

Digitisation, Knowledge and Metadata: The National Library of New Zealand as a Case Study

Steve Knight

ABSTRACT

This paper describes the National Library of New Zealand's development of a Metadata Standards Framework to guide its cataloguing/indexing/description activities as it increases its presence online. The National Library is one of the key organisations in New Zealand's culture and heritage sector, and the rapidly developing communications and technology environment has required the Library to establish a position as a prime provider of cultural and heritage materials online. Central to the ongoing viability of the Library's online initiatives is adherence to standards and ongoing compliance with interconnectivity and interoperability requirements. This paper describes the initial goal of the Library with respect to metadata, its growing appreciation of the breadth and depth of metadata initiatives worldwide and its decision to focus on resource discovery metadata as the initial output of a continually evolving Metadata Standards Framework.

Introduction

Over the last five years the National Library of New Zealand (NLNZ) has developed substantial expertise in the digitisation of a range of material and the development of online applications for the presentation of those materials.¹ The Library is also currently working on development of resources on the Ngati Toa leader Te Rauparaha, the Anglican Rangiatea Church destroyed by fire at Otaki in 1995 and on the digitisation of nineteenth century New Zealand newspapers.

Early in 2000, the Library began to look at its online applications and the implications for further initiatives of international developments in the indexing and cataloguing of online resources, now commonly referred to as metadata. A team of Library staff who represented the key interested business units within the Library (Collection Services, Alexander Turnbull Library, School Services and the Digital Initiatives Unit of Electronic Services) were tasked with researching the metadata standards environment and developing a framework for the use of metadata standards within NLNZ. This work resulted in the release, in October 2000, of the *Metadata Standards Framework for National Library of New Zealand* (National Library of New Zealand, 2000).

Background

Digitisation and presentation via the web have become major issues for cultural institutions looking to increase accessibility to the vast range of materials under their control. The mechanics of

digitisation for cultural materials have become highly developed over the last few years. However, it is becoming increasingly clear that developing and placing on the web single, discrete, stand-alone applications does not necessarily best serve the full range of potential users of our materials. There is an increasing need to ensure that those materials are discoverable and sharable not only as discrete applications but also within and across all of the user communities that might have an interest in that material. Work on the development of metadata standards is a significant contribution to understanding how this interoperability and cross-pollination of cultural materials online can be achieved.

The development of metadata standards for online delivery has a strong analogy in the development of English as the *lingua franca* for international air traffic control. There the need is obvious but it is a similar need that is driving international work on metadata standards. If interconnectivity and interoperability are to lead to 'meaningful' access to electronic resources, what is the 'descriptive' *lingua franca* that will make this achievable? How do we create seamless access to electronic resources irrespective of where they are located, or what server hardware they are loaded on, or what operating/software systems they use? This is not a hermetic exercise in information management. Access to information in an increasingly online environment is now being seen as key to the development of a knowledge economy, and, possibly, a knowledge society.

However, even within the online world, access to information is becoming increasingly difficult. More and more information is available online (some of it only online) but less than fifty per cent of that information is currently accessible through commercial search engines. That percentage is substantially higher in the case of richer, more structured (e.g. database driven) information sources. It is only this year that the *Google* search engine has developed the capability to index Portable Document Format (PDF) documents although PDF has long been the major markup language (outside of HTML) for web documents.

Although bibliographic databases, particularly within the library community, have long had a highly developed underlying structure in the MARC format (Machine Readable Cataloguing), making online resources meaningfully retrievable has been the impetus behind the development of a new range of descriptive standards. While information providers are rapidly adopting these new standards for interoperability, the commercial search engines are currently incapable of reading and indexing documents using these markup standards. We may note the significance of the State Services Commission's E-govt unit's development of a Dublin Core based metadata standard for the provision of government products and services online (State Services Commission, 2001).

Although outside the scope of this paper, issues of interoperability and access will require cooperation and partnerships, both nationally and internationally, to maximise the benefits of online information. In this context the longer-term objectives within New Zealand might include:

- promulgation of national metadata standards for creators/providers of online content;
- establishment of an umbrella body tasked with monitoring and advising on interoperability and access issues in the New Zealand context;
- monitoring the appropriateness of international developments within a New Zealand context.

Standards

Standards are "documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that materials, products, processes and services are fit for their purpose" (International Organization for Standardization, 2001). Metadata standards will be crucial for interoperability in an increasingly online environment and the Library's Metadata Standards Framework documents the Library's commitment to particular resource discovery standards for its online initiatives. Interoperability is

defined for the framework as "the ability of software, databases and hardware on different machines from different vendors to share data meaningfully". The term 'meaningfully' was a key component of the project's deliberations. The sharing of data is of little consequence if it does not maximise the user's ability to make sense of that data.

This interest in metadata and interoperability mirrors developments in the Library's technology environment. The Library's Information Systems Strategic Plan, released in early 1999, included in its statement of scope the need "to provide for a cost-effective, technologically advanced IT environment to allow simple, uniform access to the Library's information" (National Library of New Zealand, 1999: 11). The Metadata Standards Framework distinguishes between interoperability within NLNZ, i.e. for integrated access to the Library's resources and interoperability externally, as with non-NLNZ systems. This distinction has implementation implications for the Library. Integrated access might require an architecture that enables access to the full range of systems within NLNZ or records might be duplicated across multiple systems. For example, description of a collection might exist in both an Encoded Archival Description (EAD) (Library of Congress, 2001) context with pointers to associated digital objects, and within a MARC based system with pointers to associated digital objects. This last example has further implications for synchronisation of data and requires careful consideration of where the 'primary' record resides. It also needs to be recognised that adherence to standards does not in itself guarantee interoperability as standards can be implemented in different ways with consequent interoperability issues although, of course, these are not as substantial as where different standards are being used. It is important that NLNZ's implementation is in line with international best practice. It could also be argued that the New Zealand community is of sufficient size to allow for consistency in the use of implementation guidelines for describing online materials and that NLNZ might have a role in the promulgation of these standards within the wider New Zealand information community.

Substantial efforts are now being made both within NLNZ and internationally to develop implementation guidelines for metadata standards to ensure consistency of usage and enhance search and retrieval quality across multiple applications, and across multiple user communities (e.g. libraries, archives, museums) using the same standard (Research Libraries Group, 1998; Consortium for the Computer Interchange of Museum Information, 1999; Lagoze, 2001). In the implementation of multiple metadata, a possible hierarchy of metadata standards using Resource Description Framework (RDF) and Extensible Markup Language (XML) may be applied to mark up documents using Dublin Core (DC) and Encoded Archival Description (EAD).

Which standards

A standards based approach allows:

1. interoperability and interaction with other systems and services outside the boundaries of traditional libraries;
2. new kinds of services and products to be defined;
3. communities of interest and expertise to leverage their knowledge and expertise whilst sharing costs and reducing risk;
4. standardised migration paths for digital objects and their descriptions;
5. streamlined management of both traditional print and paper, and digital resources.

Metadata standards are evolving within a range of standards,² industry³ and interest groups⁴ exploring the different management requirements of different digital objects (with particular emphasis on online resources). Each of these bodies has its own particular, specific ratification processes. The framework needed to decide not only which standards to embrace but also the ratification level to be used, as defined by the particular standards authority. This decision was

generally made at a conservative level with the key criteria being that while the standard may not yet have reached its highest certification level within its proposing body, the likelihood of substantial change is minimal.

Principles

The first step was to elucidate the principles that would direct the Library's approach to the implementation of metadata standards. The following principles make an explicit commitment on the part of the Library to adopt and adhere to standards. Any deviation must be carefully considered, justified and well documented. The principles are:

1. electronic access - the Library will make access available electronically to the entire descriptive record and the digital object wherever possible. This is not as straightforward as it may seem. For example, there is an ongoing debate within the Library as to whether digitisation and the presentation of digital objects online is publication and thus usurping the Library's role of providing source material to researchers who then develop and publish the nation's narratives.
2. integrated access - the Library will provide unified access to materials within NLNZ across disparate data types and databases. As noted above this has been a strategic driver for the Library's Electronic Services Directorate for almost three years and has as its end as homogeneous as possible a technical infrastructure, providing maximum interconnectivity and interoperability to support simplified access to the Library's systems and collections.
3. interoperability - the Library will comply with international interoperability standards for the description of its collections and electronic access to them.
4. development - where a standard does not support a necessary business requirement the Library will work for the development of the standard to incorporate that requirement. The Library wants to be involved in the development of standards for the international information management community and is currently investigating how this can best be affected.
5. gaps in standards - *where there is no standard or an inadequate standard to support a necessary business requirement, the system implemented must be convertible in time to an international standard.* Where there is a gap in the standards environment the Library will ensure that application development is undertaken according to current best practice and with a view to possible standards compliance requirements in the future.
6. permitted standards - *the standards framework will define the standards to be used and their level of release.* The Library has taken a conservative approach to deciding which standards it will embrace and at what level of ratification within the proposing body. This will be discussed further below.
7. preservation - *ongoing storage and preservation of electronic material will conform to international standards.* The Library's standards framework was deliberately focused on metadata standards for resource discovery. It is currently undertaking an analysis of, and practical application of, existing preservation metadata standards in the light of international activities in this area, which could see adoption of a single schema for preservation metadata.⁵
8. language - *consistent standards and policies will be applied within and across databases where appropriate to facilitate access to all National Library databases.* This is becoming increasingly important in the online environment as compliance with a metadata standard does not of itself ensure interoperability between systems using that standard.

Resource discovery

It became clear early in the process that we would need to limit the focus of the project. This was due to two main factors. Firstly, there is an increasingly large range of online activities that are going to require their own metadata, e.g. preservation and rights management. Secondly, the metadata requirements for a number of these activities are still in development. As a result of this the Library decided to concentrate its efforts on resource discovery metadata as the most advanced and also as the most relevant to the Library at that time.

The resulting document, the *Metadata Standards framework for National Library of New Zealand*, is therefore the first phase of a dynamic and evolving metadata standards framework for the Library. It defines and describes the core descriptive metadata standards to be used by the Library for 'resource discovery' (information is locatable and receivable by users) across all media and for all the Library's collections.

Types of metadata

As noted above the project concentrated its efforts on definition of resource discovery metadata standards due to both the range of online products and services requiring different strands of metadata and because of the emerging nature of the standards concerning these different strands. There are a number of methods for categorising metadata standards depending on their purpose. For example, the Library of Congress differentiates between descriptive, administrative and structural metadata. (Library of Congress, 2000). The framework for NLNZ was developed in line with the taxonomy described in Kenney and Rieger's *Moving Theory into Practice* (2000, esp. Ch. 5). This taxonomy describes four key metadata categories for digital objects. These are:

1. resource discovery - how do we ensure that the materials we have collected are locatable and retrievable by our clients? Dublin Core is an early and still evolving attempt to provide a *lingua franca* for resource discovery in an online environment.
2. structural - how do we present our objects in context and not just as a bunch of files (e.g. the pages of a digitised book) and how do we navigate within this context (e.g. page 1 to page 2)?
3. rights management and access control - how do we ensure protection of intellectual property rights, authentication of clients and authorisation of clients to access online objects?
4. technical and administrative - what are the essential attributes of digital objects and the processes and technologies that created them which are required for long term storage, management, preservation and access?

Current position

The framework currently consists of:

1. standards for expressing and formatting data in an online environment;
2. metadata standards for use in NLNZ;
3. a usage map illustrating the possible combinations of metadata standards for specific material types and audiences;
4. content standards and resources for describing specific material types;
5. content standards and resources for Library-wide data elements.

With the framework, as it stands at present, concentrating on metadata standards for resource discovery, ongoing work will need to be undertaken to develop the Library's position on metadata standards for the other three metadata categories described in the taxonomy. This will include:

1. continuing work on preservation metadata as outlined above;
2. investigation of X.500,⁶ Lightweight Directory Access Protocol (LDAP),⁷ and vCard⁸ for directories;
3. investigation of the components of unique identifier systems, e.g. Uniform Resource Identifiers (URI),⁹ which consist of Uniform Resource Names (URN), Universal Resource Locators (URL) and Uniform Resource Characteristics (URC), Persistent Uniform Resource Locators (PURL),¹⁰ Digital Object Identifiers (DOI),¹¹ and the Handle System;¹²
4. continued monitoring of the development of the Z39.50¹³ protocol for sharing of information between remote and disparate systems, a key element in interconnectivity and interoperability for libraries;
5. implementation of the Unicode¹⁴ character set providing support for the Maori macron and the Pacific 'ayn';
6. monitoring of the consolidation of the data elements within ISO 8459¹⁵ into a single object model;
7. investigation of privacy (e.g. P3P),¹⁶ ratings services (e.g. PICS)¹⁷ and rights management (e.g. INDECS);¹⁸
8. continuing work on controlled vocabularies, e.g. thesauri¹⁹ and topic maps.²⁰

Implementation

Having made the decision to adopt specific metadata standards for the Library it is by no means a trivial exercise to actually implement this decision and reflect it in the business processes of the Library. There are ongoing issues for the Library in terms of implementation including:

1. the processes to incorporate standards (e.g. authority control) across the Library's systems as appropriate, e.g. Library of Congress Subject Headings (LCSH);
2. practical merging of multiple resources for the same thing, e.g. a consolidated iwi/hapu file;
3. relationships with standards providers, e.g. Library of Congress.

This last could become a key issue due to the in-built tension between the requirements of interoperability and international standards and the desire to appropriately reflect the 'New Zealandness' of the Library's collections, for example, in the use of subject authorities. This is a particularly important issue within NLNZ where the unpublished collections of the Alexander Turnbull Library make up one of New Zealand's richest institutional holdings of unpublished material relating to New Zealand and the Pacific. The users of these collections are predominantly New Zealanders pursuing specialist New Zealand studies research.

This leads to the possibility that compliance with international standards may not allow optimal description of these specialist materials especially in cases where:

1. a standard does not cover, or readily adapt to, New Zealand terms and usage;
2. a standard compromises the location and presentation of information in ways that clients require;
3. a standard does not contain or allow all the data elements required for the needs of collection management.

In this instance the question becomes one of to what extent, if any, will compliance to international standards allow for local variations in the use of any given standard. While this is still an issue being debated within the Library, the experience of the British Library in this regard is instructive. The British Library affirmed its commitment to LCSH, in 1995, but not without some reservation (MacEwan, 1998).²¹

Inevitably we had an initial concern about Americanisms. The idea of an American subject indexing language forming the basis for access to the collections of the British national library certainly seems to be a triumph for American cultural imperialism. And perhaps it is. But in terms of giving us problems with access I think it is a trivial issue, and certainly one which can largely be overcome through developing thesaural control on the authority file. Many cross-references from Americanisms already exist to provide access in language more natural to a British user ...

Conclusion

The Framework is the first fruit of the Library's ongoing commitment to the use of metadata standards in accordance with international best practice. Confirmation of the Library's direction was affirmed through a substantial quality assurance process with the draft document being sent to interested parties before being released. Those parties included the National Archives of New Zealand, the National Archives of Australia, the British Library, Cornell University, the National Library of Canada and the Library of Congress, the Getty Institute and the University of California at Los Angeles.

The Library recognises the rapidly shifting nature of the metadata environment and the wide range of metadata and content standards available. As a living, dynamic document, the Framework will continue to be developed, whether by taking on new standards and/or resources for use in the future or through discarding those that have been previously adopted in the framework.

Notes

1. Timeframes. <http://timeframes.natlib.govt.nz>
Te Waimano -Waters of the Millennium. <http://tewaimano.natlib.govt.nz>,
Ranfurly Collection <http://tepuna.natlib.govt.nz/abouttp/abranfurly/about.html>,
Kilbirnie-Lyall Bay Community Oral History Project. <http://tepuna.natlib.govt.nz/abouttp/ablkilbirnie/about.html>
2. International Organization for Standardization (ISO). <http://www.iso.ch/>
National Information Standards Organization (NISO). <http://www.niso.org/>
Australia/New Zealand Technical Committee IT/19. <http://www.nla.gov.au/lis/stndrds/grps/it19/>
3. World Wide Web Consortium (W3C). <http://www.w3.org/>
4. International Council on Archives (ICA). <http://www.ica.org/>
5. National Library of Australia. (1999). Preservation Metadata for Digital Collections. Exposure draft. <http://www.nla.gov.au/preserve/pmeta.html>
The CEDARS Project. (2000). Metadata for Digital Preservation. The CEDARS Project Outline Specification Draft for Public Consultation. <http://www.leeds.ac.uk/cedars/MD-STR%7E5.pdf> or <http://www.leeds.ac.uk/uldcedars/documents/Metadata/cedars.html> (HTML)
Research Libraries Group. (1998). RLG Working Group on Preservation Issues of Metadata. Final Report, May 1998. <http://www.rlg.org/preserv/presmeta.html>
Online Computer Library Centre & Research Libraries Group. (2001). Preservation Metadata for Digital Objects: A Review of the State of the Art. A White Paper by the OCLC/RLG Working Group on Preservation Metadata. http://www.oclc.org/digitalpreservation/presmeta_wp.pdf

6. International Telecommunication Union. (2001). X.500 Recommendation. <http://www.itu.int/itudoc/itu-t/rec/x/x500up/x500.html>
7. Internet Engineering Task Force. (1997). Lightweight Directory Access Protocol (v3). <http://www.ietf.org/rfc/rfc2251>
8. Internet Engineering Task Force. (1998). vCard MIME Directory Profile. <http://www.ietf.org/rfc/rfc2426>
9. Internet Engineering Task Force. (1998). Uniform Resource Indicators (URI): Generic Syntax. <http://www.ietf.org/rfc/rfc2396>
10. PURL- Persistent Uniform Resource Locator. <http://purl.oclc.org/>
11. The Digital Object Identifier System. <http://www.doi.org/>
12. The Handle System. <http://IWWf.handle.net/>
13. Z39.50 International Standard Maintenance Agency. <http://lcweb.loc.gov/z3950/agency/>
14. Unicode. <http://www.unicode.org/>
15. National Information Standards Organization. (1988). ISO 8459. Bibliographic Data Element Directory. <http://www.niso.org/sc4stand.html#8459-1>
16. World Wide Web Consortium - W3C. (2000). The Platform for Privacy Preferences 1.0 (p#pl.0) Specification. <http://www.w3.org/TR/2000/WD-P3P-20000915/>
17. World Wide Web Consortium-W3C. (1996). Rating Services and Rating Systems (and Their Machine Readable Descriptions) Version 1.1. <http://www.w3.org/TR/REC-PICS-services-961031>
18. The INDECS Project. <http://www.indecs.org/project.htm#finalDocs>
19. National Information Standards Organization. (1999). Workshop on Electronic Thesauri: Planning for a Standard. <http://www.niso.org/thes99rprt.html>
20. International Organization for Standards & International Electro technical Commission. (1999). ISO/IEC FCD 13250: 1999 - Topic Maps. <http://www.ornl.gov/sgml/sc34/document/0058.htm>
21. MacEwan, A (1998). Working with LCSH: the cost of cooperation and the achievement of access. A perspective from the British Library. A paper presented to the 64th IFLA General Meeting, August 16- August 21, 1998. <http://www.ifla.org/N/ifla64/033-99e.htm>

References

- Consortium for the Computer Interchange of Museum Information. (1999). Guide to Best Practice: Dublin Core. http://www.cimi.org/public_docs/meta_bestprac_v1_1_210400.pdf
- International Organization for Standardization. (2001). What are standards? <http://www.iso.ch/iso/en/aboutiso/introduction/index.html>
- Kenney, A. and Rieger, O. (2000). *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (Chapter 5, pp. 84-100). Mountain View, California: Research Libraries Group.
- Lagoze, C. (2001). Keeping Dublin Core Simple: Cross-domain discovery or resource description? *D-Lib Magazine*, 7(1), January. <http://www.dlib.org/dlib/january01/lagoze/01lagoze.html>
- Library of Congress. (2000). Library of Congress Digital Repository Development. Core metadata elements. <http://lcweb.loc.gov/standards/metadata.html>
- Library of Congress. (2001). Encoded Archival Description (EAD): Official Web Site. <http://www.loc.gov/ead/>
- National Library of New Zealand. (1999). Information Systems Strategic Plan. Wellington: National Library of New Zealand.
- National Library of New Zealand. (2000). Metadata Standards Framework for the National Library of New Zealand. <http://www.natlib.govt.nz/en/whatsnew/nlnz.html#metadata>
- National Library of New Zealand. Kebbell, A. (2000). Using Multiple Metadata Sets to Advantage. A paper presented to the LIANZA 2000 Conference 'Exceeding Expectations', Christchurch, October 15 - 18, 2000, Rollitt, K. E-government and Metadata and Knight, S. Metadata Standards Framework. <http://www.conference.co.nz/LIANZA2000/papers.html>
- Research Libraries Group. (1998). RLG Recommended Application Guidelines for EAD. <http://www.rlg.org/rlgead/guidelines.html#o1>
- State Services Commission. (2001). The New Zealand Government Locator Service (NZGLS) Metadata Standard and Reference Manual VI. 0. <http://www.e-government.govt.nz/guidelines/wordmetadata/html.html>