sensible that the development of digital library activities should engage all key players both within the NLNZ and in the broader community such as the participants in the National Digital Forum. Through strengthening collaborative arrangements of this kind funding might be unlocked for research into digital library issues, in general, and preservation, in particular, that the Library, as a government sponsored organisation, cannot access.

The fifty library staff, representing 12½% of the nearly 400 staff employed by the Library, who contributed to the discussions that informed this review all recognised the need for development of digital library services. They could all see effort within the Library had been devoted to this area, but saw it in discreet units of work and noted tension between objectives, expectations, and activities. Given the complexities posed by collecting,

---

[5] Academic led projects include: The Open Video Digital Library (OVDL) http://www.open-video.org (described in full in Gary Marchionini and Gary Geisler, (2002), 'The Open Video Digital Library', *D-Lib Magazine*, 8.12, http://www.dlib.org/dlib/december02/marchionini/12marchionini.html), the Alexandria Digital Library Project, http://www.alexandria.ucsb.edu/; Variation2: Indiana University Digital Music Library Project, http://dml.indiana.edu/ . International projects include the Networked European Deposit Library (NEDLIB) project, http://www.konbib.nl/nedlib/. Among the national initiatives see the National Library of Canada Electronic Collection, http://collection.nlc-bnc.ca/e-coll-e/index-e.htm which already digitally holds more 'than 9,894 titles and more than 40,000 serial issues published by both the commercial publishing sector and the government publishing sector'. Projects such as these and the National Science Digital Library are defining the expectations of digital library services, see for example Carl Lagoze, et. al., (2002), 'Core Services in the Architecture of the National Science Digital Library NSDL', *JCDL'02*, July 13-17 2002 Portland Oregon, 201-209.

[6] For example, IEEE Computer Society Digital Library, http://www.computer.org/publications/dlib/

handling, and granting access to digital objects it is hardly surprising that a coherent approach is taking time to emerge through debate. What was impressive was the eagerness of Library staff to collaborate and cooperate on developing solutions. Some staff noted that a couple of digital library projects had yet to deliver and others had been cancelled or died. For example, the difficulties of establishing digital library services were clear to members of the Project Capital Resources Control Committee and its project register records a small number of projects which were terminated.[7] This is a good sign because this arena is very experimental and evidence such as this demonstrates the Library's commitment to getting the solution right and its ability to abandon unproductive lines of inquiry and development before they become too far advanced. The Library has built on its earlier projects by putting in place a team to do the 'hard thinking' that is necessary if it is to lay the foundation for Digital Library development; this team now needs to be expanded and its work programme enhanced. The Library needs to build its future developments on the foundation of existing professional and business strengths as well as new research. During discussions in March ownership and control of digital services was a vexed question. At least the Alexander Turnbull Library (ATL), Collection Services, and Information Services all have vested interests in how, when, and where digital library services are developed. Moreover, these three divisions house staff with key expertise essential to any digital library initiative.

The Digital Library Transition Team (DLTT) has done extremely good work (see below, *Current Work* and *Appendix 2*). There are a few perception difficulties, including lack of communication, failure of the digital library programme to delivery survey objects, and tension over ENCompass because ATL curators do not feel it delivers key services needed to manage unpublished collections and facilities currently available in TAPUHI. These create credibility gaps and need to be addressed as a matter of urgency, but doing so depends upon increased resource allocation. The available resource is just too small in comparison to the objectives it needs to be used to achieve. There is general agreement that digital library activities need to progress whether or not mistakes are made. Two key steps in the process will be to minimise the risk of loss of content and the financial risk to the institution. It is essential to recognise at the outset that as there are no off-the-shelf solutions, progress will not necessarily be smooth. This is a risky, but essential development. Mechanisms for benchmarking performance in the digital library environment need to be established. This currently poses substantial obstacles because there is a need for benchmarks before the infrastructure needed to manage the content and services it is to provide have been established.

[7] For example, the NZ Electronic Information Resources Implementation Project.

Method of Working

This review was conducted in three stages. In the first stage the full range of documents listed in Appendix 2 was analysed. This was followed by research onsite at the National Library of New Zealand, with supplementary visits to Archives New Zealand, Te Papa, the National Library of Australia, and the National Archives of Australia. During a nine-day period twenty-eight interviews were conducted with fifty-eight individuals. Most interviews were semi-structured and lasted between sixty and ninety minutes. The majority involved multiple participants. Detailed notes were taken for all these interviews (see acknowledgements). In the third stage the original documents plus additional documents collected after the interviews themselves, published and unpublished materials from other digital library development activities, and the reports of the interviews themselves were examined. This report describes the future directions that the Library might take in this area to develop its digital library capacity and services.

Basic Assumptions

The review and subsequent thinking about the next steps are founded upon some basic assumptions. Digital library developments must:

- enable NLNZ to fulfil its statutory obligations under the new Act;
- enable the NLNZ to ensure long term access to the digital resources created within and about New Zealand and the Pacific region;
- foster the growth of existing and new user communities (e.g., Mäori/Pacific);
- enable the Library to respond to the changing expectations of users, who expect remote access to resources. This is especially true of researchers and university students and particularly in regard to e-journals;
- promote research using the collections;
- facilitate international visibility and access to the scientific and cultural heritage of New Zealand;
- retain the behaviour and functionality of the digital objects collected by the Library as far as is technically feasible;
- create possibilities for reuse and repackaging of the materials it generates through its digitisation programme and as far as possible with the digitally born materials that it ingests;
- generate educational opportunities, especially in the area of life-long learning;
- make it possible for the Library to show leadership to the New Zealand library and information community and to share its expertise to facilitate regional and local digital library developments;

- be responsive to the changing needs and expectations of its user communities. This will certainly require developing user evaluation strategies;
- add value to the digital library resources that it collects and to which it provides access through such services as digital library support tools (e.g., specialised retrieval tools, multimodal interfaces) and personalisation services;
- support broad information landscape opportunities and especially online delivery of services;
- help the Library to ensure that its services promote social inclusion and break down barriers to information use by current non-users;
- be constructed using internationally agreed standards;
- enable interoperability between resources both within its collections and those held by other institutions;
- leverage the Library's existing skill and resource base and promote the professional development of its staff;
- engage all staff so that they feel that Digital Library activities are the responsibility of everyone;
- not be restricted to particular media or information formats;
- assist in generating broad public interest in the holdings of the Library; and,
- promote the Library's role as a trusted repository.

In the context of broader government objectives for the public sector ensuring that New Zealand builds a dynamic and sustainable digital information landscape is essential. Digital Library developments will:

- strengthen national identity and uphold the Treaty of Waitangi through reflecting the digital identities of New Zealanders and making that visible on the international stage;
- promote the development of an inclusive and increasingly creative economy;
- facilitate public trust in government information management practices and foster the development of digital preservation practices that will form an essential foundation to the take-up of e-government and the trust of the e-citizen;
- provide a framework to encourage and assist New Zealanders to improve their technical and information literacy skills; and,
- create fairer educational opportunities.

Development of a Digital Library focus will depend upon successful efforts to:

- create effective methods of consultation and cooperation across business units in the Library;

- integrate earlier digitisation activities, including *Timeframes* and *Papers Past*, preservation programmes including the Sound Preservation activities, and future digitisation and digital content collection and generation efforts;
- generate mechanisms to maximise the exploitation of existing in-house facilities, experience, and expertise;
- involve participation of the digital library team in the management of digitisation projects to facilitate adequate scheduling, use of standards, cost estimation accuracy, and the quality of the outputs;
- develop effective systems of programme management that can sustain long term digital library initiatives;
- establish metrics that will enable the measurement of the success of the digital library programme across five to ten year horizons, but with measurable annual outcomes;
- promote an extended community of Library professionals engaged in defining and promoting the digital library service to include curators from all divisions as well as technical staff. Currently curators see these initiatives as 'a separate thing', but they sense that for digital library services to be a success they need to be viewed as an enhancement of the Library's current services;
- demonstrate a better understanding of the place that the *Discover* project has as a rung in the digital library development initiative;
- ensure a suitable technical framework to support the Digital Library development so that the core digital library team does not need to beg and borrow its services from an already over stretched Digital Initiatives Unit (DIU)[8]; and,
- construct a mechanism to ensure that the shortcomings of ENCompass, which is currently (2003) one of the best of the digital library packages on the market, can be more effectively and efficiently improved to help the NLNZ deliver its digital library services.

The Digital Library Programme has documented its work in detail and copies of these documents were reviewed (see *Appendix 2*). This material reflects the thoughtfulness, rigour, and strategic planning that the Library has brought to its efforts to develop its digital library capacity. While it might be possible to raise questions about points of detail (e.g., persistent handles as opposed to digital object identifiers [DOI], or whether the metadata guidelines can be translated into functional tools) there are no major issues that have not been well addressed and the team shows the strengths necessary to move on to the development of a digital library infrastructure and service.

[8] The DIU, itself, is understaffed to carryout its mission to deliver new initiatives, manage the upgrades of core applications, handle vendor liaison, and workout how the ideas developed by the Digital Library Transition Team can be translated into practice or represented in systems and software.

NLNZ's need to move into the digital library arena can not be underestimated. If it is successful its contribution will help to enable New Zealand's intellectual contributions to the 21st century, its economic competitiveness, and the enjoyment of its citizens. *Digital Libraries: Universal Access to Human Knowledge* identified several 'National Challenge Transformations' as essential prerequisites if all citizens (US) were to participate in and fully benefit from the Information Age.[9] In particular, transformation was considered necessary in the *Way We*:

- Communicate,
- Deal with Information,
- Learn,
- Design and Build Things,
- Conduct Research,
- Understand the Environment,
- Work,
- Practice Health Care,
- Engage in Commerce, and
- Offer Government Services and Information.

In its report PITAC recognized that digital libraries would facilitate these transformations.

[9] President's Information Technology Advisory Committee (PITAC), (2001), *Digital Libraries: Universal Access to Human Knowledge*, http://www.ccic.gov/pubs/pitac/pitac-dl-9feb01.pdf

Current Work

The work of the Digital Library Transition Team (DLTT) has raised the profile of collection, management, and provision of access to digital assets both within the Library and in the broader digital library community at national and international levels. The DLTT has had major success in developing a viable understanding of the business process workflows for handling a variety of types of digital objects. This work has been well described (see Appendix 2). They have made progress towards defining the processes from uploading materials, to storing, authenticating, and providing access to digital objects. It is likely that these processes will be the subject of revision as they are used more regularly, but their delineation has allowed the team to do some hard thinking about what is involved. The work that has been completed in researching and defining an implementation of preservation metadata has influenced other projects on an international level. Establishing the metadata model has made it possible for its implementation to be investigated and this forms a fundamental stage in the creation of the foundation for the digital library service layer.

Work on managing digital materials, and especially efforts associated with examining the processes of ingest and the issues associated with handling packaged digital objects (CD-ROMs, tapes) has not been investigated as fully as work in the area of web harvesting. The choice of the National Library of Australia's product PANDAS (PANdora Digital Archiving System) as a mechanism to support web harvesting makes sense.[10] The web harvesting work done so far has given the Library valuable experience in testing the viability of harvesting and the approaches to carrying it out. Similar effort needs to be spent on investigating the processes of ingesting and managing packaged objects.

Of special importance is the strong work relationship within the DLTT and between the DLTT, the DIU, and Electronic Services more generally. This is supported by the fact that where the DLTT has succeeded in working with other units within the Library it has done so very successfully.

Recommendations:

The Library should complete as a matter of urgency its Survey Objects Project and publicise the results. (5)

The DLTT should work harder to communicate the results of its work within the Library to ensure that the broader curatorial team are aware of it and appreciate its value.

---

[10] Indeed, the sensibility of the choice has been confirmed by other studies such as Michael Day, (2003), *Collecting and Preserving the Web: A Feasibility Study undertaken for the JISC and Wellcome Trust*, (version 1.0, 25 February), http://library.wellcome.ac.uk/projects/archiving_feasibility.pdf

Building and Delivering Collections

> Whether or not consciously, libraries and librarians have long been prominent among the few kinds of social agencies that have preserved continuity of cultural heritages. By serving as custodians of local collections, they have incidentally served a larger common good. Whatever other public benefits they provide, publishers and Internet promoters do not, and cannot be expected to, fulfil this custodianship role. (Keller, Reich, and Herkovic, 2003).[11]

The NLNZ has done much solid preliminary work to establish an understanding of the challenges and opportunities posed by collecting digital materials. ILS lists some 5000 digital items; these can be delivered in response to user requests. Most of these items have been collected as part of more traditional collecting strategies. The new mandate should quickly lead to a radical increase in the number of items that the Library will need to accession. The quantities of materials that the Library could accession are constrained by its current service and technical infrastructure, and its staff and their professional training. These two issues could be easily addressed. The greater challenge is to define what a digital library really is in the context of the National Library and what will be collected by it.

As the Library acquires an increasing amount of content in digital form the demand for access to all its holdings electronically will increase. Users will increasingly fail to understand why entire collections can not be digitised, while at the same time the access to digital representations of analogue collections will lead to increased demands by researchers to work with the originals.

Recommendations:

The Library should establish clear acquisition targets for digital materials and create sufficient infrastructure to enable it to achieve those objectives.

The Library needs to ensure that its approaches to the protection of the intellectual property rights in material is reflected in its strategic technology developments as well as in its compliant with relevant legislation.

[11] Michael A. Keller, Victoria A. Reich, and Andrew C. Herkovic, (2003), 'What is a library anymore, anyway?', *First Monday*, 8.5 (May), http://firstmonday.org/issues/issue8_5/keller/index.html

Classes of Objects

The National Library will be ingesting digital objects in one of four main classes. Each of these classes has its own unique properties which have an impact on a collecting institution's ability to select, acquire, manage, preserve, and provide access to them. The ease with which the Library can control the process by which materials are prepared for selection and acquisition and the amount of effort that will be involved in ingesting the material into its repository will depend not only on the types of objects, but also on such factors as the number of instances of the object type, the complexity of any individual instance, the file formats, and to a lesser extent the volume of the object (for example, it may prove less labour intensive and technically challenging to ingest large digital objects of low complexity in comparison to smaller digital objects created using specialised software and having a high degree of complexity). The Library will need to develop policies and procedures for dealing with each of the following four classes of digital objects.

- Packaged Objects—includes CD-ROMs, tapes, solid state devices, and other portable media that house publications ranging from databases, e-books, games, and image collections to software. The Library is likely to have little control over how this material is submitted, although some digital library activities are working with publishers to improve the process by which digital objects are created and presented for archiving. It may be feasible to encourage some publishers of electronic resources to add metadata and functionality to ease ingest, but the diversity of organisations producing packaged digital objects and the variety of types of objects means that the Library will be confronted with an expanding, rather than narrowing, range of digital objects. Some will require specialised analysis and attention if they are to be ingested into the digital repository and even then it may only prove feasible to preserve the bit stream and not the capability to render the object, or to recreate its functionality. The Library will need to decide whether ingesting the bit stream of a packaged digital object into the Library's digital repository is sufficient or whether the original media (and any packaging) that was submitted to the Library needs to be retained as well even though it is unlikely that suitable peripheral devices (e.g., tape drives) will be available in the future to access the material.[12]

[12] Johan Steenbakkers in *The NEDLIB Guidelines - Setting up a Deposit System for Electronic Publications,* (NEDLIB Report Series, 5, Koninklijke Bibliotheek, 2000) argued that digital documents should be separated from their original carrier because the carriers were intended for publishing and not for archiving. While in digital management terms he is absolutely correct, there may be some curatorial benefits to be derived from retaining the original carriers. D. Swade, Science Museum (London), has for more than a decade promoted the later view (`Collecting Software: Preserving Information in an Object-Centred Culture', in S. Ross and E.

- Website Harvesting—the results of web harvests, mainly derived from targeted selections or nominated automated deposits (NAD), rather than from 'automated trawls of the web' will include a variety of file formats and document structures (see below, *Web Archiving*). The complexity of ingesting these objects varies with file formats, the dynamic character of the site, the level of description that needs to be manually generated, and the kind of discovery tools that the Library will wish to be able to provide alongside the objects themselves. The Library will need to consider its position on ingesting web objects that it already holds (or anticipates receiving) on paper or other media (e.g., microfilm); at least one member of NLNZ staff noted that 'the preservation copy will be the physical one.'[13] The additional functionality associated with electronic versions of analogue objects suggests that it may we be worth acquiring both and deciding at a later date, maybe after reflecting on how the material is actually used, which will be the preservation copy. The increasing use of Content Management Systems (CMS) and Digital Asset Management Systems (DAMS) for storing digital objects that institutions then dynamically deliver to the web increases the likelihood that the Library will be unable to capture web publications in the not too distant future without significant investments in new technologies and establish deposit arrangements with information creators and providers that will facilitate the collection process. The implementation of the NLNZ Act 2003, which 'requires' publishers to assist the National Librarian to store and use public documents (Part 4 Section 33 of the Act), will be facilitated by arrangements to promote effective communication and exchange of digital materials between the publishers and the Library.

- Unpublished Digital Materials—these digital objects will be mainly documents (e.g., drafts of publications, emails) of authors, politicians, and other New Zealand icons. For the most part these will in the near-term be produced with fairly standard application packages and be primarily stand-alone documents. In addition, the Library will frequently have the opportunity to discuss with potential donors how the material

---

Higgs (eds.), *Electronic Information Resources and Historians: European Perspectives*, (St Katharinen, 1993), 93-104).

[13] The NLA in its recent guidelines (Appendix 1) gave primacy to the print version, by stating that where it existed it would be selected for acquisition in preference to the digital version. *Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia,* (May 2003), http://pandora.nla.gov.au/selectionguidelines.html.

should be configured and presented for deposit.[14] It would make sense if the Library were to produce some easy to follow guidelines for potential donors on *Preparing Electronic Papers for Deposit at the ATL.* For instance, where donors can produce crucial metadata or where they can note the interrelationship between particular materials the processes of ingest and cataloguing can be enhanced and the labour required reduced. Moreover the utility of the objects themselves to future researchers would be improved.

- Outputs of National Library Digitisation Programmes—by controlling how digital objects are created, the metadata that are created along with them, and the processes by which they are delivered to the Library's Digital Library Services (DLS) the effort required to ingest these objects into the Library's Digital Repository can be contained. The streamlining of this process will depend upon the Library defining its digitisation policies, and constructing and implementing digitisation procedures.

There are other classes of digital material which the Library will eventually encounter. This material may require analysis and technical effort before it can be ingested into the holdings of the Library. To facilitate its occasional development of new auditable and verifiable procedures the Library should develop a framework to ensure that the relevant documentation is produced. This approach will maximise the reusable value of each experience and gradually generate a knowledge resource on how to handle an increasing range of object classes. A publisher of online information resources (e.g., databases) might conclude that the deposit of their material in a national library has long term preservation benefits. The agreement between Elsevier Science and Koninklijke Bibliotheek concluded in September 2002,[15] which will result in the deposit of some seven terabytes of data representing 1500 journals from Elsevier at the Dutch National Library, is a spectacular example of such an arrangement. The Library's overall strategy needs to be to develop ways to respond to a limited range of digital object classes effectively, instead of attempting to construct mechanisms to respond to all possible classes of digital objects.

The nature of the objects has an impact on the effort that is involved in bringing the objects into the Library's collections, as diagram 1 makes evident. Standards in handling digital objects have an impact on their preservation, authenticity and integrity, and how they can be delivered to users. It is widely recognised that where preservation functionality can be built into the digital objects when they are created the costs of their selection, preservation, and access can be reduced.[16] The National Library

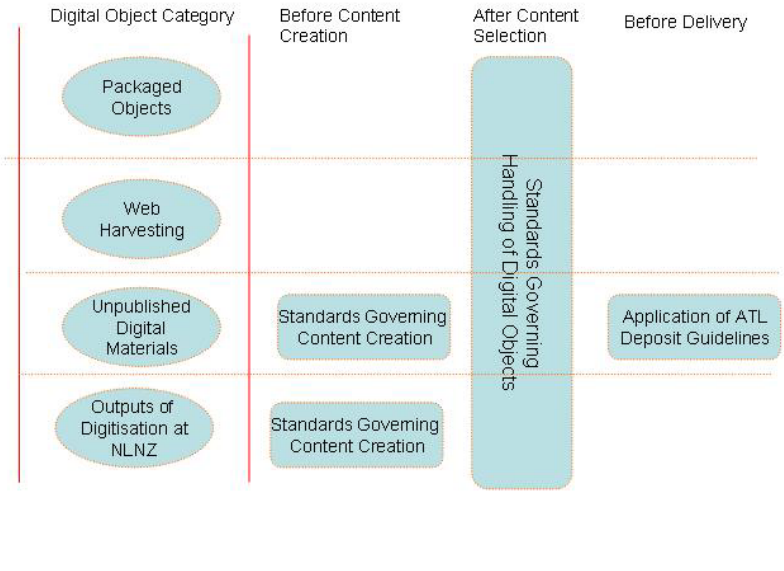
[14] Curatorial staff in the ATL noted the 'donor interview' is an essential tool in building detailed descriptions of material a donor proposes depositing.

[15] http://www.infotoday.com/newsbreaks/nb020903-2.htm

[16] This conclusion is drawn from the experience of the archives and records management community. See for example, Seamus Ross, (1999), 'Responding to the Challenges and

has little control over the construction of any of the digital objects that it will be ingesting. It will need to put most of its emphasis on standards that will enable it to manage material better once it has been acquired by the Library.

Diagram 1: Place and Role of Standards



These categories of digital objects are not likely to be homogeneous in the types of media they contain. Increasingly they will be composed of complex entities consisting of text, images, moving images, audio, and virtual reality. They will have interactive qualities. Software, both bespoke and off-the-shelf applications, will be integral to their performance. The design of interfaces for digital library users and the underlying repositories need to reflect the preponderance of the types of media held by the digital library. Increasing the amounts of time-based media, for example, has an impact on the design and optimization of the storage and delivery systems.

Web Archiving

The Library will use a number of strategies to collect digital objects including web-harvesting. Private and national institutions in an increasing number of countries are collecting material from the web. So far these projects have not adopted a consistent methodology for selection, documentation, retention, access, and disposal. Some actions, such as Kulturarw[3] (Sweden), attempt to harvest all sites within a particular

Opportunities of ICT: The New Records Manager', *Business Archives Council, Proceedings of the Annual Conference 1998* (ed. Adrian Allen), 9-25.

domain,[17] whereas others, such as Pandora (Australia), use a suite of selection criteria defined by subject specialists[18]. The Internet Archive, which has been collecting sites since the mid-1990s, has some ten billion pages and makes its holdings available through the Archive's *WayBackMachine*.[19] Although you would think a collection of this size would achieve some semblance of completeness, it does not and its coverage can be patchy.[20] Netarchive.dk, a one year project to investigate strategies of collecting and archiving Danish Internet materials, was one of the few projects to examine the process of collection and description alongside an evaluation of the research potential of the material that they had collected. This project concluded that the minimally acceptable strategy was to conduct four total harvests of the Danish domain and regular harvesting of eighty selected sites each year.[21] Ingest of the content is only the first of several challenging stages and according to the Director of the Netarchive.dk project 'it takes approximately 1 hour to make a full description of a site (30 min. for a DC-kind of description)'.[22] As the US-based Minerva Project found, harvesting websites is fraught with difficulties including problems with file formats, site boundaries, errors (e.g., missing or poorly formed links), timing, system availability, and databases.[23] Viewing the problem in a self-reflective way from the perspective of how the creating institution (e.g., the National Library of New Zealand) might preserve its own web pages is illuminating and provides an indication of some of the sorts of information, functionalities, and behaviours that may be lost when material is harvested from the web.[24] Constructing an understanding of the risks to web resources and fabricating techniques to detect them offers a complementary approach to other web object preservation strategies. Ideas investigated in Project Prism[25] at

---

[17] This statement simplifies the collecting practices because the Kultuaraw3 collects all websites ending in .se and those that end in .com, .org and .net which reside on servers based in Sweden, as well as a number of other addresses such as .nu. See http://www.kb.se/kw3/ Indeed this work showed how difficult it is to define a 'national domain space'. How will NLNZ define the 'national domain space' of New Zealand?

[18] F.A. Relf, (1999), 'PANDORA – Towards a national collection of Australian electronic publications', http://www.nla.gov.au/nla/staffpaper/ashrelf1.html. These guidelines have recently been revised (May 2003), *Online Australian Publications: Selection Guidelines for Archiving and Preservation by the National Library of Australia,* http://pandora.nla.gov.au/selectionguidelines.html.

[19] The Internet Archive and the *WayBackMachine*, http://www.archive.org/

[20] For example I have copies of pages from early versions (1995, 1996, and 1997) of the British Academy website (www.britac.ac.uk) and the HATII website (www.hatii.arts.gla.ac.uk) for 1997, 1998, and 1999 that are not available at either site or in the Internet Archive's *WayBackMachine*.

[21] The results of this project led by Birte Christensen-Dalsgaard can be found at: http://www.netarkivet.dk/rap/index-en.htm

[22] Birte Christensen-Dalsgaard, email 18/3/2003.

[23] W.Y. Arms, R. Adkins, C. Ammen, and A. Hayes, (2001), 'Collecting and Preserving the Web: The Minerva Prototype,' *RLG DigiNews*, 5.2, http://www.rlg.org/preserv/diginews/diginews5-2.html.

[24] Charles Dollar, (2001), *Archival Preservation of Smithsonian Web Resources: Strategies, Principles, and Best Practices*. (Washington, D.C.: Smithsonian Institution Archives), http://www.si.edu/archives/archives/dollar%20report.html

[25] http://www.library.cornell.edu/iris/research/prism/index.html or http://www.prism.cornell.edu.

Cornell University have been used to enable the study of this approach and using such risk pointers as content type, date, server type, content length, or date last modified.[26]

Approaches to Web Harvesting	*Countries pursuing these methods*
Selective harvest of static sites	Britain[27], Canada[28], US[29]
Selective harvest of static & dynamic sites	Australia[30]
Whole domain harvesting	Finland[31], Iceland[32], Norway, Sweden[33]
Mixed selective and whole domain harvests	France[34]

[26] see for example: Peter Botticelli, (2003), 'Risk Management for Web Resources: A Case Study on Southeast Asian Web Sites', *RLG Diginews*, 7.1, http://www.rlg.org/preserv/diginews/diginews7-1.html#feature2

[27] Currently little information appears to be available about the Britain on the Web Project. Deborah Woodyard's presentation from the *ECDL 2002* (Rome) conference *'Domain UK' Britain on the Web*, give us a useful introduction to this pilot project that selectively harvested 100 web sites (see http://bibnum.bnf.fr/ecdl/2002/uk/uk.html). There is a popular article on it at: http://www.timesonline.co.uk/article/0,,7-281852,00.html The project only considered sites where there were no problems with public access, no issues of copyright, no registration was required, and that were not government sites and not generated 'on the fly' from dynamic databases. Deborah Woodyard, kindly made available a copy of Stephen Bury's (18 March 2002) report *Domain.UK: Interim Report (March 2002)*. While several of its recommendations have a UK focus (e.g., recommendation 9 which encourages the BL to collaborate with the RDN [Resource Discovery Network], others are broadly applicable.

[28] Electronic Collections Coordinating Group, National Library Of Canada, (1998), *Networked Electronic Publications Policy and Guidelines*, http://www.nlc-bnc.ca/9/8/index-e.html

[29] see above Arms, *et. al.*, 2001. W. Y. Arms, (2001), *Web Preservation Project: final report*. (Washington, D.C.: Library of Congress), http://www.loc.gov/minerva/webpresf.pdf

[30] Warwick Cathro, Colin Webb, and Julie Whiting, (2001), *Archiving the web - The PANDORA Archive at the National Library of Australia* presented at Preserving the present for the future - Strategies for the Internet, Copenhagen (18-19 June 2001), http://www.nla.gov.au/nla/staffpaper/2001/cathro3.html

[31] Juha Hakala, (2001),'Collecting and preserving the web: Developing and testing the NEDLIB harvester,' *RLG DigiNews*, 5(2), http://www.rlg.org/preserv/diginews/diginews5-2.html#feature2

[32] Hakala (2001) reported that the Icelandic bulk harvest collected 565,169 documents from 5750 domains during January 2001. See the Nordic Web Archive, http://nwa.nb.no

[33] In his review of 'The Collection of Swedish web pages at the Royal Library - The Web Heritage of Sweden' delivered at IFLA in Glasgow, Allan Arvidson (2002) reported that 'In 1997 we harvested 6.8 million urls from 15700 web sites. The latest complete download in 2001 yielded 30 million objects from 126,000 web sites. The first download occupied 140 GBytes of data, the latest 1335 GBytes.' http://www.ifla.org/IV/ifla68/papers/111-163e.pdf He noted that the sample missed increasing amounts of material, including dynamically driven and interactive sites. See the note in *Cultivate Interactive*, 7 (2002), 'Royal Library of Sweden in the Clear' http://www.cultivate-int.org/issue7/mag-news/ A. Arvidson and F. Lettenström, (1998), 'The Kulturarw[3] project: the Swedish Royal Web Archive', *The Electronic Library*, 16.2, 105-108.

[34] Julien Masanès, (2002), 'Towards Continuous Web Archiving: First Results and an Agenda for the Future', *D-Lib Magazine*, 9.12 (December), http://www.dlib.org/dlib/december02/masanes/12masanes.html and S. Abiteboul, G. Cobena, J. Masanès, and G. Sedrati, (2002), 'A First Experience in Archiving the French Web', M. Agosti and C. Thanos (eds.), *ECDL 2002*, LNCS 2458, 1-15. ftp://ftp.inria.fr/INRIA/Projects/verso/gemo/GemoReport-229.pdf

Each of these approaches has advantages and disadvantages. In a recent review the National Library of Australia's (NLA) work in this area Margaret Philips produced a solid summary of the strengths and weaknesses of the selective harvesting approach. The labour intensiveness and high unit costs of each item were among key weaknesses she highlighted.[35] A review conducted for the United Kingdom's Joint Information Systems Committee (JISC) and the Wellcome Library by UKOLN examined the feasibility of web archiving and the various approaches currently in use.[36] In addition to its mixed selective and whole domain harvest activities the Bibliothèque Nationale de France (BNF) has also done experiments involving the acquisition of deep web content. Julien Masanès, reported at the Web Archiving workshop held just after the Roma 2002 ECDL Conference that to archive 1000 deep web sites it would take sixteen persons (five reference librarians, ten IT staff, and one coordinator). In an email response to a request for clarification he noted that this estimate was based upon extrapolation from small scale experiments, which would seem to suggest there might be some potential for streamlining the processes as they were scaled up.[37] Members of BNF staff work full-time (at least 1330 hours per year) on the initiative. It takes an average of 212 hours per deep web site to identify, contact, and manage deposit including checking the digital objects, generating the metadata, and carrying out preservation processing. The BNF team generate technical metadata (e.g., details of file formats, type data, structures, time stamps, and change history) for the web objects they acquire.

Voluntary, requested, and solicited deposit of websites/resources appear to have received limited attention, but they might be ways to address the problems associated with collecting material hidden behind firewalls, password protected sites, or dynamically drawn from databases. The information space of the 'deep web', that information landscape that is characterised by websites and associated information resources drawn from dynamic or static databases in response to specific user requests, may well be 400 or 500 times larger than the surface web.[38] As the data for the Bergman study were collected in March 2000 they are dated, but nevertheless indicative of a trend. Among his findings were that the deep web contains 7,500 terabytes of data and the surface web only 19 terabytes, the deep web contained 550 billion documents and the surface web only 1 billion, on average deep web sites 'receive fifty per cent greater monthly traffic than surface sites and are more highly linked to than surface sites',

---

[35] Margaret E Phillips, (2003), *Collecting Australian Online Publications,* Balanced Scorecard Initiative 49, http://pandora.nla.gov.au/BSC49.doc, 4.

[36] Day, 2003.

[37] Julien Masanès, Bibliothèque Nationale de France, email 24 June 2003.

[38] Michael K. Bergman, (2001), 'The Deep Web: Surfacing Hidden Value', *The Journal of Electronic Publishing*, 7.1, http://www.press.umich.edu/jep/07-01/bergman.html. Even some of the citations in Bergman's own paper (e.g., footnote 47) are now hidden behind password protected websites.

the sixty largest deep web sites were 40 times larger than the surface web, the quality content in the deep web is far greater than that in the surface web, and 95% of the deep web is publicly accessible information.[39] As a result as attractive as comprehensive web harvesting may be it is far from comprehensive because it does not reach the hidden web.[40] Moreover when the NEDLIB project examined the material returned by the NEDLIB Web Harvester's harvest of the Dutch web domain, it found that only 20% of the sites harvested were of value to the Koninklijke Bibliotheek (KB).[41] This conclusion suggests that the KB had established a relationship between value and retention.

Selective web harvesting of static and dynamic pages might provide the backbone of a mixed harvesting strategy. However, in light of the rights accorded the Library and the obligations on publishers in the new Act it is likely to be even more effective when combined with nominated automated deposit (NAD).[42] A nomination would not, necessarily, need to mean that the Library would select the material for acquisition, a decision that would ultimately be made in light of the Library's Collections Policy. Among the types of potential nominated deposit four seem to me worthy of mention:

- solicited/requested deposit—where selectors contact information owners and asked them to deposit;
- contracted deposit—where publishers or other resource creators agree to deposit material with the NLNZ on a regular basis.
- creator nomination initiated deposit—where resources and sites are proposed for ingest by the creator/owner. The creator would provide associated metadata and the necessary hooks to facilitate ingest; and,
- public nomination initiated deposit—here the general public could nominate sites they believe to be worthy of preservation. This strategy gives way to either the solicited or contracted approach as the Library needs to contact the site owners and negotiate deposit.

All four approaches can be streamlined through varying degrees of automation. In general by using NAD the NLNZ may better position itself to ensure that it acquires and is capable of ingesting cost efficiently a

---

[39] *ibid*.

[40] Though for another perspective on the issue it is worth seeing S. Raghavan and H. Garcia-Molina, (2001), 'Crawling the hidden Web', *VLDB 2001 Proceedings of 27th International Conference on Very Large Data Bases,11-14 September 2001, Roma.* http://www.dia.uniroma3.it/~vldbproc/017_129.pdf. There is research interest in developing tools to overcome some of these obstacles.

[41] Neil Beagrie, (2003), *National Digital Preservation Initiatives: An Overview of Developments in Australia, France, the Netherlands, and the United Kingdom and of Related International Activity*, (Washington, D.C., Council on Library and Information Resources and the Library of Congress).

[42] The NLNZ Act 2003 is powerful vehicle to enable the preservation of New Zealand's documentary heritage. For example see Part 4 Section 31 for an appreciation of its reach and scope.

broader range of the national scientific and cultural heritage of New Zealand as it exists in the online environment.[43] The NAD approach does not require that the Library actually ingest all material that is nominated, but at least it creates an environment were selectors will have access to and information about elements of the deep web which they might otherwise not discover or which might be technologically inaccessible to them. Some thought might be given to developing NAD tools and nomination mechanisms and tools.

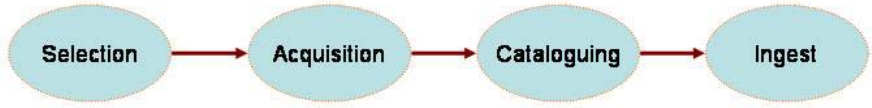
A session spent with NLNZ staff selecting websites for harvest indicated just how time consuming and intellectually demanding the process of selection is. This was the case even though the actual domain from which material was being selected was narrow—NZ government websites. The selectors at the time of the demonstration (March 2003) used cumbersome tools (i.e., spreadsheets) for tracking their work; they need access to productivity enhancing applications. Watching the process demonstrated that even in this arena additional planning needs to go into the development of practices and guidelines to formalise and clearly document how Library staff should collect material. For instance, the selectors do not collect digital representations of material that the NLNZ already holds in print. While the NLNZ wish to avoid duplication in its collections makes sense, the print and digital representations of the same material are very different entities. A user can do things with a digital version that can not be done with a print one; it seems a pity to lose that functionality. Also currently material is harvested at the very end of the collection cycle (see diagram 2) and this has resulted in the staff selecting and cataloguing some material that disappears by the time its harvesting has been requested and carried out.[44] Diagram 2 suggests a way the process be re-ordered to avoid this pitfall.

---

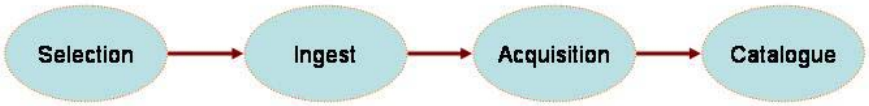
[43] Estimating the costs associated with and the viability of harvesting even the New Zealand segment of the Internet reminds one of the conundrum posed by the Walrus on seeing as he walked with the Carpenter 'along the briny beach' 'such quantities of sand'… 'If seven maids with seven mops'/Swept it for half a year/Do you suppose,' the Walrus said/'That they could get it clear?'. *The Walrus and the Carpenter*, Lewis Carroll.

[44] Brewster Kahle reported that in 1996 there were '50 million web pages with the average page online for only 75 days' (1997), 'Archiving the Internet', *Scientific American* (March), http://www.archiving.org/sciam_article.html. Peter Lyman reported that 'the average life span of a Web page is only 44 days, and 44 percent of the Web sites found in 1998 could not be found in 1999' see *Archiving the World Wide Web*, (2002), http://www.clir.org/pubs/reports/pub106/web.html

Diagram 2: Digital Content Activity Models



Current Digital Content Model



Future Digital Content Model

S Ross 2003

One scenario the National Library might consider is concentrating on collecting material created at national level only. The Danish experience with national and regional radio and television archiving has shown that regional material is little used. Of course, this argument may unravel overtime; genealogists and local historians have strong and documented interests in regional materials. Furthermore we can think of examples of historical studies that were made possible by access to regional records (e.g., Le Roy Ladurie's *The Peasants of Languedoc*). This would have the further downside that it would not necessarily give a representative record of the cultural and social diversity that makes New Zealand such a rich society.

Recommendations:

The Library should combine active selection of materials for acquisition with mechanisms to enable automated deposit of digital materials.

In the process of selecting material for acquisition selectors reject certain materials, perhaps consideration should be given to listing material not selected for acquisition.[45]

Consideration should be given to collecting material whether or not it is already held in print.

[45] PANDAS is designed to manage the metadata about titles that have been both selected and rejected for inclusion in the archive. Use of the PANDAS tools would address this recommendation.

Trusted Repositories

Whatarangi Winiata has reminded us that the Māori put tremendous value on knowledge. They were concerned about the nature and character of the repositories that held it, how it was transmitted, and to whom it was transmitted. '[A]mong the whānau, hapū and iwi were people trained to be the repositories of their knowledge.'[46] These repositories could

1. receive the information with the utmost accuracy;
2. store the information with integrity beyond doubt;
3. retrieve the information without amendment;
4. apply appropriate judgement in the use of the information; and,
5. pass on the information appropriately.

These are the expectations that we have of emerging digital repositories, and these are attributes for which we must test these repositories.[47]

The National Library of New Zealand is already a Trusted Repository—it has demonstrated that it is the place of record for New Zealand's printed heritage and that it has the capacity to maintain the National Bibliography. There is a general expectation that it will develop and continue to fulfil this role in the digital environment.[48] To do this it will need to build its capabilities through developing redundant and geographically distributed secure digital repositories. Over the past couple of decades the Library has invested heavily in developing digital records about its holdings and constructing services to provide access to these. The Library has constructed a formidable infrastructure managed by some impressive technical staff. However, the current NLNZ infrastructure is too fragmented to sustain the level of service that would be required if the Library is to put in place a viable digital library service. A key stage in laying the foundation for a digital library will be the consolidation of the current server technologies and the move of as much of the current information assets to ENCompass as possible[49]. The server and the

[46] Whatarangi Winiata, (2002), *Repositories of Rōpū Tuku Iho: A Contribution to the Survival of Māori as a People*, (Wellington: Library and Information Association of New Zealand Aotearoa Annual Conference, 17-20 November 2002), http://www.confer.co.nz/lianza2002/PDFs/Whatarangi%20Winiata.pdf

[47] I am grateful to Professor Derek Law at the University of Strathclyde for calling this to my attention and Steve Knight of the National Library of New Zealand for finding the citation.

[48] Reed-Elsevier's decision to delete some articles that it had published has prompted much debate. J. O'Donnell in a posting to *Liblicense-L* on 29 January 2003 (Subject 'Re: vanishing act' http://www.library.yale.edu/~llicense/ListArchives/0301/msg00118.html) wondered 'What guarantors other than libraries do we realistically have?'

[49] The ResetGroup report to the National Library of New Zealand on *Server Consolidation—Business Case* (February 2003) argues a very sound case for consolidation. Although this was not an area covered by the remit of this review, the conclusions of in the ResetGroup's report about the current server configuration and deployment makes sense. The tables on pages 14-16 in that report provide an exceptional and robust summary of the benefits of consolidating not only the servers but also the information assets of the Library.

information resource service consolidation will reduce management costs and should free technical support services to support the management of a digital repository. This will be essential if the Library is to contain some of the budget costs associated with putting a suitable infrastructure in place. The OAIS-standard (ISO 14721:2002), a high level reference model for supporting the design of systems to ensure long-term access to digital objects, provides a good framework, as the NLNZ has found, for defining a digital preservation environment.

The problem here is that the OAIS model is high level and the National Library will need to move to design its own trusted repository. OCLC and RLG's report on Trusted Repositories (RLG/OCLC *Attributes of a Trusted Digital Repository*[50]) provides a high level model for the design, delivery, and maintenance of a digital repository. Its strength lies in the fact that it lays down aspects of the process that need to be certified and auditable if an institution is to be said to be running a trusted digital repository. For example, they press for clear statements by repository owners on such matters as policies and assumptions (e.g., practices, environment, and security), definition of processes in place to manage fidelity checks for ingest, and metadata creation and management processes. The RLG/OCLC model has several drawbacks, among them the fact that it does not explicitly support de-accessioning of objects from the repository and recent experiments have indicated that it is extremely expensive to deploy.[51] But the key here is that whatever repository model the Library may wish to adopt the processes need to be well documented and auditable. In the definition of the attributes of the trusted repository the RLG/OCLC model has much to recommend itself, although there may be a variety of ways to implement repositories.

Repository management can be a highly complex task as every type of object must be treated differently. To alleviate this problem, a common set of operations needs to be defined to perform basic repository management functions such as depositing, copying, storing, and archiving disparate types of data. For the objects and metadata it holds, a digital repository must provide secure storage, facilitate the maintenance of integrity and authenticity, and permit the authorised destruction of items. Five primary repository functions that must be enabled at an administrative level are

[50] http://www.rlg.org/longterm/repositories.pdf

[51] Actually the OCLC Digital Archive cost some 2.4 million US$ to design, pilot, and develop the requirements and software for. The hardware and software components used cost a further 675,000 US$. (Meg Bellinger, October 2002, 'Cost and Business Models for Digital Preservation: Developing Digital Lifecycle Management Services at OCLC, http://www.dpconline.org/graphics/events/presentations/pdf/BellingerDPCForum_CostsBusinessModels.pdf). There are at least two drawbacks to the model that was adopted. First the application layer depends upon proprietary software (Oracle) and secondly they used hierarchical storage systems. The project might have better used open source products such as Postgres or MySQL and a Storage Area Network (SAN) or a Network Attached Storage (NAS) solution as these are more robust that an HSM (Hierarchical Storage Management).

ingest, retrieve, track, verify, and destroy[52] and at a user level retrieval and verification are the key services that are needed. There have been a number of projects focused on laying down the foundations for the long-term storage of digital objects. DSpace and Fedora are two models that have much to recommend themselves.

Flexible Extensible Digital Object and Repository Architecture (Fedora)[53] supports interoperability and extensibility of digital library systems and institutional repositories. To do this it handles the definition of key digital library services and their interfaces, allowing flexible interaction of existing services and their integration with new services. The architecture of the system is based on interoperable digital objects and repositories.[54] It has defined an XML schema for the Fedora digital object model and this can be mapped to METS (Metadata Transmission and Encoding Standard). In the context of the project, different services and components can communicate with each other through open interfaces, and clients can interact with them. One strength of Fedora is that it treats digital objects as a package of internally stored files with references to remote files and attaches the software to the objects, whenever appropriate, along with administrative, technical, and descriptive metadata (Dublin Core). Of course, the difficulty is that while Fedora addresses the issues associated with managing objects it really does not appear to address fundamental issues associated with the storage of these objects quite as well.

An alternative model DSpace[55] provides the framework for a digital repository designed to capture, distribute, and preserve the intellectual output of MIT staff, with the view to its adoption by, and federation with, other institutions. A DSpace repository should support the management of digital assets (e.g., publications, digital objects) to make them visible and accessible over time. The system's information model is built around the concept of 'communities', each with its own distinct information management needs. DSpace identifies two levels of preservation: bit preservation, and functional preservation. Bit preservation involves the preservation of the digital file exactly as it was deposited without any change or future alteration. The functional preservation level guarantees a 'useable' file as technology formats, media, and paradigms change. Some formats can be functionally preserved more easily than others, such as TIFF images or XML documents, which can undergo format migrations.

---

[52] The decision as to whether items should be destroyed after they have been ingested into a trusted digital repository will need to be made in light of clear rules as to what can and can not be deleted from the repository and under what conditions. Such possibilities might be planned for in the Library's Digital Library Policy which itself would need to reflect NLNZ Act 2003, which refers to the retention of material in perpetuity, and other Library policies.

[53] http://www.fedora.info An excellent discussion of Fedora can be found in 'The Mellon Fedora Project: Digital Library Architecture Meets XML and Web Services', (S. Payette and T Staples, in M. Agosti and C Thanos (eds.), *ECDL 2002*, LNCS 2458, 406-421).

[54] See *Mellon Fedora Technical Specification*, (December 2002), http://www.fedora.info/documents/master-spec-12.20.02.pdf

[55] http://dspace.org/index.html.

However, many other formats are proprietary or unsupported, and so are much more difficult to preserve in this manner. The overall process of accessing formats could be greatly eased if preservation specialists had access to a registry of file formats.[56] The Digital Library Federation (DLF) has begun work to investigate the possibilities of creating a file format registry to assist the process of digital preservation.[57] The NLNZ could benefit from being an early partner in this effort.[58]

There are other models besides.[59] An exciting model for the digital repository structure could be derived from the work that has been conducted at the National Archives of Australia (NAA) in defining its digital preservation repository for public records.[60] The functional requirements at repository level can be made quite simple where the need is to link the digital object, which might be stored as a bit stream in the file system with an XML or SQL database containing the metadata about the object. The digital repository structure proposed here is designed to minimise access to the underlying repository layer by users. Users interface with the OPAC and the digital objects that they discover are packaged on request and delivered to them via an access storage service rather than directly from the repository. (This is the digital equivalent of closed stacks with resources being delivered to the reading room for use.) Another key feature is the decision by the designers to enable the objects to be examined at ingest (e.g., checked for viruses), processed, and wrapped before being placed into the repository. This processing stage will be crucial in every digital library service and in the near term in will probably require manual as well as automated intervention.

It would be useful to test these models against the requirements set out above; to do this the Library will need to build its capabilities through developing redundant and geographically distributed secure digital repositories. The risk exposure of having no redundancy in the digital repository structure could be catastrophic. Although we know that the greatest likely cause of data loss from the repository would be human error (46% or greater) other causes of loss include hardware and software failure (36%), and natural disasters (3%).[61] As the digital holdings of the Library

---

[56] Seamus Ross and Ann Gow, (1999), *Digital Archaeology: Rescuing Neglected and Damaged Data Resources*, (London: Joint Information Systems Committee and the National Preservation Office), http://www.ukoln.ac.uk/services/elib/papers/supporting/pdf/p2.pdf.

[57] Stephen L. Abrams and David Seaman, (2003), 'Towards a global format registry', IFLA Berlin 2003, http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf.

[58] Information about the progress of the DLF discussions on a *Registry for Format Representation Information* can be found at: http://hul.harvard.edu/formatregistry/. See especially the 'Proposal for a Format Registry for Digital Library Preservation' available at the site.

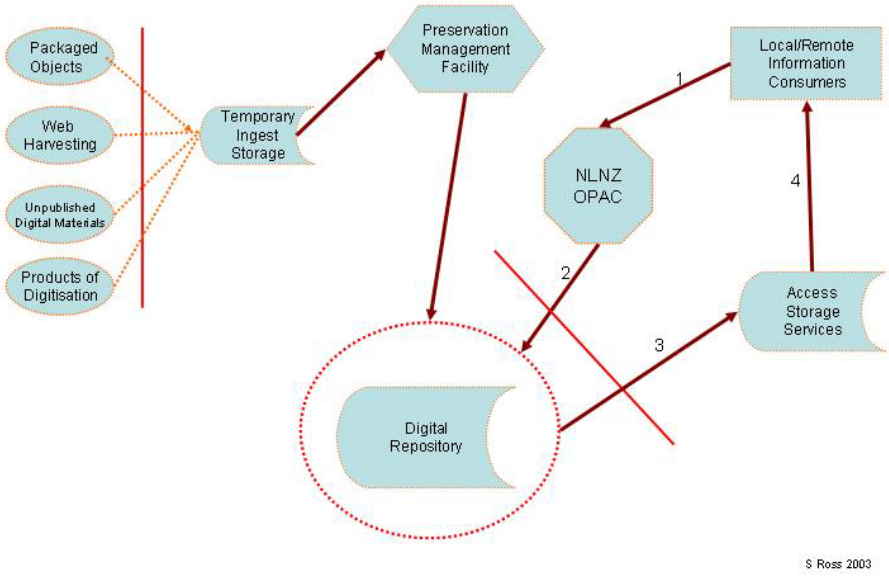
[59] The National Archives has implemented a system to store electronic records as bit streams in conjunction with metadata about structure, content, integrity, and provenance. It is currently only accessible locally. http://www.pro.gov.uk/about/preservation/digital/archive/default.htm

[60] Stephen Ellis and Andrew Wilson of the NAA described the plans during the March interview.

[61] NLNZ might wish to raise the risk associated with natural disasters as Wellington sits on multiple geological fault lines.

continue to expand and begin in their number and extent to reflect the prevalence of digital documents in society, their loss would have an increasingly catastrophic impact on the Library's core activities as well as on record of the cultural and scientific heritage of New Zealand and South Pacific. Redundancy offers the Library a way to mitigate the risk both to the heritage of New Zealand and to its core business. Some risk analysis work might be done to justify the decision to develop geographically redundant storage.

Diagram 3: Digital Repository Structure



The interoperability of the preservation layer with the Library's catalogue system needs to be a central element of any system. Of course it is true that the repository development models that are cited here are conservative and reflect traditional data management strategies. A potentially valid alternative is LOCKSS (Lots of Copies Keeps Stuff Safe).[62] LOCKSS implements a peer-to-peer network of persistent web caches, which can be configured to be self-managing with robust integrity checking and to auto-balance cache sites for trust. This too could be developed in such a way that it interfaced with the NLNZ OPAC. The difficulty with the LOCKSS model is that the actual assigning to it of trusted repository status in ways that will be certifiable and auditable may prove very difficult.

On the basis of this review it would appear that one trusted repository is needed for New Zealand. The National Library, Te Papa, and Archives New Zealand should team up to develop a shared design and shared maintenance and staffing infrastructure. If this should turn out not to be achievable the National Library must move forward independently to implement a trusted digital repository. Estimates (see below, *What will a Digital Library Programme Cost*) of the costs associated with developing a

[62] http://lockss.stanford.edu/

digital repository are unproven. The Koninklijke Bibliotheek, with 260 full-time equivalent staff and an annual budget of 36.5 million euros, collects multimedia publications but not games or software. To support its digital collecting it has invested heavily in developing practical experience and the physical infrastructure necessary to establish its digital accessioning and preservation capacity. For example, between 1998 and 2001 it invested 1.45 million euros to construct a new deposit system[63] and in 2003 anticipates spending some 1.14 million euros creating its archiving capacity for these materials.[64]

Recommendations:

The Library should move to consolidate its underlying computer-based storage systems to maximise efficiency and minimise costs.

The Library should ensure that there is a level of distributed redundancy in its systems to ensure that the loss of one location would not put its entire digital library at risk.

In specifying, designing, and implementing its digital repository the Library needs to consider using products that will interface with its existing public access catalogues.

In specifying, designing, and implementing its digital repository the Library should use an open source solution where that solution has achieved a broad user community in the Library as opposed to the techie environment.

That the Library explicitly flag at ingest those materials that it intends to preserve at bit-stream level and those that it will guarantee long term access to at functional level.

The Library should specify, design, implement and deploy a digital repository, if possible in conjunction with other national memory institutions (e.g., Te Papa and Archives New Zealand), but if necessary independently of them. (6)

[63] On the deposit system see B. Feenstra, (2000), *Standards for the Implementation of a Deposit System for Electronic Publications (DSEP),* NEDLIB Report Series 4. (The Hague: Koninklijke Biblioteek). Electronic version available at: http://www.kb.nl/coop/nedlib/homeflash.html

[64] Beagrie, 2003 (see above).

Digital Library Services

Repositories are, of course, only one element of a digital library. Access to our scientific and cultural heritage depends upon services 'to unlock cultural resources by offering personalised, highly interactive, stimulating, hybrid environments and shared spaces'.[65] Key to the success of the digital library programme is deciding the full shape of the digital library services that the library will offer. As the DELOS San Cassiano report demonstrated, digital library service development needs to be the focus of substantial co-ordinated research and development.[66] A digital library offers integrated services to information that is available for study through reading, listening, viewing, or virtual handling (i.e., haptics).[67] It provides a platform that will enable that information to be discovered, retrieved, enhanced, extended, linked, packaged, and personalised. Digital library information systems should support multiple sources of related information, rich and complex data sources, diverse information sources, multimedia resources, task-orientation, domain-oriented cross-lingual access (e.g., Māori and English), collaboration, and sustainability. Their effectiveness at delivering these service levels should be continually evaluated with the results of these evaluations fed back to designers and developers to improve the quality of the digital library itself.[68] Researchers, such as David Alsmeyer of BT Advanced Communications, have shown that '[m]aking more information available on the intranet increases library usage both by local users choosing online access in preference to using the Library in person and by remote users who previously had no practical means of access.'[69] The relationship between local and remote access to the born digital holdings of NLNZ requires attention.

---

[65] DigiCULT Report, (2002), *Technological Landscapes for Tomorrow's Cultural Economy—Unlocking the Value of the Cultural Heritage*, (Luxembourg: Office of the Official Publications of the European Communities).
http://www.digicult.info/pages/report2002/dc_fullreport_230602_screen.pdf

[66] http://delos-noe.iei.pi.cnr.it/activities/researchforum/Brainstorming/brainstorming-report.pdf

[67] The work of the University of Michigan to develop its Digital Library Extension Service system, which enables user access to the diversity of digital objects that the UofM Library has been collecting, might be worthy of further evaluation by the NLNZ.

[68] S. Choudhury, B. Hobbs, M. Lorie, and N. Flores, (2002), A Framework for Evaluating Digital Library Services, *D-Lib Magazine*, 8.7/8,
http://www.dlib.org/dlib/july02/choudhury/07choudhury.html

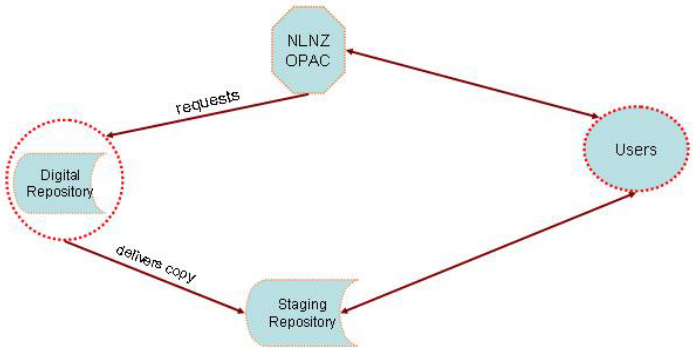
[69] David Alsmeyer (2000), *Economics and usage of a corporate digital library*,
http://www.si.umich.edu/PEAK-2000/alsmeyer.pdf. By 1999 the BT digital library provided access to some 800 online journals. Initially the digital library services had been provided to the 3500 users at the Adastral Park Research Centre and in 1994 they made up 90% of the users, by 1999 they constituted only 40%. The 1999 study of the Adastral Park users found that 1,091 online users read 9,108 journal articles 12,919 times whereas 1500 library users borrowed less than 8,000 articles in the same time period. (The online user figures were probably underestimated because it proved impossible to count use of some online publications such as those provided by a link to the ACM Digital Library). These usage figures are slightly lower than those produced by Jeffrey K. MacKie-Mason, Maria S. Bonn, Juan F. Riveros, and Wendy P. Lougee in their 1999 study 'A Report on the PEAK Experiment: Usage and Economic Behavior', *D-Lib Magazine*, 5.7/8, http://www.dlib.org/dlib/july99/mackie-mason/07mackie-mason.html.

The digital library will need to serve many communities. The Library has demonstrated that it can promote access to its holdings broadly. One of the early digital library experiments, *Papers Past*, is accessible through a bilingual, Māori and English, website. This success was made possible because the Library co-operated with the New Zealand Translation Centre (who themselves work closely with the Māori Language Commission) and it contributed to the enriching of the site as well as the language. The Library, in conjunction with LIANZA and Te Rōpū Whakahau, is already developing a database to sit alongside Te Puna to enable Māori subject headings, but the searcher will not need to know the difference.

A problem facing the NLNZ, and every other library wishing to develop digital library services, is that commercial developers have not delivered digital library management systems of sufficient richness, robustness, and sustainability to meet the needs of the potential digital library communities. It is widely observed that none of the products on the market provide the services necessary to implement a sustainable digital library service of any diversity or scale. Those requiring digital library implementation mechanisms can identify the shortcomings of the products, but do not know with certainty what an adequate product would look-like and what services it would support. Here though the Library should not lose sight of the fact that it manages both analogue and digital collections and its users will need to navigate seamlessly across both kinds of material. The Library will remain a hybrid environment where digital and analogue collections will be used side-by-side: the fully digital library like the paperless office will remain for some time to come the stuff of science fiction. Indeed some of the concepts expressed in *A Generic Resource Discovery Interface for the National Library* stress the importance of both categories of collection and the fact that as far as the user is concerned at discovery level it should not matter whether they are physical or virtual.[70] It is widely agreed that at the point of information discovery users do not need to know whether the resource held by the Library is a printed book, a painting, a manuscript, a website, an image file, or a database. For some items (which could be either an analogue or a digital item) a visit to the Library may then be required and for others online delivery might be feasible.

[70] Version 3.0, 25 February 2003.

Diagram 4: NLNZ Led User Access



S Ross 2003

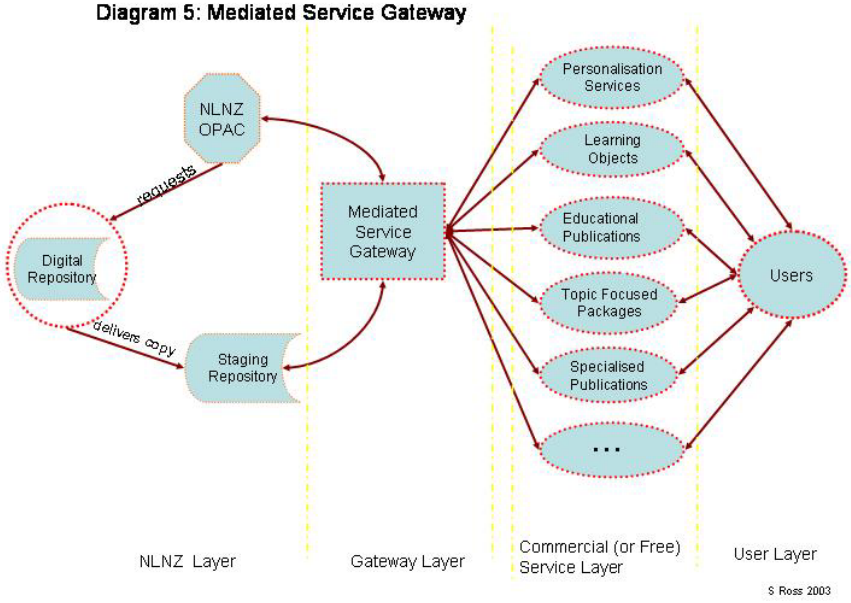
The Library could decide to provide a basic level of digital library services (Diagram 4) and permit companies to develop service levels on top (Diagram 5). In this model the Library would undertake such tasks as selection, acquisition, cataloguing, storage, preservation, and access. It would focus its attention on the first five of these six activities and provide only rudimentary access mechanisms. But through its use of standards, rich metadata, and its adherence to open description of its services the Library would enable other organisations to develop chargeable services that would take advantage of the digital holdings of the Library.[71] This avoids the need for the Library to engage as a commercial actor, ensures that researchers have a minimal level of access to the content in the digital library, but enables commercial enterprises, charities, and public bodies to develop levels of service on top of the library's infrastructure.[72] For example there is an expectation that digital libraries will provide personalisation services and this is a kind of service that commercial firms could develop and market as mechanisms for ensuring the mixed exploitation of the resources

[71] NLNZ could either charge these firms a fee for creating hooks to the content it held or allow these commercial ventures free access under the premise that by enabling these services it would reap the income as a share of the tax revenue that the government gained from the success of such emerging industries and the employment opportunities they generate.

[72] Many authors have increasing recognised that heritage institutions need to adopt more robust business models if they are deliver and sustain online digital content. Recently, Gerry Hall summed up the issue, 'For cultural and heritage institutions to successfully deliver online digital cultural content in the future, with significantly less reliance on federal government funding, they would need to adopt a more commercial approach to creating digital products and marketing those products.' Gerry Hall in (2003), 'Business model issues in the development of digital cultural content', *First Monday*, 8.5 (May), http://firstmonday.org/issues/issue8_5/wall/index.html. He is mistaken in his judgement about capacity building; cultural institutions need substantially more and continually re-newed investment in infrastructure if they are to fulfil their roles in the digital age.

that the Library was creating.[73] In another scenario, researchers might wish to create interpretation layers for use by students at primary and secondary levels. The tremendous strength of the opportunities offered by the digital library is that other organisations can assist the Library in improving access to the information assets it holds through providing service layers. The provision of services of this kind does not undermine the Library as a place of record. It places the Library at the core of the information chain. It would be feasible for service mediators to create and own metadata about digital objects in the Library's digital repository that was not owned by the Library itself. For instance, metadata that make digital objects in the NLNZ digital library usable as learning objects need not be created by the Library, but might be provided (and even owned by) researchers, educational publishers, or teachers. (See diagram 5 for an example of how this might work.) Digital library resources, services, and delivery must not be allowed to perpetuate existing barriers to access to library services, such as those felt by users with limited vision, hearing, or mobility.

Diagram 5: Mediated Service Gateway



Recommendations:

The Library needs to define what it intends to deliver through its digital library programme and establish a ten year vision for constructing its services.

---

[73] Recent research supported by the European Union's DELOS Network of Excellence and National Science Foundation in the United States has indicated how digital library services might develop to meet the needs of a diversity of user communities and the kinds of services that they could provide. DELOS/NSF Working Group on *Reference Models for Digital Libraries: Actors and Roles*, Final Report (2003), http://www.dli2.nsf.gov/internationalprojects/working_group_reports/actors_final_report.html and DELOS/NSF Working Group on *Personalisation and Recommender Systems in Digital Libraries*, (2003), http://www.dli2.nsf.gov/internationalprojects/working_group_reports/personalisation.html

The Library should continue to work closely with its Library catalogue software provider to ensure that the product develops in ways that will enable the Library to deliver its projected services.

The Library should consider what services it would wish to provide itself and those that it would permit other organisations to deliver as service layers.

That all digital library developments reflect the needs of the communities creating digital materials and the diversity of the backgrounds and needs of potential users.

The development of the Library depends upon definition of the services that the Library intends to provide and those that it could contract out either for financial reward or for free.

.

NZ Digital Library -- digitisation

The Library has remarkable collections that can be better exposed for research and education purposes through digitisation. Digitisation projects led by the Library have demonstrated that where it is possible to represent analogue objects in digital form, access and public appreciation of the heritage, including the possible ways that heritage materials can be used, are enhanced.[74] *Timeframes*, for example, provides an unparalleled way into the collections. This and other activities have demonstrated that access to even a limited range of types of materials (e.g., images of Māori on the Whanganui River or Heaphy paintings) gives potential users a glimpse of the holdings of the National Library and encourages both online and visit-led use of the collections. Those resources made available by the Library meet the needs of multiple audiences and support national and international partnerships; for instance, access to digital resources can be used to promote creativity among children in schools. While the current and completed digitisation activities, such as *Timeframes*, *Discover*, and *Picture Aotearoa* demonstrate the potential of digitisation to improve access to collections, these initiatives do not appear to be part of a coherent strategic programme to expose and make accessible the richness of New Zealand's heritage cared for by the National Library.[75] This lack of a strategic approach and policies governing digitisation means that the Library is neither building its digitisation capacity coherently nor ensuring consistency of activity across all its business units. Some digitisation work will always be done in response to public user requests and any coherent policy should reflect and support the provision of such services and the addition of materials created by it to the digital resource-base of the Library. The work so far completed, it could be argued, enabled the Library to demonstrate and better understand how it can use the technology to enhance public access to the collections. Further investment in this area should depend upon establishing a digitisation policy and implementing guidelines for digitisation within the Library.

Examination of digitisation initiatives at other institutions[76] indicates that the access to collections for educational purposes, life-long learning, and

---

[74] The Library may wish to consider the Lund Principles developed by the European Commission and the Member States, http://www.cordis.lu/ist/ka3/digicult/en/eeurope.html, as the backbone of a national reference model for digitisation.

[75] The access to digital representations of Māori material will be challenging, because the Māori would wish to see some restrictions on who can see what material.

[76] Projects such as the BNF's work to digitise thirty-five million pages at 300 ppi to make available 100,000 books to readers at the new BNF library. Technical Advisory Service for Images (TASI), http://www.tasi.ac.uk/; *The NINCH Guide to Good Practice in the Digital Representation and Management of Cultural Heritage Materials* (2002), http://www.ninch.org/guide.pdf. There is a more detailed examination of the initiatives at thirty sites considered in preparing the *NINCH Guide* to be found at: http://www.nyu.edu/its/humanities/ninchguide/interviews/; See also http://www.minervaeurope.org/ for guidelines on digitisation developed from European Projects (2003).

public enjoyment of the heritage are the strongest imperatives behind digitisation. Typical among such activities is the American Memory Project.[77] There was a feeling among some staff that the needs of the primary and secondary school environment were best served by providing better access to collections for researchers who would take advantage of the collections to create materials for the educational sector (see the service layer model above which would encourage such mediated approaches). However, those projects that have successfully carried out digitisation to enable research using particular materials tend to have done so with a particular research agenda in mind. It is difficult to point to digitisation activities where material digitised to enable research by one scholar has also facilitated research by others. This anomaly may well arise because the little scholarship that has been done using digital resources has been done using those that were created to enable particular researchers to answer specific research questions and there has been little opportunity for the resources to be reused by other researchers. The Library may wish to digitise material in anticipation of scholarship that might be conducted using the resources it has created, but the effectiveness of this strategy is difficult to predict as the research agendas of disciplines are notoriously difficult to anticipate far enough in advance to ensure that investment can be appropriately targeted for digitally reformatting collections. Recent scientific work, such as the human genome project or protein studies have demonstrated the central role of large datasets in enabling new kinds of science. In a similar vein we may find in the environmental and social sciences and the humanities that a similar transition will take place as researchers increasingly are recognising that the 'canons of resources' on which their disciplines are based are too narrow and new research demands substantially greater quantities of information. This creates a tension between the clear rewards of digitisation for educational purposes and that done to enable research. It may well be that by identifying areas where there is an overlap between educational and research needs could ensure that investment in this activity would be maximised (e.g., newspaper collections, late 19th and early 20th century photographs, passenger registers from ships, poster collections, and other ephemera).[78]

A number of improvements in practice and infrastructure might lead to the more effective application of digitisation within the Library. These include:

- the development of selection criteria, and consistent and widely used metrics for documenting how material is selected;

- the establishing of digital representation certification processes (e.g., mechanisms for demonstrating that the digital representation is a faithful copy of the analogue original) to

[77] http://memory.loc.gov/

[78] In the educational arena the American Memory Project provides a good starting point. See American Memory Learning Page: http://memory.loc.gov/ammem/ndlpedu.

provide a way to demonstration the authenticity of the digital copy;

- coherent long term digitisation planning to reduce costs and eliminate the practice of digitising the best treasures of the collection or 'end of year' digitisation to make effective use of unanticipated revenue savings;

- the definition of workflow, standards, and infrastructure developments and implementations that can be consistently adopted across the organisation;

- reduction of the number of units within the Library engaged managing the in digitisation of holdings;

- establishing and adopting standards for the creation and representation of content across the Library;

- outsourcing of digitisation activities to achieve the economies of scale where an in-house service can not deliver these;

- development of metrics to measure use and impact of the products of digitisation;

- enabling stronger communication between the content curators and staff involved in specific digitisation projects (e.g., *Discover*) about the materials that might be digitised which should lead to cross-unit collaboration and the digitisation of more complementary materials;

- the creation of a cross business unit committee to manage the digitisation activity and moving the digitisation activity to the Corporate Services unit of the Library; and

- the need to extend digitisation activities beyond documents, photographs and ephemera, to audio and video items, despite the fact that these latter items may raise further complexities, e.g., related to intellectual property rights, storage, and delivery.

A key unanswered question is whether the Library should continue to make substantial annual investments in the microfilming of holdings for preservation purposes.[79]

[79] In its examination of Cyberinfrastructure the NSF Blue-Ribbon Advisory Panel commented that 'major research libraries have switched from microfilm to digitization for both preservation and access.' National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, (2003), *Revolutionizing Science and Engineering Through Cyberinfrastructure,* (January), 42. http://www.communitytechnology.org/nsf_ci_report/report.pdf

This microfilming has limited impact on access and certainly does not reformat the analogue assets in ways that improve how users can manipulate, study, and present them. The Library needs to emphasise how new ways of working are enabled by its activities to promote access to its collections. Analysis of research conducted at the University of Yale[80] showed that digitisation from microfilm produced digital images with less fidelity, legibility, and completeness (three widely accepted measures of digital image quality) than were generated when the digital images were created by digitising from the original object. Indeed the University of Cornell[81] project showed that the most effective way to create microfilm versions of digital materials was through a combination of digitisation and computer output to microfilm (COM) using e-beam technologies.[82] It is true that digital preservation technologies have not demonstrated that they can ensure long term access to digital assets in the same way that microfilm is expected to, but as the amounts of digital material that need to be secured for preservation purposes in national libraries and archives are as yet, in all but a few exceptions, small, the incentive to ensure long term access has been limited (see *Trusted Repositories* above).

It is widely recognised that during the lifespan of a technology a particular work is rarely reformatted more than once.[83] It is, therefore, essential that, as materials that are digitised are unlikely to be digitised again, reformatting is done to the highest standards available. By bringing consistency to digitisation practices the Library will ensure that the results of digitisation have long-term asset value, can be made interoperable, can serve the needs of multiple communities, and can be assessable across time. Another result of harmonising the practice surrounding digitisation will be to reduce the costs and risks associated with ingesting the products of the process into the digital repository for long-term storage, delivery, and study. Just as we have classified the types of digital objects that the Library is likely to ingest, it would be useful to classify the types of materials that it is likely to reformat through its digitisation programmes and matched these formats against reformatting technologies (e.g., document fed scanners, digital cameras), workflow schedules, and delivery and preservation strategies.

---

[80] Paul Conway, (1996), *Conversion of Microfilm to Digital Imagery: A Demonstration Project. Performance Report on the Production Conversion Phase of Project Open Book*, Yale University Library. See also, Paul Conway, (1996),'Yale University Library's Project Open Book: Preliminary Research Findings,' *D-Lib Magazine*, February, http://www.dlib.org/dlib/february96/yale/02conway.html.

[81] A.R. Kenney, (1997), 'Digital to Microfilm Conversion: A Demonstration Project, 1994-1996, Final Report,' Cornell University Library, Ithaca, NY, www.library.cornell.edu/preservation/com/comfin.html.

[82] For a description of electronic beam imaging technologies see http://www.igraph.com/PressReleases/WP_ebeamfilm.htm. This was the technology that Cornell used for its COM work.

[83] Professor Andrew Prescott of Sheffield University has considered the use of reformatting technologies in the case of the Domesday Book and the Beowulf manuscript, for example.

Digitisation Scenario[84]

Advanced tools and improved workflow models are needed to reduce digitisation costs. This scenario constructs a comparison between microfilming and digitisation.[85] It is designed to be as much of a like-for-like comparison as possible and does not consider the use of high speed document fed scanners as a way of automate the conversion of printed material, from auto disbinding, to scanning, to auto metadata creation, to rebinding.[86] Instead this comparison uses digital cameras and should be applicable for manuscripts and other documents (up to almost A3 size). Three-short area array digital cameras, such as the Atmel, can achieve throughput of about four images a minute on consistent material (where the camera does not need to be reset between shots) and do so digitising at between 300 ppi and 400 ppi. Examples of such materials include serials and manuscripts. In practice it is possible to create about 240 scans an hour or 1440 in a six-hour day. A camera operator working 225 days per year could deliver 324,000 images. So for argument sake, let's say the operator costs NZ$62,000 per year (a base salary of NZ$50,000 plus on costs of 25%). At this throughput and staff cost relationship we could get the process down to 0.19 NZ cents per page. The problem is that the Library (or commercial organisation conducting the work) will have to make significant capital investment in equipment. Each digitisation station, consisting of a camera, camera stand, motorised table, PC with a speed tape backup device[87], chair, blackout space, and lights costs about GBP 35k or roughly NZ$115,000.

If we assumed that the total cost of the setup over three years was NZ$171,000 (consisting of initial set up, maintenance at 20% per year for two years as the first year should be included in the initial purchase price, and other consumables of about 5k per year per station [although this may not be necessary were we to use portable drives]) and the costs of one staff member to run the camera six hours per day would be NZ$195,000 (including inflation increases). The total cost of the camera station and operator over three years would be NZ$366,000. To be on the safe side, let's assume a downtime of 10% on top of our already low throughput figure; we

---

[84] This is one example scenario indicative of the scenarios the Library will build and consider if it decides to conduct a review of the relative merits of digitisation and microfilming under different circumstances.

[85] This scenario is based on the experiences of SCAN, The Scottish Archives Network, which is achieving a throughput of at least 1000 images per day per digitising station. It takes 11 seconds to setup and image a page provided the camera itself does not need to be reset, if this is necessary it takes much longer. See http://www.scan.org.uk or http://www.scottishdocuments.com.

[86] See the work of META-e and the papers about it presented at the April 2002 conference on 'New Perspectives for the Automated Digitisation of Printed Material', http://heds.herts.ac.uk/METAe/Articles/art05_1.htm

[87] This is essential because the quantity of data means that it is not feasible to move the data across the network without slowing the network down dramatically. The alternative approach, if you were to operate this system 24/7, would be to use hot-swappable disk drives, which would actually be cheaper as well as meaning that no machine was down for more than three minutes for the switch.

actually only achieve 290,000 images per machine per year. Over three years we should hit 870,000 images at NZ$0.42 per image including the cost of the capital equipment. Accommodation costs are missing as is an element profit margins if we outsource—even if we gave very generous profit margin of 15% it would still only be about NZ$0.48 a page.[88] These costs are not too divergent from the actual costs of digitising 8000 volumes covering some 2.5 million pages by The Making of America IV. The costs associated with preparation, shipping, quality control, generation of page level metadata, OCR and SGML generation, scanning (US$ 0.13 per page), and process management came to about US$ 0.27 per page.[89]

There are a number of assumptions that underlie this scenario. The material must be consistent[90], the technical metadata needed for each image must be automatically collected, the descriptive and administrative metadata must be available and hierarchical, and minimal quality assurance required. Adding metadata creation costs to the overall cost model would detract from the like-for-like comparison between the microfilming and digital imaging because no additional metadata is typically required for the outputs of microfilm projects. Technically it would be possible to produce the digital images without adding much metadata beyond that which the image system automatically creates and existing bibliographic information, but that would mean not taking full advantage of the capabilities of the environment. The European Commission Fifth Framework Programme project META-e has investigated the possibilities of automated metadata creation and demonstrated the emerging potential of OCR (optical character recognition) to handle document typing and page classification.[91] If metadata creation can not be automated, and the additional costs of their creation need to added in it does push up the cost of the reformatting beyond that of microfilming, but adding the metadata brings significant advantages to users that may outweigh the additional financial outlay. This approach, of course, depends upon the development of the delivery mechanism separately and the establishment of the suitable storage repository, but the Library will need both these elements anyway. The cost of delivering different types of content across the underlying service will be marginal. Migration of images overtime will be a cost the Library will need to carry, but the results of work done at the *Archivo General des Indias (Sevilla)* project[92] and research done at the University of Leeds with 'migration on request' suggest that where material is created to consistent

[88] If the Library could run the station eighteen hours a day it could produce 2.6 million pages over three years at a total cost 708,000 NZ$ or 0.27 NZ cents per page. Even if it were to assume this work was outsourced and outsource costs would incur a further 15% profit margin (paid at 10% per year with and additional 5% paid at the end of the 3 year contract on delivery of the total 2.6 million pages) it would still only cost 0.324 NZ cents per image.

[89] M. Bonn, (2001), 'Benchmarking Conversion Costs: A Report from the Making of America IV Project', *RLG Diginews*, 5.5), http://www.rlg.org/preserv/diginews/diginews5-5.html#feature2

[90] You can not scan an A4 page followed by a sheet from a newspaper, followed by a piece of ephemera followed by an A4 page and achieve this. You pretty much need to digitise by format and size as well as content.

[91] http://meta-e.uibk.ac.at

[92] See http://www.erpanet.org/www/products/toledo/Toledo%20Report%20v5.pdf , 20-26.

standards the costs of reformatting terabytes of data will be minimal after the first object has been transcoded[93]. Migration is increasingly becoming a key strategic approach to the curation of large data sets.[94]

These figures compare favourably to microfilming costs. NLNZ staff provided cost information for microfilming based on the costs of manuscript and serial filming undertaken in 2000. They updated the costs to 'reflect current direct labour rates and film cost, and the depreciation charge has been modified to reflect a three-year working life for the equipment. Other costs have not been updated, including estimated overhead cost.'

Cost of Creation of Microfilming Objects

	Estimated Current Cost c per frame
Direct Labour (10.25 hours per film of 430 frames)	0.4105
Direct Materials	0.0307
Machine Usage	0.0248
Overhead	0.0300
Total Variable Costs	0.4960
Fixed Costs ($2.43 per film of 430 frames)	0.0057
	0.5017
Estimated cost per frame	$ 0.50
This compares with the current contract costs for outsourced filming of	
Microfilming manuscripts	$ 0.69
Microfilming Newspapers & Serials	$ 0.39

The costs on microfilming newspapers and serials are very close to those from digital imaging the same material. Comparison of the manuscript costs are a little more dangerous to do accurately because they will depend upon the consistency of the material, problems of selection and preparation for scanning, and the costs of returning material to the shelves. The one area where they fail to compare is long-term storage. Microfilm is much cheaper to store, but it also tends to have low, but high, cost usage (even if those costs are carried by the user). The digital resources have higher storage costs, but they offer readers quicker retrieval times and greater flexibility in how they may be used. They produce higher quality prints than you can obtain from microfilms, can be accessed by multiple

[93] Phil Mellor, Paul Wheatley, and Derek Sergeant, (2002), 'Migration on Request, a Practical Technique for Preservation', M. Agosti and M. C. Thanos (eds.), *ECDL 2002*, LNCS 2458, 516–526.

[94] M. Lübeck, A. Valassi , *et. al.*, (2003), 'An Overview of a Large-scale Data Migration', Twentieth IEEE/Eleventh NASA Goddard Conference on Mass Storage Systems & Technologies, 7-10 April 2003, San Diego, http://www.storageconference.org/2003/papers/06-Lubeck-Overview.pdf.

individuals in a variety of both local and distant locations simultaneously, and they will be in colour where that is an advantage.

For text-based materials digital representation has the further advantage that optical character recognition (OCR) applications can be used to improve the ways users can investigate the content. The sophistication of some OCR software packages means that not only are these packages able to recognise textual elements, but they can handle articles split across columns. Work at the University of Tel Aviv digitising the *Palestine Post* has shown just how the combination of digital imaging with OCR-based indexing can transform the way researchers can use collections.[95] Resources such as the *Papers Past* could similarly benefit from the application of such technologies.

Recommendations:

The Library needs to complete and adopt its digitisation policy within this financial year. (1)

A digital library policy needs to be established by the end of the first quarter of the next financial year. This must include a sketch as to the nature of the digital library that it is trying to develop. (2)

The Library should consolidate its digitisation activities within Corporate Services and manage these through a cross-disciplinary oversight committee to ensure that resources are effectively allocated and there is consistency to its digitisation work whether this is conducted in-house or outsourced. (3)

This oversight committee should also consider the relationship between the digitisation programme and the microfilming activities and consider whether or not newer technologies provide suitable security to enable it to discontinue microfilming.

Wherever access to digitised textual sources can benefit from the application of appropriate optical character recognition technologies it should be applied (e.g., Papers Past).

[95] Ronald W. Zweig, (c. 1999), *Retrieving Text from Digital Images Lessons from the Palestine Post Project*, http://kipp.tau.ac.il/lessons.htm

What will a Digital Library Programme Cost

> A surprising observation made during my study trip was that with one or two exceptions, national libraries have done very little long-term corporate planning for their new roles in the digital age. Most recognise that they have inadequate technical infrastructure in place to support their digital collections but are unsure what to do about this. There was little evidence of attempting to integrate new activities and roles into strategic planning or mainstream operations, and there is no understanding of the costs entailed in digital archiving. (Pam Gatenby, Assistant Director General, Collections Management, National Library of Australia, 2002).[96]

The NLNZ will be ingesting at least four classes of content (a) products of its own digitisation programmes, (b) packaged resources whether these are delivered on CD-ROMS, online (e.g., data services that wish to lodge copies of their material with the library), or tape, (c) digital manuscripts, and (d) web harvested materials. The costs associated with each of the objects classes will vary both across object classes and within classes as some instances of these classes are more difficult to handle than others. Despite numerous attempts to quantify the costs of building digital libraries the costs of selection, acquisition, ingest, and cataloguing of digital content remain a matter of guesswork. Where organisations have attempted to produce detailed costings they have done so mainly at the macro level and against an array of assumptions and guesses that can not easily be verified or replicated. This in part reflects the lack of understanding of processes and workflow involved in this activity. It also echoes differences of opinion as to what is involved in digital library content collection. For example, even with the benefit of the extremely thoughtful work that the National Library of New Zealand has produced to assist it with understanding workflows[97], these models do not provide us with sufficient information to allow us to estimate the effort associated with each process that they stipulate. Process-based costing information is likely to be available only when the Library has run its digital library programme for some time. Then their detail will reflect the way it is monitored.

A survey of Digital Library Federation members, mainly research libraries at large US academic institutions, showed that 'the average spending on all aspects of digital library programs was $4,341,798 ($2,641,798 if costs of acquiring access to commercial electronic content are excluded). Fewer than half of the DLF libraries surveyed invest primarily in digital reformatting

---

[96] Pam Gatenby, (2002), 'Report on Senior Executive Fellowship to Research Digital Archiving in National Libraries', http://www.nla.gov.au/nla/staffpaper/2002/elect.html. In the report 'the term "digital archiving" is used as shorthand for the full range of activities associated with collecting and managing publications in online form for access – for instance; selection, collection, quality control, cataloguing, file management and preservation for on-going use'.

[97] *WKOF* 2002, *WKS1* 2002, *WKS2* 2001, *WKS3* 2001, *WKS4* 2001, *WKON* 2002, *WKWK* 2002.

programs. Most have oriented themselves toward the development of technical infrastructure and of various reference and other end-user services.'[98] These libraries have found the scale of investment involved in generating the framework and service infrastructure to be substantial. The University of Harvard has focused on developing a digital library infrastructure and in the four years since its inception has invested some $12 million dollars (US) in the exercise.[99] While many institutions now engage in digital library development and service provision there is a lack of consistency in objectives and services. Where financial data is available it is either not provided at a sufficient or the comparable level of granularity to enable cross-programme assessments.[100]

Examples of costing models come from or can be derived from the work of the National Digital Archive of Datasets (NDAD)[101], the Research Libraries Group, OCLC, NEDLIB, and National Archives Records Administration (NARA). But most of these organisations are not dealing with the diversity of media that the NLNZ will need to handle. The digital library may find its closest parallel in the concept of data centres, those organisations that have been developed in many countries to preserve social science and scientific data sets. They receive and provide access to data sets online and on portable media. The process that these organisations must undertake to bring these data into the collection is relatively comparable as well. Based on a review of fifteen data centres SGT has begun developing a micro-level cost modelling tool for NASA. ''The SEEDS (Strategic Evolution of Earth Science Data Systems) cost estimation model and coupled requirements sets developed to support the SEEDS Formulation team in estimating the life cycle costs of future ESE data service providers and supporting systems, where 'data service provider' is used as a generic term for any data/information related activity.''[102] The SEEDS model details the cost elements involved in digital preservation for all stages from ingest through to delivery.[103] (It excludes the processes of selection and de-accessioning, probably because these are not activities which the data curators control.) In its current form the model is fairly abstract and has yet to be

---

[98] Daniel Greenstein and Suzanne E Thorin, (2002), *The Digital Library: A Biography*, (Washington, D.C.: Digital Library Federation and Council on Library and Information Resources, December revision), 2.

[99] For comparison purposes the Library holds some 14 million volumes, has an annual budget of US$80 million (2001), and just over 1000 employees. The University of Michigan holds about half as many items, spends roughly 50% less, and has just under 500 employees.

[100] Perhaps we should not be surprised that the costs for digital library remain vague when we consider that 'most organisations in today's increasingly IT-dependent economy are unable to use cost accounting systems to accurately determine "real-time" IT processes and activity costs' (James Gerlach, Bruce Newmann, Edwin Moldauer, Martha Argo, and Daniel Frisby, (2002), 'Determining the Cost of IT Services', *Communications of the ACM*, 45 (9), 61-67.

[101] Kevin Ashley, (2000), 'Digital Archive Costs: Facts and Fallacies', *Proceedings of the DLM-Forum on Electronic Records: European Citizens and Electronic Information: the memory of the information society*, (Brussels, 18-19 October 1999), INSAR Supplement IV, 121-126. Details of NDAD and its work can be found at http://ndad.ulcc.ac.uk.

[102] G. Hunolt, (16 January 2002), *SEEDS:* Requirements / LOS & Cost Model -- Working Paper, http://lennier.gsfc.nasa.gov/seeds/LOS_020116.pdf .

[103] *ibid*., 20-37 (Sections 4.3 through 4.5).

reformulated in a way that will make it useable as a tool; SGT anticipates completing this process by September 2003.

There are a number of revenue cost areas that will need to be covered and the amount of specialist effort involved in each of these stages or functional areas will be reflected in the per unit cost of taking an item from selection phase to distribution and preservation.

Management (M)
Costs included in this category are management and administration costs associated with delivering of the digital library service, including strategic planning, management of staff (e.g., human resources costs), logistics, supplies, facilities, security management, and property inventory and management.

Services (F)
These are the non-staff costs associated with maintaining the capital equipment in functional order, such as system maintenance contracts. In the case of the NLNZ these might be carried elsewhere, but need to be reflected in the model.

Selection (S)
The effort involved in selecting objects. How will selection be handled? Will it be possible to define a clear set of policies which are simple and inexpensive to apply? Will it be feasible to broadly apply these policies or will decisions involving expensive curatorial input need to be made on a case-by-case basis? It will only be through automating the process that it will be possible to reduce the costs. Indeed, for some classes of digital material, such as packaged objects, specialised staging and viewing environments for the selectors may be required and their use may require technical support.

Acquisition (A)
What effort will be involved in the acquisition of the digital object? In the case of New Zealand where the Library Act 2003 enables the Library to define and clear materials at category level through Notice of a Requirement, the costs of acquisition might be contained and indeed front-loaded (in the research and analysis necessary to develop an informed Notice).

Ingest (I)
Ingest will involve the reception of the digital object whether on a fixed carrier or online, integrity checking of the digital object itself and any associated items (e.g., packaging, software), ensuring that the object functions as expected, checking that the objects carries no hazards (e.g., viruses), its insertion into the digital repository, and verification of the copy and its completeness[104]. For some digital

[104] It is especially crucial with web-based materials that verification be done immediately.

objects it may be necessary to carry out format conversion, migrate the data to a new environment, and extract metadata. Where ingest of a digital object will require access to a technical support specialist the costs will be pushed up.

Cataloguing & Metadata Creation (CMC)
Digital objects will need to be catalogued and have a range of additional administrative and preservation metadata associated with them as defined by the work completed so far by the DLTT.[105] Metadata creation puts huge overheads on digital collection and presentation. Significant reductions in cost could be achieved where digital object originators can be encouraged to supply their digital objects with the appropriate metadata or where the necessary metadata can be automatically extracted from the materials. The costs of creating the metadata from scratch, therefore, will dramatically push up the costs of acquiring a digital object.

Processing (P)
Where objects are associated with applications or other software and these are retained by the Library as the primary way to access the electronic object, it may be necessary to process the resource on a regular basis to ensure that the applications continue to run as the underlying repository environment changes or the kinds of available system technologies change.

Documentation (D)
For some digital objects the metadata may not be sufficient to enable users to access the objects and additional documentation may be required. For instance where resources are linked to applications user manuals may be essential.

Archiving (V)
This macro cost involves a number of sub-costs including placing the digital object in the repository, managing it, securing the link between the metadata and the object, regular quality checks, and the creation of redundancy (e.g., backup copies). This cost point might be analysed in more detail to account for costs associated with longer-term accessibility of the digital object, such as migration or emulation. These latter costs are impossible to estimate at this stage however.

Access (AD)
Costs associated with providing user access will vary depending upon the structure of the repository and the type of service layer the Library wishes to offer potential users. Online access will probably be the lowest cost, but this will depend upon the level of resiliency that the Library aims to achieve and whether online access is local or remote. (Many of the costs of providing access mechanisms – the interface

[105] e.g., MSMP 2000, MSFW 2000.

layer – might already be hidden in the digital library software that the Library has put in place. Alternatively if the Library adopts a service layer model they may be carried by Commercial Access Providers.)

User Support (US)
The level of user support to be provided to digital library users will also impact on costs. Users will expect some level of user support. How much will be free? What will be charged for?

Technical Coordination (TC)
A portion of the costs that the Library will need to carry associated with coordination of standards, interfaces between applications, development and maintenance of system architectures, and contributions to format registry activities will need to be attributed to each object.

Implementation (Im)
There will be costs related to developing the digital repository that will need to be amortised across all the digital objects in the repository across time. These will not be one-off-costs either, but every three to five years it will recur.

The combination of all the costs associated with each of these areas will determine the costs of the digital library service on an item-by-item basis – Item-level Digital Library Costs (IDLC)— as well as on aggregate annual basis—Annual Digital Library Costs (ADLC). Using this model you might arrive at these costs as **IDLC = M+F+S+A+I+CMC+P+D+V+AD+US+TC+IM**. The difficulty is that we do not have sufficient information about the actual costs at micro-level to allow us to arrive at the underlying unit costs. Since we have no metrics for establishing the per unit elements of the costs an alternative method for arriving at the costs must be found. There are other cruder ways of arriving at costs. For a narrow class of digital objects they might be calculated on a per document basis as Dürr and Meer did. They argued that 'for metadata assignment by the library plus administration and quality control plus the infrastructure of the operational electronic archive the estimated costs are now estimated to be € 10 + € 10 + € 9 = € 29 per document' but over a 50 year period they calculate that the cost will rise to 40 euros per document.[106] If costs are to be reduced we need to be able cost the processes.

The Library will need to make a crucial decision as to whether it will provide access to the digital objects that it acquires or whether it will merely engage in bit-stream collection and preservation. As we noted above this decision may need to be taken on an item-by-item basis, whether by the selectors alone or on the basis of advice from technical advisors (e.g. Digital

[106] Eugene Dürr and Kees van der Meer, (2001), *Emulation and Conversion: Organisational and Architectural Overview - way of working, costs, methods*. Report at the E-Archive project; version 1.2; http://www.library.tudelft.nl/e-archive/Documenten/Resultaten/roquade2.pdf, see 7-9.

Object Analyst or Content Ingest Support Specialist). For many objects bit-stream preservation will be the most cost effective way to secure the object itself, but this is a critical strategic decision because it focuses the digital library on preservation at the expense of access. Users may be frustrated that while they can discover some material they are unable to access it and the Library exposes itself to the possibility that it might need to undertake significant research and development work in the future to access these bit-streams.[107]

Increasing the number of potential depositors will increase the costs. If the Library could reduce the number of digital publishers with which it deals it could reduce the costs of acquisition, but it might achieve a better result if it could encourage publishers to supply material for addition to the archive in specific formats and with essential metadata elements already attached in ways they could be automatically extracted. Volume of material may be a factor, but it is more likely that costs will be related to a combination of volume, number of files, frequency of accessions (1000 files in one load cheaper than 1000 files in 1000 deposits), the inter-relationship between different digital objects, and the difficulties associated with linking the digital resources to associated software. The Library will wish to ensure that it evens out the influx of digital objects during the year by controlling the timing of material arriving to ensure that it maximises the throughput that its staff can achieve.

[107] New Zealand has some excellent University departments of Information and Computing Science. There is a possibility for capacity building by promoting research in digital preservation within these departments.

Storage Costs

The costs of providing the storage for an archive consist of three core elements: (a) actual online or near-line storage costs, (b) the costs of backup media and securing it, and (c) technical support staff to manage the servers. Cost (a) is typically classed as a capital cost, and (b) and (c) as revenue costs. Without an indication of the quantities of storage we are looking at any estimate of storage costs is difficult, but we know that they have declined dramatically and constantly since the late 1950s. Disk storage cost in 1992 roughly US$2 per megabyte and in 2002 about US$2 per gigabyte.[108] It is by far the cheapest type of storage available. In 2001 for instance, disk storage cost between US$0.004 and US$0.02 per megabyte compared with US$0.054 per megabyte for DLT and US$0.42 for optical media. It is worth bearing in mind that you will probably need to replace your storage system every three to five years and that during this period for every unit (dollar, pound, or euro) spent on the purchase cost at least 2 further units will be spent on maintenance. OCLC has instituted a charging structure for its repository with amounts of less than 100 gigabytes costing US$60 per gigabyte per year and costs falling back to US$15 per gigabyte per year where the organisation leases a terabyte or more per year.[109] The discrepancy between these costs (100 gigabytes of storage would cost US$200 if purchased and US$6000 per year if leased) is explained by the fact that the costs of the raw storage do not include costs for facilities (e.g., air conditioned space), staff, technical services, and backups. Even when these costs are included the OCLC costs do appear to include a hefty profit margin—but in such a high-risk arena this is necessary. The other problem with raw storage is that the costs do not actually scale well and this relates to the complexity of system infrastructure needed to manage effectively a couple of terabytes of data; it is far greater than that required to handle fifty gigabytes. A comparative set of costs from the UK[110] comes in (in June 2003) at around 135,000 GBP for a Storage Area Network (SAN), although the Library would have to decide whether this particular storage model is the right one for their digital repository (contrast with NAS or HSM). This consists of an archive server (SUN) [20K GBP], SAN storage (4.3 terabytes), controller and switch [60K GBP], a backup system with four hundred tapes [35K GBP], and elements related to connectivity (hub/switch/fibre) [20K GBP]. Additional storage comes in at about 7.5K GBP per terabyte. The initial system comes in at about 31 GBP a gigabyte with additional storage at about 7.5 GBP per gigabyte. These figures compare favourably to the costs of leasing. Hendley and others have argued that data storage costs are not major costs

[108] http://romulus.gsfc.nasa.gov/msst/conf2002/PPT-PDF/c04e-gart.pdf

[109] S. Chapman, (2003), 'Counting the Costs of Digital Preservation: Is Repository Storage Affordable?', *Journal of Digital Information*, 4.2, June, http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Chapman/

[110] This is probably a more valid comparison than with US costs because it seems that Britain and New Zealand suffer from similar island country mark-ups.

in the model.[111] Experts on the National Science Foundation Blue-Ribbon Advisory Panel Cyberinfrastructure reported that '…the cost of data repositories (done correctly) will be dominated by the recurring costs of personnel performing curation, maintenance and upgrade, and providing user advice, assistance, and support.'[112]

The costs of the storage are only a small part of those associated with building a digital repository and may be increased if an adequate and desirable level of redundancy is incorporated into any design.

[111] Tony Hendley, (1998), *Comparison of Costs and Methods of Digital Preservation*, (London: British Library Research and Innovation Report 106), http://www.ukoln.ac.uk/services/elib/papers/tavistock/hendley/hendley.html; Mary Feeney, (1999), *The digital culture: maximising the nation's investment* see especially chapter 5: Estimating the costs of digital preservation, (London: National Preservation Office). http://www.ukoln.ac.uk/services/elib/papers/other/jisc-npo-dig/

[112] National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, (2003), *Revolutionizing Science and Engineering Through Cyberinfrastructure*, (January), page 77. http://www.communitytechnology.org/nsf_ci_report/report.pdf

Cost Planning Analysis

In the end the estimates vary and the cost models are not consistent. This is especially true when we just look at the areas of web archiving[113] and digital preservation.[114] It is unlikely that any model that might be developed at this stage would give a true and accurate estimate of the costs involved in building a digital library. For the moment it seems safer to define the shape of the team that would be required to support the process and to determine the staffing costs in much the same way that the large data archives examined in the *ESDIS Data Center Best Practices and Benchmark Report* have done,[115] on the basis of workload and staffing. They investigated fifteen data centres each of which managed between 4 and 730 terabytes of data a year. The consultants examined the staffing against a suite of metrics that included terabytes ingested annually, complexity of ingest, numbers of products brought in, number of products made accessible, and number of users. This model is risky because the Library will embark on developing its digital library without any prediction of costs per object. But the digital library projects that have been carried out so far have only given us an understanding of the functional areas in which effort is needed. The cost structures will relate in this model to the functional areas in which staffing is required and the costs of technical infrastructure (e.g., storage system, desktop PCs for staff, a peripheral device library).

The costs are likely to vary over time, but are unlikely to move downward. In the development phase they are very likely to involve substantial investment in both capital and revenue categories. The decline in the capital costs is likely to occur as storage costs continue to fall, but any declines in storage costs may be cancelled out by rises in the amounts of storage required (e.g., as the Library begins to bring online its audio collections or it begins to ingest objects rich in moving image material the demand for storage will rise). Declines in capital costs might also be offset by increases in revenue costs. The staff costs will be a function of the time it takes to select, acquire, ingest, process, and catalogue each object, as well as the number of objects that the Library intends to admit to its collections in any given year. Achieving a precise relationship between the numbers of

[113] For example, Day, 2003, 23-24 noted that the British Library had made a bid to UK government for an additional funding line of 600,000 GBP (beginning in 2004) to enable it build on the results of the Britain on the Web originally known as Domain UK (see Woodyard above) and undertake regular archiving of websites. It was estimated that 10,000 websites would be selectively collected and a further annual bulk harvest would be carried out.

[114] For comparison see: Ashley 1999 (above); Stewart Granger, Kelly Russell and Ellis Weinberger, (2000), *Cost elements of digital preservation* (version 4), http://www.leeds.ac.uk/cedars/colman/costElementsOfDP.doc, Shelby Sanett, (2002), 'Toward Developing a Framework of Cost Elements for Preserving Authentic Electronic Records into Perpetuity', *College & Research Libraries*, 63.5, 388-404; *Making of America IV (The American Voice 1850-1876): Assessing the Cost of Conversion*. University of Michigan, Digital Library Services, July 2001, http://www.umdl.umich.edu/pubs/moa4_costs.pdf.

[115] G. Hunolt and A Booth, (9/2001), *ESDIS Data Center Best Practices And Benchmark Report*, (Science Operations Office, Earth Science Data and Information Systems Project, Goddard Space Flight Center, NASA Contract NAS5-00154).

staff engaged in building the Library's digital collections and the numbers of items added to the collection in any year will never be precise. In contrast to the paper world, where there was little risk that if items arrived at the Library but were not assessed immediately that they would decay (or become inaccessible) the same can not be said for digital objects. In the longer term the revenue costs of running digital library services will need to be balanced against the other costs of the Library. That is if the numbers of print items acquired by the Library begins in the longer term to decline the revenue associated with the acquisition of print items will probably need to be transferred to digital library activities. The model for defining staff costs in this early digital library is a *time -1* costing model (see below).

Recommendation:

The Library should improve its economic modelling of the costs associated with collecting and maintaining digital objects whether coming from digital objects on fixed media, digitisation, or web harvesting. This should include consideration of the economic impact of these resources on building the creative economy of New Zealand, encouraging tourism and services to industry. (8)

Developing Digital Library Capacity: An Options Appraisal

DIGITAL LIBRARY SERVICES GROUP

The major challenge facing the Library will be how to establish the capability to select, acquire, catalogue, manage, conserve and preserve, and provide access to, at least, the four classes of digital objects identified as target classes for inclusion in the Library's collections: packaged objects, website harvests, unpublished digital materials, and the outputs of the Library's digitisation programme. Few institutions anywhere have experience managing the acquisition and preservation of these classes of objects individually and none can be said to be proficient in managing all four classes.

The Library currently does not have adequate capacity to deliver a Digital Library Service. How can a digital library activity be developed within the Library so that the digital library is part of the Library as a whole? With the help of staff from across the Library five options were considered.[116]

- do nothing;
- monitor initiatives in other national libraries and wait until they have solved the problem through developing tools, benchmarked and costed workflow strategies, and transferable policies and procedures;
- maintain the current approach of building capacity through a small team focused on developing the intellectual framework for the digital library;
- develop an integrated approach, with different digital library activities happening in different parts of the Library but with improved communications between the units delivering aspects of the digital library service; or
- adopt a team-based approach, a Digital Library Services Group which would handle everything from digital library level policies and procedures to selection through to access to digital materials. This team would have a limited life with the longer-term objective of reintegrating the activities into business units across the Library once standards, workflows, best practices, and tools had been established for handling our four classes of digital objects.

Each of the options has merits and the objective of this options appraisal is to look at each of the options in turn and to identify the strengths and weaknesses associated with the option. This is followed by a brief discussion recommending one of the options as a way forward.

[116] All scenarios need to be cognisant of the relationship between the National Library and the Alexander Turnbull Library (ATL).

OPTION ONE: DO NOTHING (MODEL 20TH CENTURY LIBRARY)

The assumption behind this option is that the digital revolution is still in its formative stages and that it is difficult to predict both what future technological developments might hold and what of our current culture that is represented digitally is worth incorporating in a digital library and preserving for future researchers.

The strengths of this approach include that:

- the option is low cost and can be pursued within the current budgetary constraints;
- it requires no changes in staff duties;
- it lets other national and international agencies take the risks and carry the intellectual development costs associated with establishing digital library services;
- it ensures that the Library maintains its existing work programme and continues to meet current commitments, because it does not require other priority areas to cut back on their funding and activity; and,
- it promotes the concept of the Library as Museum.

The main weaknesses to this approach are that it:

- does not enable NLNZ to meet its obligations under the National Library of New Zealand Act 2003;
- exposes the NLNZ to the risk that other institutions might usurp its traditional role as the trusted repository for New Zealand's published and unpublished heritage;
- could lead the NLNZ to fail to meet its obligation to provide New Zealanders with access to their information heritage;
- would undermine the work of the National Digital Forum (NDF) and the expectations of the community for the direction that the NLNZ should show in the area of digital libraries;
- does not enable the NLNZ to fulfil its leadership role in the areas of digital library development and preservation;
- would represent a failure to enable the professional development of NLNZ staff;
- would lead to a downward trend in the amounts of published and unpublished material collected by the Library;
- would act as an obstacle to the Library in its efforts to service its policy advice roles; and,
- the Library would miss a unique opportunity to rethink its current commitments and develop a service that responds to the changing information landscape.

OPTION TWO: MONITOR TO ADVISE – ACTIVE NOT PASSIVE

In many ways this approach is similar to option one, but it involves the more active monitoring of the work that is going on in digital library service design, implementation, and development. This option depends upon the Library establishing an active mechanism to monitor initiatives in other National Libraries and conducting regular re-assessments.

The strengths of option two are that:

- the costs are low as only one or two staff members would be required to fulfil this role of monitoring and keeping the Library informed about developments at other institutions;
- no specialised up-skilling would be required and therefore no immediate investment in staff development would be necessary;
- it lets other national and international agencies take the risks and carry the intellectual development costs associated with establishing digital library services;
- it enables the Library to maintain its existing portfolio of activity;
- the monitoring team could share the results of their watching brief with colleagues at other New Zealand institutions; and,
- it limits the immediate risk exposure of the Library.

On the negative side of the equation this approach has a range of limitations, all of which are shared with Option One. This option:

- does not enable NLNZ to meet its obligations under the National Library of New Zealand Act 2003;
- exposes the NLNZ to the risk that other institutions might usurp its traditional role as the trusted repository for New Zealand's published and unpublished heritage;
- could lead the NLNZ to fail to meet its obligation to provide New Zealanders with access to their information heritage;
- would undermine the work of the National Digital Forum (NDF) and the expectations of the community for the direction that the NLNZ should show in the area of digital libraries;
- does not enable the NLNZ to fulfil its leadership role in the areas of digital library development and preservation;
- would represent a failure to promote the professional development of staff in a key emerging area of Library activity;
- would lead to a downward trend in the amounts of published and unpublished material collected by the Library;
- would act as an obstacle to the Library in its efforts to service its policy advice roles;
- means that the Library would miss a unique opportunity to rethink its current commitments and develop a service that responded to the changing information landscape; and,

- there may be no good time to start a digital library service.

OPTION THREE: MAINTAIN CURRENT APPROACH

The Digital Library Transition Team has been very successful (see Section2 above) over the past couple of years in laying the foundations for a digital library development and it may be that there is much more to do before the Library is in a position to begin development of a fully functional digital library service. The premise behind this approach is that before actually launching a programme the foundation work should be completed.

The positive benefits of this approach are that it:

- supports the continued development of an already successful programme of work;
- involves some degree of implementation;
- engages cross disciplinary activity;
- has been very successful so far in establishing visibility for its successes at local, national and international levels;
- builds on work that has already been completed;
- contributes to national, and international work in developing digital library fabric, such as standards;
- enables the Library to incorporate development of its digital library capabilities in an incremental and gradual way; and,
- it is cost neutral as the infrastructure needed to deliver this level of service is already in place.

The weaknesses of this approach include:

- people working in different areas engaging in these activities can lead to poor communication, problems about prioritisation, and synchronisation across units;
- the continued slow progress towards developing the digital library infrastructure;
- resource constraints that are driven by supply;
- lack of organisational integration and support for this activity;
- the number of units within the organisation that are involved in contributing to the current lines of development and claim ownership of the activity;
- the lack of clarity of the Library's digital library development strategy to the outside world;
- the length of time it will take to develop a service may be so long that the digital library service is not delivered in time to meet the Library's need;
- failure to ensure the professional development of Library staff;
- lack of visibility of the significance of the issue within the Library;

- the failure to enable services to meet the information needs of the New Zealanders; and,
- its inability to enable NLNZ to meet its obligations under the National Library of New Zealand Act 2003.

OPTIONS FOUR and FIVE

These two options focus on the development of a sustainable digital library infrastructure that ensures that the Library can deliver a full array of digital library services. They share four characteristics that are worthy of mention at this stage:

- they both require additional new staff;
- each of them only works if the problem of how they are managed can be addressed;
- they both need to be well integrated into the overall business planning of the National Library and can not be conducted as additional business; and,
- the impression of their success in the long term will depend upon more than measuring such outputs as ingests into the collection but will require that broader impacts on society (e.g., economic growth, tourism, educational benefits) are measured if the costs associated with the programme are to be justified.

OPTION FOUR: Integrated APPROACH

Option four builds the digital library team on the back of existing Library services and uses instruments such as Service Level Agreements (SLAs) (or Statements of Service Levels) and collaboration tools to ensure that the distributed team works as a coherent unit.

Among the strengths of this option are that it:

- enables the NLNZ to meet its obligations under the National Library of New Zealand Act 2003;
- provides New Zealanders with access to information resources and guarantees the long term survival of the Nation's heritage;
- builds on the current skills of existing staff;
- encourages the Library to tighten up the processes surrounding the digital library services that it provides from selection to preservation;
- depends upon only minimal restructuring as it relies on the federation of staff rather than their redeployment; and,
- engages staff in many parts of the Library in the activity which will contribute to the development of an internal culture and to the widespread realisation that digital library services are part of the larger range of services that the Library must provide.

The downsides of this approach are that it:

- depends on Service Level Agreements (SLAs) and/or delegation of activity which is difficult to manage and to enforce internally,
- requires an increase in staffing;
- depends upon rigorous auditing of processes and deliverables, fragmented management, and distributed control;
- makes prioritisation of activity difficult because the staff have conflicting allegiances;
- creates a complex framework that will make workflow management and control difficult;
- will be difficult to benchmark the performance of the effort because it will need to be measured across separate units;
- will be tempting to meet the requirements of the initiative by asking existing staff to take on additional duties rather than by expanding the effort dedicated to digital library development and preservation which will mean that customer needs are unlikely to be fulfilled; and,
- the integrated model will make collaboration with other institutions complex.

OPTION FIVE: Team-based (TIME BANDED, 5 years)

A further alternative is to develop a Digital Library Services Group to manage the development of digital services for the Library. This would involve everything from establishing digital library policies through to managing the development and deployment of the digital repository. Digital library policies will be created against the framework of other corporate level policies, such as collections, access, and preservation policies. The Group would be established for a limited term while the practices were defined and the requirements established. The view would be to have defined the services and the mechanisms for delivering them within the first twelve months and then to move on to benchmarking services with the eventual goal at the end of sixty months of having the services re-integrated into existing Library units.

The strengths of Option Five are that it:

- enables the NLNZ to meet its obligations under the National Library of New Zealand Act 2003;
- provides New Zealanders with access to information resources and guarantees the long term survival of the nation's heritage;
- builds on the current skills of existing staff;
- is highly visible as a unit both nationally and internationally and brings substantial publicity benefits to the Library;

- will be easier to engage other organisations in collaborative efforts because the Library's profile and commitment to the activity is clear;
- contributes to developing a digital services culture within the library and promotes further skills and a knowledgeable support community;
- is relatively easy to manage because conventional management approaches can be adopted;
- would be highly interdisciplinary because the Group would include a wide range of staff with different specialisms working as a team to develop digital library services strategies; and,
- will be easier to benchmark performance because the deliverables/outcomes are directly tied to the team conducting the work.

The downsides of this approach are that it:

- will require additional staffing;
- changes the shape, working conditions and focus of some business units, if only on a temporary basis;
- exacerbates the feeling of fragmentation between the traditional library teams and the digital library team;
- takes staff (some of which might be cherry picked) away from some business units and may require that these posts be backfilled for the period that the staff member is seconded;
- may prove difficult to deconstruct such a Group once it has been established because of the territoriality that will naturally emerge; and,
- depends upon senior management buy-in and action to ensure that the organisation can deliver its promise.

Selecting the most viable and sustainable option

The key here is to identify the option that will most readily benefit the Library by enabling it to develop a sustainable digital library service. The most compelling reason for rejecting Options One and Two is that they do not enable the Library to address its mandate under the new Act. Option Three also does not enable the Library to respond effectively and flexibly to the changing information landscape; it puts in motion a slow process and does not deliver for the Library all the pieces of the digital library service coherently. While there is no reason to produce a New Zealand Digital Library as a distinct organisational entity a more coherent approach is needed than can be provided under options one, two, or three. During the short term it may be necessary to develop the services as an independent unit. The management complexities of Option Four and the numerous interfaces it requires between operational and technical service units within the Library indicate that this option has a higher than acceptable risk of failing to deliver a suitable digital library service in the near term. Option Four also makes it very difficult to ensure that the necessary policies and procedures are put in place.

Option Five offers the Library the most effective way to develop its digital library capacity, while containing the risks associated with such a development. It guarantees the mechanisms and steps it will take for NLNZ to fulfil its mandate under the act are visible and measurable. Any politics that might exist in the Library must not be allowed to impinge on the development of an adequate Digital Library Programme. The impact of politics on digital library service development and management might even be eliminated by reducing competition for control of these services. Implementing option five, would establish a group not too dissimilar to the team at the National Library of Australia (NLA) that selects, catalogues, and manages digital assets, poses challenges. In the National Library, the Alexander Turnbull Library (ATL) has responsibility for selection, cataloguers sit in Bibliographic Services, and the current DLTT is based in the Electronic Services directorate. Option five has some genuine advantages: it contains the Library's risk of failure, establishes a clear unit of responsibility for delivering the Library's new mandate under the Act, ensures that the investment of capital, revenue and intellectual effort are ring-fenced, and creates an environment where it is relatively easy to observe and measure the impact of the Group.

There are benefits to be gained from investing in a dedicated Digital Library Services Group. These include:

- cross-disciplinary fertilisation of thinking through integration of expertise of the selectors, acquisition specialists, cataloguers, technologists, standards developers, and workflow and process designers working as a team;

- taking a fresh look at the processes, functions, and infrastructure that are required to make the digital library work;
- establishing a high profile digital library service which would deliver the same profile of, say, special collections to users and others in the Library community;
- coherence of effort;
- facilitating a proactive approach to donors and content creators - e.g., the technologists, standards developers, and selectors can work together to develop guidance to improve the way material that is selected for ingest is presented, and to enable the organisation to reduce the costs associated with ingest, management, and preservation;
- eliminating fragmentation of effort;
- promoting the documentation of processes and the establishing of benchmarking measures to measure performance[117];
- simplifying the application of mechanisms to measure performance; and,
- ensuring that the remit and focus of the team is well defined and can be measured.

It will be essential to ensure that any digital library service developed by the Library is not perceived either internally or externally as something separate from the Library itself. All emerging services must be seen as core activities of the Library. This will require a culture of inclusion, communication, and commitment at all levels to ensure that the digital library service delivers results of value to the Library and its users.

[117] The remit and focus of the Digital Library Services Group needs to be well defined and all its activities designed so that they can be measured (see below, *Evaluation Metrics and Impact Measurements*).

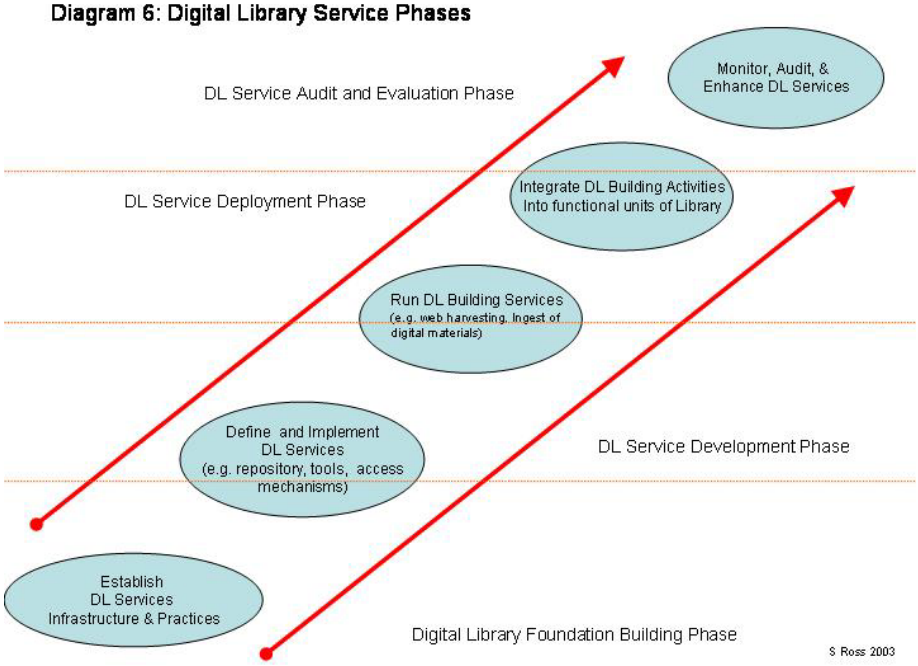
Responsibilities of the Digital Library Service Group (Option 5)

The Digital Library Service Group needs to take responsibility for the development of standards, definition of technical resources, selection, acquisition, cataloguing, provision of user access, and establishment of a trusted digital repository. The Group will need to document and work out how it can transfer these procedures into the Library more generally. The Group needs to report at a senior enough level within the Library that competition for resources during the brief period of its existence can be avoided. An adequate level of staffing will be required to ensure that it can enable the Library to fulfil its mandate to ensure long-term access to New Zealand's documentary heritage.

The key test will be whether the Library can put in place a strategy that takes the digital library programme from its infrastructure development phase and transforms it into a service that is fully integrated into the Library itself. One view of digital library development would divide it into five phases (as in diagram 6). The aim is to move from planning and design to a position where the NLNZ is monitoring, auditing, and enhancing digital library services that are integrated into the fabric of the Library. The transition from intellectual framework to digital library service depends upon the Library first defining what kinds of collections it will be developing and what services it will be providing. Two primary challenges will be to define (a) a realistic collection development strategy that responds to the kinds of material that the Library could collect and the technological infrastructure that it can realistically develop to enable that collection to take place (and be secured), and (b) the need to ensure that adequate levels of access can be provided to the material both now and in the longer term. Diagram 6 attempts to show the stages involved in digital library development as a progression from planning through to audit of services. The Library has so far completed substantial work on laying the foundations for a Digital Library Services infrastructure and practices and done initial work in the defining and implementing digital library services—basically substantial work in stage one is complete and the beginnings of stage two have been carried out. This is par for the course. The DLF study could not identify, among its sample of twenty-one leading university research libraries in the US, a single mature digital library. In fact the authors concluded that 'The digital library is an organisational form that is in flux…It is becoming apparent that the adult digital library program will no longer be organisationally or functionally distinct from the library as a whole.'[118] The model proposed in this report should enable the Library to develop its digital library services in a protected environment until they are mature enough to be spread back into the overall services of the NLNZ itself.

[118] Greenstein and Thorin, 2002, 22-23.

Diagram 6: Digital Library Service Phases



What would a digital library services team cost

We have reviewed above the issues associated with costing a digital library service and how the distribution of these costs changes over the five-year timeframe considered here. What we have not done is to define the structure of the team that would be required to carry out the task and what the annual cost of this team would be. We estimate that it will take a manager plus at least a team of twelve or thirteen to provide a viable staffing structure[119], but this is an estimate that depends upon the following seven key assumptions:

1. the numbers of digital materials will continue to increase year on year[120];

[119] Greenstein and Thorin (2002) found that 'staffing levels for digital library initiatives vary across DLF member institutions (from 7 FTEs to about 48 FTEs). The average is 18 FTEs' (p 68). These figures should be used with some caution because the report does not specify what roles the staff is playing; for instance, this could include staff supporting the delivery of subscription based materials. William Arms reporting on the Library of Congress pilot study into web archiving, Minerva, noted that web archiving was 'not worth beginning without a dedicated team of librarians and technical staff, about ten people initially', W.Y. Arms, (2001), *Web Preservation Project: interim report*, (Washington, D.C.: Library of Congress), http://www.loc.gov/minerva/webpresi.pdf

[120] The key here is that the numbers of digital objects will increase. Increases in size are far less significant. In an environment of declining storage costs larger digital objects will cost only marginally more than smaller digital objects. Indeed it is probably easier to manage a single 1 gigabyte object than to mange 1000 1 megabyte objects, although the more management processes can be automated the more overall costs can be reduced. 'The How Much Information Project', http://www.sims.berkeley.edu/research/projects/how-much-info/index.html, provides one of the most comprehensive analyses of the way information will continue to expand.

2. the Library will selectively acquire materials (even where they are brought forward or proposed as part of the nominated automated deposit [NAD] mechanism described above)[121];
3. it will increase staff proportionately to increases in selection and acquisition targets unless it can reduce effort per unit of work;
4. the Group will take advantage of hierarchical description and automate as much of the process of as possible;
5. the Group will have access to suitable digital repositories and will play a role in the definition of the access layer that will be developed by technical services or sourced from a commercial provider;
6. the Group will have access to skills development through training opportunities that are strategically planned to ensure that the staff has the capabilities to carry out their duties in the face of rapidly changing technology and best practices;
7. the Library will wish to integrate the processes of digital collection development into existing business units at the end of the initial five-year period.

In addition to the Digital Library Services Manager, the digital library services Group should consist of[122]:

- *Two selectors*: responsible for identifying a mix of content in digital form including a balance across the digital object classes. As some material will be brought forward for acquisition as e-manuscript collections held by notable New Zealanders the selectors will also need to support technical preparation of some material for ingest.

- *Four cataloguers* (one in year one, and three further ones, two by the end of year two, and the fourth at the end of year three): The Library should set itself a target of cataloguing a minimum of 4000 distinct digital objects each year. This is about doable with four cataloguers. (Stress the word distinct because some items will be complex objects which will take longer to catalogue because they will consist of multiple elements.)

- *Acquisition and Access Specialist* who can ensure that classes of material are identified under the Act for selection and suitable notification published. The Acquisition and Access Specialist may need to handle those special cases that will arise where the Library will need to gain particular permissions to acquire material (e.g., where digital materials about New Zealand are published [and printed or disseminated electronically] from

[121] Work on the Minerva Project showed that selective collection was 'at least 100 times as expensive as bulk collection.' Arms, *et. al.*, 2001.
[122] A team of this size would also require at least one staff member to provide administrative support.

outside New Zealand and therefore are not covered by the NLNZ Act 2003). It is possible that in this model we have underestimated the scale of the acquisition work necessary because this estimate is anticipating the power of the Notice of requirement for public document concept as recognised under the NLNZ Act 2003 (see Part 4 Section 31). If the Library aims to acquire material from the dynamic (or deep) web it may require substantial action to track down, contact, and arrange for the material to be deposited. Facilitating the acquisition of items put forward under the nominated automated deposit (NAD) scheme would also add to the work of the acquisition specialist.

- *Content Harvester*: The team will require at least one staff member who will operate the harvesting tools and ingest the harvested items into the digital repository.

- *Content Ingest Support*: For items such as packaged objects, unpublished manuscripts, and nominated automated deposit (AND) the process of ingest will require technical support.

- *Digital Object Analyst*: the management of ingest, storage, and delivery of digital objects will depend upon the Library continuing to develop an understanding of emerging object types and environments. While the work of the Content Harvester and Content Ingest Support officer should be fairly straight forward as each class of object that the Library will need to handle is better understood, they will still need to work to the guidance of the Digital Object Analyst.

- *Digital Repository Manager*: Although the actual technical process of setting up, optimising, migrating, backing up, and ensuring the running of the digital repository will be handled by the Library's technical support team (possibly using resources freed through server and resource consolidation) the team needs to have a digital repository manager. This post will be responsible for ensuring that the design of the repository responds to the changing needs of the digital library itself. Also this post will monitor access mechanisms to ensure that as many of the objects that are in the repository can actually be delivered in a functioning way to users.

- *Business Analyst and Strategic Planner:* The Digital Library Services Group will need continually to review its approaches to streamline them and push down the per unit cost for ingest, cataloguing, and preservation of digital materials. To do this it will need access to a business analyst who can look at the workflows, processes, and trends to identify steps for improvement. There will also be an essential need for capacity

building through experimentation and this post could use 50% of its time for this purpose.

- *Standards Officer and Documentalist*: Keeping up to date with international standards and best practices will be central to the development of this service. As was noted above (see above, *Current Work*) the Library has done an excellent job in guaranteeing that its development work reflects internationally agreed standards. The development of best practice guidelines and the recording of working practices will be essential to ensure that the work of the Digital Library Services Group can be transferred back into the Library itself.

While it would be possible to create a hierarchical structure for this team, the nature of the activity and the anticipation that they will be reintegrated into traditional Library departments suggests a flat structure for the team offers little risk. The process of management could be facilitated by tightly defined job descriptions with clearly stated performance benchmarks that are assessed on a six monthly basis. This approach reflects the proposals for evaluating the work of the Digital Library Services Group more generally (see below, *Evaluation Metrics and Impact Measurements*). The actual relationship between the members of the team reflects the provision of services to one another. While overall outcome measures, rather than output measures, should be used to assess the progress of this team, detailed automated tracking mechanisms could be used to monitor their workflow (e.g., numbers of items selected, numbers handed off for acquisition, numbers of items acquired, numbers of items handed off for cataloguing, numbers of items entering the digital repository). This would facilitate the measurement of process costs and ease the identification of bottlenecks. Identifying and employing staff for these positions may prove challenging; we know, for instance, that there are specific skills required to be able to select and to catalogue electronic material. The Library may find that it needs to work hard to achieve the right balance between the costs of employing under-skilled staff and training them up, and buying in expertise at higher costs.

This staffing structure means that during years two to five the total cost of salaries (including on costs) for the Digital Libraries Services Group is likely to be in the order of NZ$ 1 million to 1.5 million, about 5% of the libraries budget. This figure is much lower than the expenditures of other institutions on their digital library teams. A team of this size should be able to ingest, describe, and make accessible about 4000 digital objects per year.[123] The cost would be between NZ$250 and NZ$375 per object.[124] As

[123] A figure based on the fact that we know that it takes about an hour to catalogue a digital object and the constraints on ingest are not those caused by selection or the technical processes of ingest but are based on the actual number of objects that can be catalogued in a year. We used the figure based in part on the work of the Danes of 1 hour per object, which would mean that a cataloguer working 6 hours per day could manage to describe 1330 objects in a 225-day working year. Because this is unknown territory we have reduced the actual target to about

we made clear in the section on cost modelling the Library will only be able to verify the per unit costs on a *time – 1* basis using an approach not too different to that used to evaluate the costing of scientific data archives. The difficulty is that the definition of an object is variable in this instance. It could vary from a single document of a couple of kilobytes to a dataset of a terabyte. So depending upon how you count the cost per information unit these figures could be very low or high. This sort of variation in the types of entities with which we are dealing mean that comparisons with Library estimates for heritage cataloguing (currently NZ$121.00 per item) and general cataloguing (about NZ$53.00 per item) are not sustainable.[125] The Library might wish annually to analyse the success of its workflow models. By identifying the effort attached to each process it might be possible to establish how individual processes could be streamlined to drive down the per unit cost of collecting, describing, and preserving digital objects so that they are closer to the costs associated with acquiring printed products.

Where web harvesting is seen to be an increasingly central aspect of the digital library group's activities we recommend caution. As we noted above, over the coming two to five years the way web sites deliver information will continue to change. They will become increasingly dynamic with database back-ends and this will make them increasingly difficult to harvest. This is why the Library may wish to investigate the possibilities of a mixed selection and acquisition strategy, some based in part on trawling the web and some based in part on nominated automated deposit (NAD) of information resources. This mixture of push and pull strategies would provide a complementary way to facilitate comprehensiveness of coverage. Nominated deposit mechanisms could be reused or extended to provide tools that could ease the submission of materials by publishers in accordance with the Notice of a Requirement under the Act. (In the latter case, reducing the costs of ingest by setting standards governing the formats that the Library will accept into its repository as it can do with digitised materials would be prudent.) Conceivably the Library may wish to develop mechanisms with publishers to streamline the process by using the Act as a lever to promote the creation

---

1000 items per year. The targets for the selectors are actually low at less than 10 items (which will of course vary in size) per day, but it will be essential to find a balance between the four types of digital objects from which they will need to select.

[124] An assessment of the Australian web archiving costs indicates that they are just too high. The very recent review by Margaret E. Phillips, (2003), *Collecting Australian Online Publications,* Balanced Scorecard Initiative 49, http://pandora.nla.gov.au/BSC49.doc, page 4 sheds much light on their work and points the way to improvements in practice, although there is no evidence that it will reduce the per unit cost. Reducing per unit cost is essential. NLNZ , might perhaps adopt a different management approach by defining an acceptable per unit cost and attempting to streamline the processes to deliver the items within that per unit cost.

[125] NLNZ provided the comparative data.

of procedures and tools that would reduce costs of deposit for publishers and the Library alike.[126]

Recommendations:

The Library should establish a time-constrained Digital Library Delivery Service with responsibility for strategic developments in the areas of selection, acquisition, cataloguing, providing access to and preserving digital materials. (7)

The Library needs to ensure that its collection management strategy is realistic and reflects the technological services that it can deploy.

[126] As we have seen above Part 4 Section 33 of the NLNZ Act 2003 enables the National Librarian to request assistance from '…a publisher of an electronic document to which a requirement relates…' '…to enable the NL to store and use an identical copy of the document.'

Evaluation Metrics and Impact Measurements

The programme must put in place mechanisms to measure its success. These measurements must not be restricted to evaluation of the outputs of the programme itself but need to reflect the impact that the activity has on society at large.[127] At least five types of evaluation are relevant to digital libraries: *front-end*, *formative, summative, iterative,* and *comparative.* However, digital libraries are difficult to evaluate due to their richness, complexity, and variety of uses and users. Efforts made by the US Digital Library Initiative (DLI) have been effective in establishing the need for evaluation of Digital Libraries, in identifying some of the areas most likely to be productive, and in demonstrating the effectiveness of small-scale evaluation efforts. However, they also showed the limitations of current evaluation efforts. Careful thought needs to be brought to bear on defining suitable evaluation metrics and data collection methods (e.g., focus groups, questionnaires).

In addition to mechanisms to evaluate the success of digital library initiatives measuring its impact is essential. Such an approach is supported by the move from output measurements to outcome measurements. These could include assessment of the impact of the developments on tourism, economic growth (e.g., emergence of new industries), and education.

Recommendations:

The Library should put in place mechanisms to evaluate the success of its Digital Library Services Group at all stages of their work, including an evaluation of users, both external and internal, of digital library services at the outset.

Methods for measuring the impact of the digital library services on society need to be agreed.

[127] Comparison between digital service delivery and search room delivery throws up anomalies. Peter Kaufman on Innodata at the April 2003 NINCH meeting on costs of dgitisation reported that 'the Library of Congress receives two million requests per day for digital files, compared with 2 million requests per year for items to be delivered to the search rooms'

Outreach: Dissemination and Community Building

NLNZ is leading digital library development activities in New Zealand. The Library should share its expertise as it develops, but this will require capacity development and an outreach function.[128] For example, the Digital Library Services Group will need to ensure that its guidelines and best practices are widely adopted and used by the broader library community in New Zealand. But there are other ways the Library can disseminate its expertise. For instance the National Preservation Office could play a role in the dissemination of guidelines for digital preservation. There are three areas in which guidelines could benefit the broader library communities:

(a) guidelines governing how digital objects should be created (expanding on the work currently underway in Copying and Digital Services which could be extended to recognise the preservation issues);

(b) guidelines for potential depositors of digital manuscripts on how they could improve the likelihood that the library could accession, catalogue, preserve, and provide access to the materials that they wished to deposit (e.g., formats, documentation); and,

(c) guidelines on handling digital materials that are presented to the library (e.g., the digital manuscripts of authors, politicians, and others).

Dissemination could best be done through online and printed, easy to follow guidelines, workshops, training courses, and advice services.

Recommendations:

The Library should ensure that in developing its digital library activities it exercises its leadership role and supports public and academic libraries in the area of digital preservation. (4)

Consideration should be given to establishing measurable approaches to ensuring the dissemination of expertise including developing the NPO services in this area.

[128] Part 2 Section 9(2) (c) of the National Library of New Zealand Act 2003 mandates that 'The National Librarian must take reasonable steps to—'provide assistance in the development of libraries to institutions and persons in New Zealand…'.

Conclusion

The National Library of New Zealand has a mandate to act to secure New Zealand's digital heritage. A number of Library staff acknowledged that development of digital library services were core to the business of the library and an essential step in the transition from a print based infrastructure to a digital one. Most importantly there was a widely shared sentiment that it was time to act and that the Library, to quote one interviewee, should 'get on and do it, make mistakes, learn as we go.' The key to success will be to manage objectives, establish achievable and measurable targets, not to focus solely on the cost per item or numbers of items collected but on impact measures, and to manage the risks associated with any programme of development through rigorous project management. The Library will need to be transparent about its objectives and should ensure that sufficient information is accessible to the public to facilitate broad understanding of the challenges that NLNZ faces in building a record of New Zealand's documentary heritage in the 21st century.

External activities, such as Project PROBE will enable schools by the end of 2003 to have broadband access. This will transform what educators and learners can do with digital resources.[129] The role of the Library in providing the content to maximise the benefits of that infrastructure will be greatly enhanced by its positive response to the new Act. In developing its digital library services the NLNZ should consider the level of expenditure that it would take to build a physical extension to its existing library. In many ways that is what it will be doing. The Library is conducting digital collection development alongside implementation of the technical infrastructure. In the longer term the technical infrastructure will be akin to the shelving for analogue materials, but in the short-term a certain amount of experimentation and research and development will be necessary.

[129] http://www.minedu.govt.nz/index.cfm?layout=document&documentid=7887 Not only will penetration of services be high in Schools, but computer use at home passed 47% in 2001 and home internet use 37% (*Information Technology Use in New Zealand 2001* [May 2002]). Compare this to the US where in 2002 54% had access to the Internet, Nua Internet Surveys, 2002. It is worth considering these figures in light of projections of growth in demand for bandwidth. K. G. Coffman and A. M. Odlyzko, (2002), 'Growth of the Internet', I. P. Kaminow and T. Li, (eds.), *Optical Fiber Telecommunications IV B: Systems and Impairments*, (Academic Press) 17-56, http://www.dtc.umn.edu/~odlyzko/doc/oft.internet.growth.pdf.

Recommendations

Throughout this report a number of recommendations of detail have been made (such as the need for further consideration of handles vs. DOIs, microfilming, that if the Library develops a selective web harvesting programme that it maintain information about the process including details of sites rejected, the need for server consolidation, the need to add a dynamic OCRing functional layer to *Papers Past*). There are eight key recommendations (and these appear in bold) which are essential if the Library is to ensure the long-term success of its digital library programme.

(1) The Library needs to complete and adopt its digitisation policy within this financial year.

(2) A digital library policy needs to be established by the end of the first quarter of the next financial year. This must include a sketch of the shape and character of the digital library that it is trying to develop.

(3) The Library should consolidate its digitisation activities within Corporate Services and manage these through a cross-disciplinary oversight committee to ensure that resources are effectively allocated and there is consistency to its digitisation work whether this is conducted in-house or outsourced.

(4) The Library should ensure that in developing its digital library activities it exercises its leadership role and supports public and academic libraries in the area of digital preservation.

(5) The Library should complete as a matter of urgency its Survey Objects Project and publicise the results.

(6) The Library should specify, design, implement and deploy a digital repository, if possible in conjunction with other national memory institutions, but if necessary independently of them.

(7) The Library should establish a time-constrained Digital Library Delivery Service with responsibility for strategic developments in the areas of selection, acquisition, cataloguing, providing access to and preserving digital materials.

(8) The Library should improve its economic modelling of the costs associated with collecting and maintaining digital objects whether coming from digital objects on fixed media, digitisation, or web harvesting. This should include consideration of the economic impact of these resources on building the creative economy of New Zealand, encouraging tourism and services to industry.

These recommendations will enable the Library to build on its remarkable successes in laying the foundations for the elements of a digital library that the National Library will need to have in place if it is to respond to the changing information landscape and to fulfil its duties under the new Act.

Appendix 1: Acknowledgements

Staff at the National Library of New Zealand, Archives New Zealand, and Te Papa participated in discussions about digital library developments, needs, and possibilities. Energy, excitement, originality, and team spirit are all terms that describe the support they brought to the process.

I would like to extend special thanks to Catherine Vriens, whose logistical assistance, administrative support, and effective management skills made this review viable.

Finally, Steve Knight (Manager, Digital Library Transition Team, Electronic Services, National Library of New Zealand) who made this report possible through vision, encouragement, and persistent questioning deserves the biggest vote of thanks.

Seamus Ross, 30 June 2003.

Appendix 2: List of Documents Submitted

MNEMONIC	BIBLIOGRAPHIC ENTRY
MS	
MSMP	National Library of New Zealand. 2000. *Metadata Standards Framework for NLNZ.*
MSFP	National Library of New Zealand. 2000. *Metadata Standards Framework for National Library of New Zealand.* http://www.natlib.govt.nz/files/4initiatives_metafw.pdf
MSFW	National Library of New Zealand. 2000. *Metadata Standards Framework for National Library of New Zealand.*
DLP	
DLPO	National Library of New Zealand. 2001. *Digital Library Programme. Programme Overview.*
DLPP	National Library of New Zealand. 2001. *Digital Library Programme. Programme Proposal.*
DLPG	National Library of New Zealand. 2001. *Digital Library Programme. Programme Governance.*
DLPS	National Library of New Zealand. 2001. *DLP Scope Definition.*
DP	
DPFF	National Library of New Zealand. 2001. *Digital Preservation Project. File Format Policy.*
DPHA	National Library of New Zealand. 2001. *Digital Preservation Project. Handover Report.*
DPLI	National Library of New Zealand. 2001. *Digital Preservation Project. Longevity Issues: Migration vs. Emulation*
DPPM	National Library of New Zealand. 2001. *Digital Preservation Project. Preservation Metadata.*
DPON	National Library of New Zealand. 2001. *Naming Digital Objects for Management through Digitisation to the Digital Library.*
DPSL	National Library of New Zealand. 2001. *The Software Library as a Model for Preserving the Functionality of Digital Objects in the National Library of New Zealand.*
PP	
PPWS	National Library of New Zealand. 2001. *Project 251 – Digitisation of Newspapers. Web Specification.*
PPTD	National Library of New Zealand. 2001. *Papers Past Website Technical Documentation*
PPQA	National Library of New Zealand. 2001. *Papers Past Quality Assurance Documentation*
PPPR	National Library of New Zealand. 2000. *Project 251 – Digitisation. Project Proposal.*
PPPL	National Library of New Zealand. 2001. *Project 251 – Digitisation of Newspapers. Project Plan.*

MNEMONIC	BIBLIOGRAPHIC ENTRY
PPRV	National Library of New Zealand. 2001. *Project 251 – Digitisation of Newspapers. Project Review.*
DL/DLI	
DLIA	National Library of New Zealand. 2001. *Digital Library Infrastructure. DLI Architecture.*
DLBS	National Library of New Zealand. 2001. *Software for Digital Library Infrastructure. Business Case.*
DLBC	National Library of New Zealand. 2001. *Hardware & Software for Digital Library Infrastructure. Business Case.*
DLPL	National Library of New Zealand. 2001. *Digital Library Infrastructure. Project Plan.*
DLPR	National Library of New Zealand. 2001. *Digital Library Infrastructure. Project Proposal.*
DLIO	National Library of New Zealand. 2001. *Digital Library Infrastructure. Storage for Digital Objects.*
WK	
WKOF	National Library of New Zealand. 2002. *Digital Workflows: Published Offline*
WKS3	National Library of New Zealand. 2001. *Unpublished Offline Digital Object Workflow.*
WKS2	National Library of New Zealand. 2001. *Published Offline Digital Object Workflow.*
WKS1	National Library of New Zealand. 2001. *Online Digital Object Workflow.*
WKS4	National Library of New Zealand. 2001. *Treat Digital Material Workflow.*
WKON	National Library of New Zealand. 2002. *Workflows: Published online*
WKWK	National Library of New Zealand. 2002. *Unpublished workflow*
PI	
PIPW	National Library of New Zealand. 2002. *A National Library Persistent Identifier Scheme for Digital Objects Workshop Presentation*
PIPP	National Library of New Zealand. 2002. *Presentation to CNRI*
PISA	National Library of New Zealand. 2002. *A National Library Persistent Identifier Scheme for Digital Objects. Draft Proposal.*
PM	
PMEB	National Library of New Zealand. 2002. *Entity Breakdown*
PMDI	National Library of New Zealand. 2002. *Digital Archive Workflow (Draft)*
PMF2	National Library of New Zealand. 2002. *Metadata Standards Framework – Preservation Metadata.* http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta

MNEMONIC	BIBLIOGRAPHIC ENTRY
DLT	
DLTC	National Library of New Zealand. 2002. *Digital Library Transition Team. DLTT Charter*
DLTG	National Library of New Zealand. 2002. *Digital Library Programme. Programme Governance.*
DLTH	National Library of New Zealand. 2002. *Digital Library Programme. Programme Handover.*
DLS	
DLSO	National Library of New Zealand. 2003. *Digital Library Infrastructure. File Structure for Digital Objects.*
WH	
WHTR	National Library of New Zealand. 2002. *Web Harvesting Trial. Interim Report.*
SP	
SPEP	National Library of New Zealand. 2002. *Selection Guidelines for New Zealand and Pacific Online Publications*
SPOV	National Library of New Zealand. 2002. *Acquiring And Making Accessible New Zealand And Pacific Digitally Born Publications*
SPSD	National Library of New Zealand. 2001. *Electronic Information Resources. Report on the suitability of the Voyager to meet the needs of a Selection Database as identified by the E1 Pilot Project*
SPCP	National Library of New Zealand. 1996. *Collections Policy. July 1996; updated December 1997 and June 1999* http://www.natlib.govt.nz/en/about/1keypolcollections.html
SPEW	National Library of New Zealand. 2002. *E-selection file process for digital archive material (third draft)*
OT	
OTNA	National Library of New Zealand. 1965. *The National Library Act.* http://www.natlib.govt.nz/en/about/1keypolnlact.html
OTNB	National Library of New Zealand. 2002. *National Library of New Zealand Bill (Te Puna Mātauranga o Aotearoa Bill) Government Bill* http://www.natlib.govt.nz/files/nlnzbill.pdf
OTCA	1994. *Copyright Act 1994*
OTMR	PriceWaterhouseCoopers. 2002. *National Library of New Zealand. Project Methodology Review*
A	
ACAS	The Next Generation Internet. 2002. *Collaborating at Speed : Innovation Infrastructure for a Knowledge Economy* http://www.internetnz.net.nz/public/working-group-reports/wg-internet2/ngi021007report.pdf ICT Taskforce. 2002. *Breaking through the Barriers. Draft Report.*

MNEMONIC	BIBLIOGRAPHIC ENTRY
AC	
ACAR	National Library of New Zealand. 2002. *Annual Report 2002* http://www.natlib.govt.nz/files/1pubannual.pdf
ACIM	National Library of New Zealand. 2002. *Briefing for the Incoming Minister* http://www.natlib.govt.nz/files/1pubbim.pdf
ACDF	National Library of New Zealand. 2003. *Departmental Forecast Report For The Year Ending 30 June 2003* http://www.natlib.govt.nz/files/dfr.pdf
ACSF	National Library of New Zealand. 2001. *The 21st century: The strategic direction of the National Library of New Zealand Te Puna Mātauranga o Aotearoa: A revised framework for planning.* http://www.natlib.govt.nz/files/framework.pdf
ACMP	National Library of New Zealand. 2001. *Te Kaupapa Mahi Tahi A Plan for Partnership.* http://www.natlib.govt.nz/files/policy-maori.pdf
AC21	National Library of New Zealand. 1998. *Towards the 21st Century: Strategic Plan of the National Library of New Zealand*
ADE	
ADEN	National Library of New Zealand. 2002. *COR Implementation Phase Evaluation. Meeting Notes.*
ADET	National Library of New Zealand. 2002. *COR Implementation Phase Evaluation. Transcript.*
ADER	National Library of New Zealand. 2002. *Discover: Implementation Phase Review and End of Project Review*

Appendix 3: List of Interviewees

National Library of New Zealand Te Puna Mātauranga o Aotearoa

- Penny Carnaby - Chief Executive and National Librarian
- Dave Adams - Team Leader, Copying and Digital Services, Alexander Turnbull Library
- Lynn Benson - Co-ordinator Unpublished Collections, Alexander Turnbull Library
- Keitha Booth - Manager, Information Centre, Corporate Services
- Margaret Calder - Chief Librarian, Alexander Turnbull Library
- Douglas Campbell - Business Development Analyst, Digital Initiatives Unit, Electronic Services
- Brian Cleaver - HeXad Consultant, consultant on various aspects of Digital Library Programme and Project Manager for Turnbull Systems Project
- Graham Coe - Director, Electronic Services
- David Colquhoun - Curator, Manuscripts and Archives, Alexander Turnbull Library
- Jocelyn Cuming - National Preservation Officer, Alexander Turnbull Library
- Mike Dalton - Senior Unix Administrator, Technical Services
- Alison Elliott - Director, Collection Services
- Linda Evans - Curator, Oral History Centre, Alexander Turnbull Library
- Sue Guest - Director, Corporate Services
- Geraldine Howell - Director, School Services
- Graeme Jackett - Manager, Digital Initiatives Unit, Electronic Services
- Adrienne Kebbell - Business Development Analyst, Digital Initiatives Unit, Electronic Services
- Steve Knight - Manager, Digital Library Transition Team, Electronic Services
- Pamela Najar - Manager, Preservation, Alexander Turnbull Library
- Joan McCracken - Reference Librarian, Pictorial Collection, Alexander Turnbull Library
- Jenny McDonald - Manager, Te Puna Support, Electronic Services
- Deirdre McFarland - Team Leader, Acquisitions Team, Bibliographic Services
- Janet McFadden - Service Centre Manager/Acting Manager School Services National, School Services

- Ruth Miller - Customer Support Consultant, Te Puna Support, Electronic Services
- John Mohi - Director, Services to Mäori
- Bronwyn Officer - Conservator Sound Recordings, Alexander Turnbull Library
- Christine Pullar - HeXad Consultant, Project Director for the Digital Library Programme and Turnbull Systems Project
- Philip Rainer - Manager, Research Centre, Alexander Turnbull Library
- Nonnita Rees - Senior Policy Analyst, Policy and Strategic Development
- Emma Roache - Research Librarian, Alexander Turnbull Library
- Sam Searle - Digital Library Projects Leader, Digital Library Transition Team, Electronic Services
- Kirsty Smith - Digital Archivist, Alexander Turnbull Library
- Lockie Stewart - Database Administrator, Technical Services, Electronic Services
- Clark Stiles - Curator, NZ and Pacific Published Collections, Alexander Turnbull Library
- John Sullivan - Curator, Photo Archives, Alexander Turnbull Library
- Dave Thompson - Digital Library Resource Analyst, Digital Library Transition Team, Electronic Services
- Chris Todd - Team Leader, Books Plus Team, Bibliographic Services

Archives New Zealand Te Whare Tohu Tuhituhinga o Aotearoa

- Dianne Macaskill - Chief Executive and Chief Archivist
- Jeremy Cauchi - Group Manager Access Services
- Matthew Hockey - Manager Description
- Michael Hoyle - Group Manager Government Record Keeping
- Jonathan London - Head of Preservation Services
- John Roberts - Group Manager Archives Management
- Adam Stapleton - Team Leader GLADIS Project

Museum of New Zealand Te Papa Tongarewa

- Dr Seddon Bennington - Chief Executive
- Gillian Andreae - Manager Technical Services
- Bridget Popplewell - Business/Systems Analyst

The University of Auckland Library, Te Tumu Herenga[130]

- Janet Copsey - University Librarian

The National Library of Australia

- Jan Fullerton - Director-General
- David Toll - Deputy Director-General
- Pam Gatenby - Assistant Director-General, Collections Management
- Roxanne Missingham - Assistant Director-General, Resource Sharing
- Warwick Cathro - Assistant Director-General, IT
- Jasmine Cameron - Assistant Director-General, Executive and Coordination Support
- Tom Ruthven - Director, Interlending Services and Digitisation Project
- Tony Boston - Director Digital Services

National Archives of Australia

- Dr Stephen Ellis - Acting Assistant Director General, Collection Management Branch
- Andrew Wilson - Assistant Director, Digital Preservation

[130] Telephone interview.

Appendix 4: List of all Recommendations

The Library should complete as a matter of urgency its Survey Objects Project and publicise the results. (5)

The DLTT should work harder to communicate the results of its work within the Library to ensure that the broader curatorial team are aware of it and appreciate its value.

The Library should establish clear acquisition targets for digital materials and create sufficient infrastructure to enable it to achieve those objectives.

The Library needs to ensure that its approaches to the protection of the intellectual property rights in material is reflected in its strategic technology developments as well as in its compliant with relevant legislation.

The Library should combine active selection of materials for acquisition with mechanisms to enable automated deposit of digital materials.

In the process of selecting material for acquisition selectors reject certain materials, perhaps consideration should be given to listing material not selected for acquisition.

Consideration should be given to collecting material whether or not it is already held in print.

The Library should move to consolidate its underlying computer-based storage systems to maximise efficiency and minimise costs.

The Library should ensure that there is a level of distributed redundancy in its systems to ensure that the loss of one location would not put its entire digital library at risk

In specifying, designing and implementing its digital repository the Library needs to consider using products that will interface with its existing public access catalogues.

In specifying, designing and implementing its digital repository the Library should use an open source solution where that solution has achieved a broad user community in the Library as opposed to the techie environment.

That the Library explicitly flag at ingest those materials that it intends to preserve at bit-stream level and those that it will guarantee long term access to at functional level.

The Library should specify, design, implement and deploy a digital repository, if possible in conjunction with other national memory institutions (e.g., Te Papa and Archives New Zealand), but if necessary independently of them. (6)

The Library needs to define what it intends to deliver through its digital library programme and establish a ten year vision for constructing its services.

The Library should continue to work closely with its Library catalogue software provider to ensure that the product develops in ways that will enable the Library to deliver its projected services.

The Library should consider what services it would wish to provide itself and those that it would permit other organisations to deliver as service layers.

That all digital library developments reflect the needs of the communities creating digital materials and the diversity of the backgrounds and needs of potential users.

The development of the Library depends upon definition of the services that the Library intends to provide and those that it could contract out either for financial reward or for free.

The Library needs to complete and adopt its digitisation policy within this financial year. (1)

A digital library policy needs to be established by the end of the first quarter of the next financial year. This must include a sketch of the shape and character of the digital library that it is trying to develop. (2)

The Library should consolidate its digitisation activities within Corporate Services and manage these through a cross-disciplinary oversight committee to ensure that resources are effectively allocated and there is consistency to its digitisation work whether this is conducted in-house or outsourced. (3)

This oversight committee should also consider the relationship between the digitisation programme and the microfilming activities and consider whether or not newer technologies provide suitable security to enable it to discontinue microfilming.

Wherever access to digitised textual sources can benefit from the application of appropriate optical character recognition technologies it should be applied (e.g., *Papers Past*).

The Library should improve its economic modelling of the costs associated with collecting and maintaining digital objects whether coming from digital objects on fixed media, digitisation, or web harvesting. This should include consideration of the economic impact of these resources on building the creative economy of New Zealand, encouraging tourism and services to industry. (8)

The Library should establish a time-constrained Digital Library Delivery Service with responsibility for strategic developments in the areas of selection, acquisition, cataloguing, providing access to and preserving digital materials. (7)

The Library needs to ensure that its collection management strategy is realistic and reflects the technological services that it can deploy.

The Library should put in place mechanisms to evaluate the success of its Digital Library Services Group at all stages of their work, including an evaluation of users, both external and internal, of digital library services at the outset.

Methods for measuring the impact of the digital library services on society need to be agreed.

The Library should ensure that in developing its digital library activities it exercises its leadership role and supports public and academic libraries in the area of digital preservation. (4)

Consideration should be given to establishing measurable approaches to ensuring the dissemination of expertise including developing the NPO services in this area.

Appendix 5: Terms of Reference

NATIONAL LIBRARY OF NEW ZEALAND
DIGITAL LIBRARY DEVELOPMENT REVIEW
Terms of Reference

Purpose of digital library activities

The National Library of New Zealand is currently developing its approach to the management of electronic material through the development of a digital library for both digital originals and digital objects created through the Library's digitisation programme. Through this work it has become clear that preservation of digital materials will be a significant new business requirement within the Library.

NLNZ is committed to the establishment of a digital library and to the management and preservation of its contents. The digital library will enhance access to the Library's digital resources for all New Zealanders now and in the future and is necessary if the Library is to achieve its mandate *'to collect, preserve and make available recorded knowledge'* in an environment increasingly characterised by electronic resources, both offline and online.

The work undertaken within NLNZ reflects international research and development in this area where there is as yet no majority view. It is to a large extent being developed in the light of current best practice.

The Library is aware that the nature of its response to the requirements of digital objects will evolve over time and that this may necessitate changes to business processes and user attitudes and may also involve considerable cost. As a result of this the Library wishes to be assured that the approach it has taken to development of the capability needed to meet these requirements to date is consistent with mainstream thinking and capable of being extended through to a successful conclusion for the Library.

Consequently, the Library is requesting an external review of its digital library activities to date with a view to either validating current progress or highlighting areas of deficiency and to provide a base document from which to continue implementation through to successful incorporation of business processes related to collection of digital material into the Library's core business.

Background

The National Library of New Zealand (NLNZ), in common with other cultural institutions world-wide, is undergoing a period of intense change in the context of rapidly developing information technologies which are radically transforming collections and services.

The Library's work on digital library activities has been a key component in responding to this changed environment and reflects an acceptance of the necessity for a range of initiatives to be undertaken if the goal of long term storage and preservation of digital objects, to meet the future needs of national and international users of that information, is to be achieved.

The Library's core services are supported through two primary collection management systems. These systems will continue to provide the resource discovery metadata for electronic material. The published collections are managed through a MARC based Integrated Library System (ILS) and the unpublished systems are managed through an ISAD-G based system.

Elements of the unpublished collections have been digitised and are stored in and accessible through a stand-alone system called Timeframes (using Hyperwave), from which order processing of copies is managed[131].

Need for a digital library

The National Library of New Zealand has a legislative mandate 'to collect, preserve and make available recorded knowledge, particularly that relating to New Zealand ,to supplement and further the work of other libraries in New Zealand, and to enrich the cultural and economic life of New Zealand and its cultural interchanges with other nations'.

The Library's mid-term strategic planning document *the 21st Century* reflects the need for the Library to take account of new developments relating to the collection and accessibility of digital material.

The Library is currently addressing these requirements through a series of projects related to:

- collection and management of electronic material
- development of a Digital Library
- enhanced access to its collections via digitisation.

This work has been undertaken in the light of similar initiatives internationally. However, what is different in the National Library of New Zealand's approach is the attempt to develop holistic end-to-end processes for incorporation of business processes related to collection of digital material into the Library's core business. This approach has been dictated by the need to deal with digital objects now as a collecting agency and also the requirement to implement digital library activities within current baseline funding.

---

[131] Access to all the Library's online services (excluding those that are subscription only) is available from the
Library's web site at http://www.natlib.govt.nz/

The work programme currently underway within NLNZ includes the following activities:

- development and implementation of business process workflows for incorporating digital objects into the Library's business processes, eg selection, acquisition, care and handling, transformation of physical originals
- development of infrastructure for digital material, eg upload process, storage, authentication, access
- purchase and implementation of a metadata repository for provision of portal services to NLNZ applications
- creating a testbed application for researching and implementing a range of digital library activities, eg resource discovery metadata, preservation metadata (schema, extraction, storage), persistent identifiers
- evaluate feasibility of and pilot web harvesting for the capture and preservation of New Zealand based and related web sites
- delivery of a sustainable 'production line' process for bulk digitisation enabling ongoing large scale digitisation similar to the numbers being delivered by Project 251 (ie 251,000 images).

Expected deliverables

By putting this work out for review we hope to gain further insight into the current status of the Library's digital library activities, gain a benchmark for ongoing development and consequently strengthen the eventual outcomes for digital library development.

Outcome of the review will be a formal report presented to the National Library of New Zealand. The review should provide comment and recommendations on:

1. the basic assumptions of the Library's approach to building the digital library infrastructure and business processes
2. the validity or otherwise of the Library's approach
3. the completeness or otherwise of the Library's approach
4. the conformity of the Library's approach with all relevant emerging international standards and trends in this area
5. the current status of digital library activities within the Library, ie where the Library is up to in the development and implementation of a digital library
6. the status of the Digital Library Transition Advisory Committee and the Digital Library Transition Team
7. recommendations for future activities (including an estimate of the funding required) to ensure successful incorporation of business processes related to collection of digital material into the Library's core business.

Approach

The following approach will be followed:

1. A package of material providing the background necessary to put the Library's digital library activities in context will be prepared.
2. Terms of Reference will be emailed, supporting material couriered, to the reviewer.
3. Reviewer will be situated on site for a portion of the review period.
4. Interviews will be arranged by the Library with selected staff
5. Reviewer will provide a draft report to the Library
6. Reviewer will provide a final report to the Library.

Confidentially

Material provided by the Library and conclusions drawn from that material are to be kept confidential. Information can only be released with the permission of the Library.

Timeframes

The following timetable is envisaged for the review process:

1. Draft Terms of Reference emailed to reviewer – 14 October
2. Final Terms of reference emailed to reviewer – 18 October
3. Supporting material couriered to reviewer – 18 October
4. Reviewer onsite research and interviews – 9-20 December
5. Draft report presented – 17 January 2003
6. Final report presented – 31 January 2003.

Contact information

Contacts are:

Management of the review:

Steve Knight, Digital Library Transition Co-ordinator

0064 4 4743142, steve.knight@natlib.govt.nz