Preservation Metadata: National Library of New Zealand Experience

Steve Knight

Abstract

Development of approaches to preservation metadata has been an integral component of international efforts in the field of digital preservation. The focus of the community engaged in this work is currently shifting, and there is, as yet, no formal agreement around a conceptual framework and identification of required data elements. At the same time attention is now turning to the more complex task of building sustainable technical, infrastructure, and policy frameworks that will enable organizations to implement preservation metadata strategies practically at a local level.

The National Library of New Zealand, Te Puna Mātauranga o Aotearoa, has been actively engaged in work on preservation metadata. This has involved development of a preservation metadata schema, a more granular implementation-ready data model/XML schema, a software application for programmatically extracting preservation metadata, and finally a repository for storing the gathered preservation metadata. This article contextualizes the National Library of New Zealand experience by discussing the purpose of preservation metadata and the ways that organizations may use this type of information in the future to support their long-term goal of preserving digital assets in perpetuity.

Introduction

It is not entirely clear where the phrase "preservation metadata" was coined. In their seminal 1996 report, Donald Waters and John Garrett noted that "metadata, which refers to information about information, is sometimes used as a generic term for systems of reference" and that "the preference for the term metadata . . . appears to flow from the felt need to emphasize the

LIBRARY TRENDS, Vol. 54, No. 1, Summer 2005 ("Digital Preservation: Finding Balance," edited by Deborah Woodyard-Robinson), pp. 91–110

special referential features needed in the digital environment and to distinguish those special features from those of more traditional systems of citation, description and classification" (Garrett & Waters, 1996, p. 47). A year later, in 1997, Lorcan Dempsey and Rachel Heery described a situation

> where a digital representation of the file exists, physical characteristics of the representation (file size, format, information documenting the capture process, etc.) will reside in the header of the digital representation file, or if it is maintained separately, in a separate metadata format and syntax (e.g. a digital representation of a letter written by Mark Twain; with separate physical characteristics and capture information on each page-image). (Dempsey & Heery, 1997, p. 30)

Also in 1997, Michael Day posed the question whether "as the archives community are seriously considering using metadata to ensure the integrity and longevity of records, it might be useful to investigate whether a similar approach would be useful for digital preservation in a library context—and in particular for networked documents" (Day, 1997). In 1998 the same author, under the auspices of the CEDARS project (Curl Exemplars in Digital Archives), began to answer that question by producing "a review of metadata formats and initiatives in the specific area of digital preservation," in which he notes "a growing awareness that metadata has an important role in digital resource management, including preservation" (Day, 1998). So within two years the notion of preservation metadata went from obscurity to center stage in the digital preservation work plan, where it has remained for the last six years or so.

This article describes the work undertaken by the National Library of New Zealand Te Puna Mātauranga o Aotearoa (the Library) in this context of international developments in the preservation metadata arena. In addition, it both answers some questions regarding our ability to deal with an "uncontrollable and unmanageable flood" (University of Heidelberg, 2005) of digital materials through a series of pragmatic, staged steps (Thompson & Searle, 2003) and asks some questions about the development of an international approach regarding preservation metadata and why it has taken so long to arrive at a consensus.

Organizational Context

In 2003 the Library's governing legislation was revised with the passing of the National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003 (New Zealand Government, 2003). The act defines the purpose of the Library as "to enrich the cultural and economic life of New Zealand and its interchanges with other nations by . . . collecting, preserving, and protecting documents, particularly those relating to New Zealand, and making them accessible for all the people of New Zealand, in a manner consistent with their status as documentary heritage and taonga."[1]

The act also, for the first time, provides the Library with the mandate to engage fully with digital material, both online and offline, and to ensure that we accord digital material the same degree of responsibility and care we show our nondigital collections. Part 4 Section 29(1) defines an electronic document as "a public document in which information is stored or displayed by means of an electronic recording device, computer or other electronic medium, and includes an Internet document," which is further defined as "a public document that is published on the Internet, whether or not there is any restriction on access to the document; and includes the whole or part of a website" (New Zealand Government, 2003, p. 14). A public document is also defined elsewhere within the act.

It is within this context that the Library is undertaking a program of linked initiatives to ensure the incorporation of digital material into the Library's core business processes with a view to the long-term accessibility of those resources. The goal of the program is to develop holistic, end-to-end processes for the handling of digital material. The program includes the following activities:

- Developing and implementing business process work flows for incorporating digital objects into the Library's business processes; for example, selection, acquisition, care and handling, and transformation of digital originals
- Developing infrastructure for digital materials; for example storage, authentication, and access
- Researching and implementing "components" of the digital archive; for example, preservation metadata (schema, data model, extraction, storage) and persistent identifiers
- Implementing Web archiving for the capture and preservation of New Zealand-based and related Web sites
- Implementing a portal service for provision of access to all the Library's applications

The progress of the Library to date has thrown up a number of major areas of need that will require continued attention if the Library is to successfully confront the challenge of digital preservation. These include

- Recognition that while information in all formats is still increasing, more and more is being produced digitally and the gap between digital and print production is constantly increasing
- Engagement with the wider information community will become increasingly important as it is unlikely that any one organization is going to be able to do it all
- The need for allocation/reallocation of resources to digital preservation and developing the appropriate skill base

- Ensuring that we have the necessary technology infrastructure, including redundancy
- Development of appropriate strategies, policies, processes, and procedures
- Ensuring that our selection, acquisition, and description processes are in sync with the requirements of digital preservation

Preservation Metadata

The Library's work on metadata began in 2000 and was based on the taxonomy described in Anne Kenney and Oya Reiger's *Moving Theory into Practice: Digital Imaging for Libraries and Archives* (2000): resource discovery, structural, rights management and access control, technical and administrative. Initial work concentrated on metadata for resource discovery (National Library of New Zealand, 2000) and described the core descriptive metadata standards to be used by the Library for resource discovery across all media and for all the Library's collections.

Schema

The Library released the first version of a logical model for preservation metadata online in November 2002 (National Library of New Zealand, 2002), with a revised version incorporating learning since the original version being made available in June 2003 (National Library of New Zealand, 2003c). As is usual in these types of endeavors, the Library's efforts built on progress already made by earlier initiatives—for example, the work undertaken by the National Library of Australia (1999), the CEDARS program (Cedars Project, 2002), Online Computer Library Center/Research Libraries Group (OCLC/RLG) activities (OCLC, 2003a), and the shared language provided by the Open Archival Information System (OAIS) Reference Model (Consultative Committee for Space Data Systems, 2002)—but with a view to practical implementation. We have attempted to minimize the degree of overlap with other metadata and focused on that metadata necessary for preservation, including the notion that the preservation metadata record itself is an integral part of the preservation process. The Library's schema is not regarded as fixed. It is our current iteration of a minimum set of metadata for digital preservation, and it is expected that it will change over time as the requirements for preservation metadata become clearer.

Data Model

The Library then developed a data model to inform the implementation of the schema (National Library of New Zealand, 2003b) along with an XML schema version of the data model (National Library of New Zealand, 2003a). The data model extends the schema into an implementable framework increasing the granularity of the schema.

Repository

A repository for preservation metadata is currently being developed. It is expected that it will integrate with our existing metadata systems, creating a comprehensive metadata framework for resource discovery, preservation, rights, etc.

Extract Script

In parallel with the work on the schema and the data model, the Library has developed a tool to automatically extract metadata from the headers of a range of file types. Automation is essential to the success of any preservation metadata strategy given the number of file types we have to grapple with and the complexity of the associated metadata.

The script produces an initial XML output of everything available in the header of the file. An XSL style sheet transformation is then applied to produce an output of that metadata identified as important to preservation. This is then uploaded to the metadata repository. The script has a flexible modular architecture to allow the addition of adapters for new file types and for the fine-tuning of the XML output as required. The extract tool will be discussed in more depth below.

Organizational Concerns

While it may be reasonably clear what the organizational impact of digital preservation might be, there are still significant concerns as to how a sustainable outcome will be achieved in this arena. For the Library this includes the following:

- The low-level awareness of the need for digital preservation within the community of "memory institutions" and more widely
- The lack of metrics regarding the scope of the challenge
- The lack of skill sets for implementing digital preservation; for example, the multiplicity of software involved and digital conservation/archaeology
- The lack of agreed international approaches to digital preservation
- The lack of practical models to match the high-level conceptual work already undertaken internationally
- The lack of cooperation/collaboration among the wider range of agents potentially able to assist in developing digital preservation solutions; for example, the computing industry

Appropriate mitigation strategies for these concerns and for promoting the need for and importance of digital preservation would usefully include the following:

- Promotion of a more coordinated international approach to the development of solutions to challenges relating to digital preservation such as preservation metadata, persistent identifiers, and implementation models

- Programs to raise the awareness of the need for digital preservation within the community of "memory institutions" and more widely
- Studies designed to provide accurate metrics on the scope of digital material needing preservation, including extrapolations for sizing purposes

Standards

The lack of international consensus on preservation metadata is a key inhibitor to full implementation of a preservation metadata strategy at the Library. This lack of consensus reflects to some degree a catch-22 implicit in the notion of preservation metadata. There is no way to test the effectiveness and efficiency of the metadata approach to digital preservation without suffering some catastrophic loss of digital objects against which to test the metadata approach.

There is a significant degree of faith involved in the development and implementation of a preservation metadata program (which might also explain, at least partially, why it is that the library community has been at the forefront of developments in preservation metadata—metadata is a natural and integral component of our normal business practice). In making any decision on whether to implement a preservation metadata process, organizations must bear in mind the potential costs of data recovery. The risks and associated costs of data loss are as yet unknown. In a recent publication on preservation metadata, Wendy Duff from the University of Toronto states that "reliable authentic digital objects will not be preserved across time without adequate preservation metadata" (2004, p. 27). Yet, what is there in our experience of the digital environment that makes this so?

Why do we not just wait for a technological response to rescue us from our quandary? For the National Library of New Zealand this is not satisfactory. We have a legislative mandate and a professional duty to begin collecting, preserving, and making accessible digital materials now and into the future. Our legacy to the future is minimal loss of our digital heritage. To that end the cost of preservation metadata today can be considered negligible compared with cost associated with a catastrophic loss of digital material in the future that might have been mitigated had preservation metadata been available.

We also need to remember that preservation metadata has other uses within our organizations in terms of

- collection management: how many Word 2, Word 6, TIFF objects, etc. are there in our archive?
- management information: metrics for sizing, costing, etc.
- helping drive preservation decisions by knowing what is there (for example, technical input decisions for preservation activities such as migration or for making output or responsibility decisions based on curatorial expertise).

Digital preservation is an immature field, and there is no silver bullet. Even if preservation metadata is purely insurance or risk mitigation, this is sufficient justification for the present.

Standards compliance is a key operational principle for the National Library of New Zealand, and it is imperative that the Library does not go down a cul-de-sac in pursuing solutions to digital preservation issues. In this regard, while the Library has in place what it considers the main building blocks for preservation metadata (that is, schema, data model, tools, and repository) the work of the OCLC/RLG PREMIS project will be crucial to the ongoing implementation of a preservation metadata program (OCLC, 2003b).

The lack of an agreed standard is important as it makes it difficult for any organization to commit the resources required to move from the conceptual development to a practical implementation. What will happen should a common approach or standard not be able to be agreed upon by the preservation community? How are we to accommodate the specter of multiple preservation metadata standards/specifications/implementations? What would be the interoperability issues that would arise from such a situation? On the other hand, it may be that the first successful, cost-effective implementation of preservation metadata will become the de facto standard.

Implementing Preservation Metadata Processes

As standards evolve and agreement is reached regarding the schema elements and their implementation through specifications of a data model and repository architecture, and tools for capturing the agreed metadata become available, questions arise regarding the mechanics of implementing preservation metadata processes. There is genuine uncertainty as to when preservation metadata is to be captured. Should it be captured as part of an agreement with the publishing/creation community, at acquisition, or at ingest into the archive? When does preservation metadata get updated and by whom during the life cycle of the object within the archive? What is needed now is a real, fully functional system in place in order to evaluate cost, sustainability, funding, staffing, etc., and thus determine both the impact and the long-term viability of preservation metadata as a component in the digital preservation space.

The Role of Automation

The question of funding preservation metadata is not yet resolved at the Library. However, as noted above there is a very real tension between the need for this new type of metadata (and structural and rights metadata) and the ability of the traditional cataloguing function to deliver these services from within their normal staffing establishments—thus, the necessary ac-

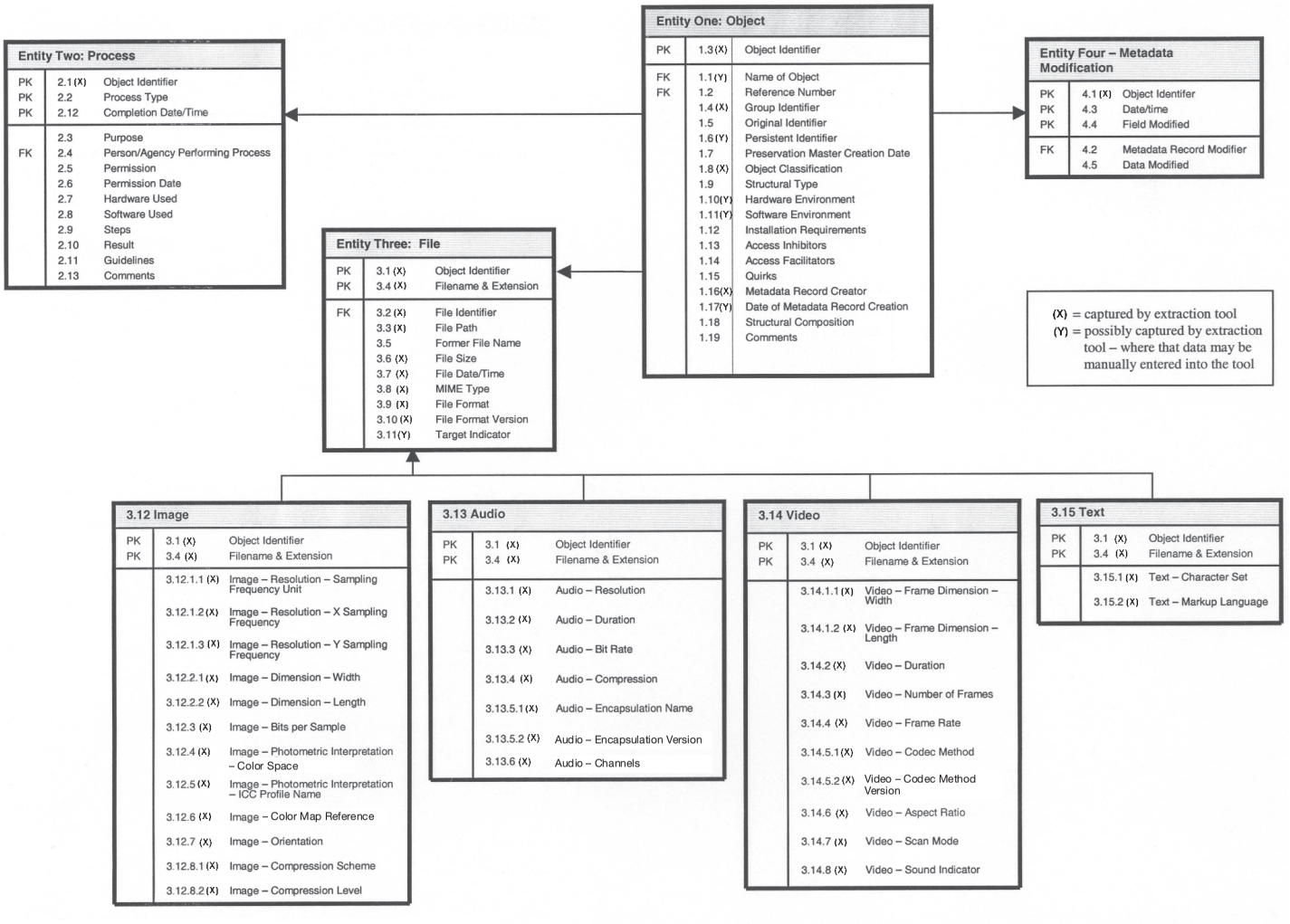

Figure 1. Data Elements to Be Captured by Extract Tool

cent on whether preservation metadata gathering can be automated and to what extent. Effectively, the more digital preservation activities can be undertaken by means of automation, the more achievable our objectives will become.

The Library's work on preservation metadata has always been predicated on two simple questions: Is what is being proposed absolutely essential or core "preservation" metadata? Is what is being proposed achievable programmatically? The first of these ensures that the focus is on preservation metadata and does not include metadata that is either unnecessary or more properly situated elsewhere; for example, descriptive metadata or rights metadata. It is important in the preservation context to be clear and that we are collecting what we *need* not what we *want*. The second question explicitly recognizes the need to incorporate automated routines as fully as possible into digital preservation solutions.

Figure 1 shows the proportion of the Library's data model that we expect to be captured automatically with the extract tool (the elements marked with an *X*). The elements marked with a *Y* we hope to be able to extract programmatically or at the least be able to feed into the tool as parameters. It is clear from this that a significant amount of the required metadata should be obtainable programmatically.

Preservation Metadata Extract Tool

As noted above, the Library has developed a tool for automatic extraction of metadata from files. It consists of a base generic extract process with "adapters" for extracting metadata from specific file types. To date, fifteen adapters have been written—for Microsoft Word 2 and Word 6, TIFF, WAV, JPEG (including the EXIF data), BMP, HTML, Open Office, Excel, PowerPoint, Microsoft Works, Word Perfect, PDF, GIF, and MPEG.

The tool works as follows (see figure 2):

1. User selects files and invokes the extract tool.
2. The tool automatically selects the appropriate adapter to use for any given file type.
3. The tool outputs a native XML format containing all the information it was able to find in the file headers.
4. An XSL style sheet is run over the native format file to create an XML file generated on the basis of an XML DTD schema version of the Library's data model. That output is the final preservation metadata as it is understood to date.
5. If there is no adapter for a file, a default set of metadata is generated based on the file attributes recorded from Entity 3.1 to 3.11 of the Library's data model. This ensures that, if an unknown file type is encountered, a minimum set of metadata can be extracted.

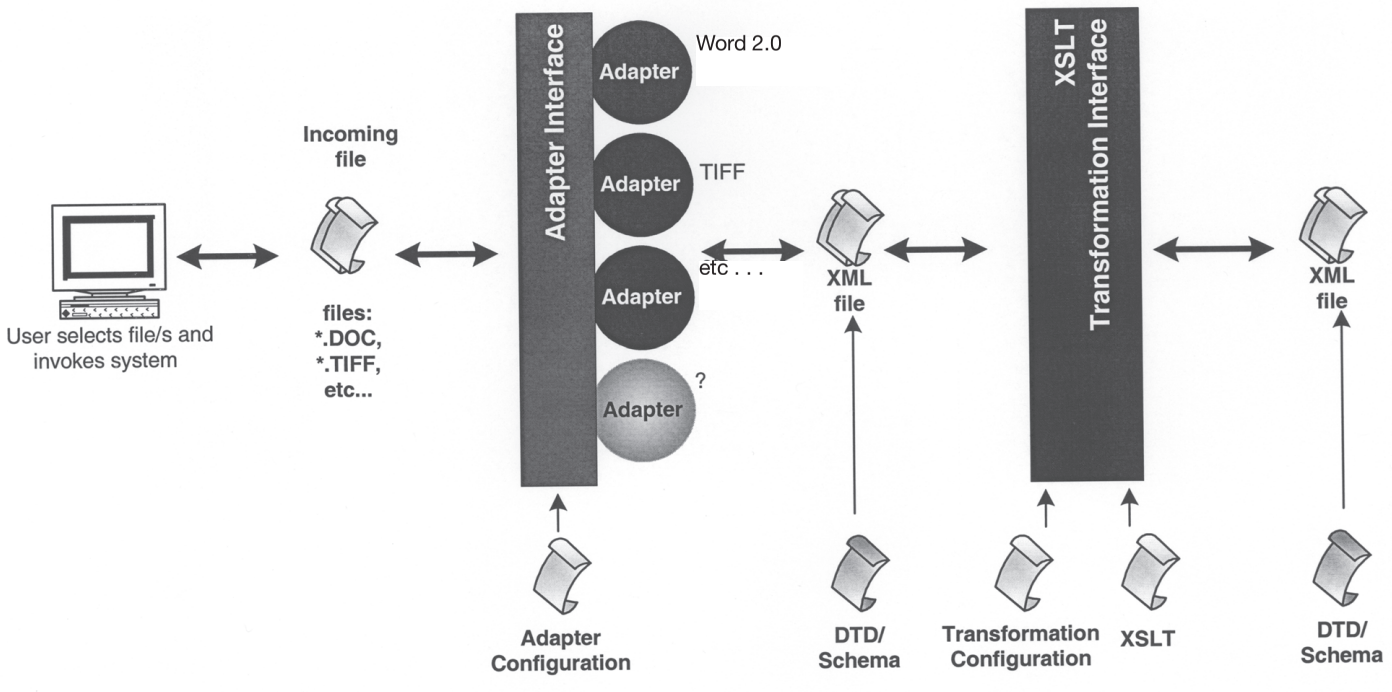

Figure 2. Preservation Metadata Extract Tool

More adapters are planned. The tool has been developed with sufficient flexibility that functionality around the application can be developed separately from the adapters and more adapters can be plugged into the application. The tool is also customizable for other institutional purposes; for example, the XML output of the application can become the input for another application or be written directly to a repository. The extract process itself is very fast as the files themselves are not opened, only the header is read. From a preservation perspective this allows for the metadata extraction process to be done in a secure, read-only environment. This begs the question "If the information is already in the header, why don't we just leave it there and get it later if/when we need it?" Our response, as in the argument for preservation metadata above, is that we are in this business for the long term and that we need to be very conservative with regards to the potential for a catastrophic occurrence where the objects and/or their internal data may not be available.

The next step for the extract tool is to make it available to the wider preservation community with a view to it becoming one component of a suite of tools supporting organizational preservation metadata strategies. One good example of how this may work is the development of JHOVE (JSTORE/Harvard Object Validation Environment), a collaborative venture between JSTORE and the Harvard University Library. JHOVE "provides functions to perform format-specific identification, validation, and characterization of digital objects" (JHOVE, 2005). Potential use cases for JHOVE include

- Identification: I have an object; what format is it?
- Validation: I have an object that purports to be of format F; is it? I have an object of format F; does it meet profile P of F? I have an object of format F and external metadata about F in schema S; are they consistent?
- Characterization: I have an object of format F; what are its salient properties (given in schema S)?

It is clear that answering these questions is a natural precursor to making the decision to extract data from any given file and to using that data as authenticated preservation metadata. It would seem a logical next step to establish an environment to facilitate further development of tools such as JHOVE and the Library's preservation metadata extract tool.

Preservation Metadata Next Steps

Key steps to further progress on preservation metadata include the following:

- How to formulate a "standard" international element set that organizations can pick up fully or partially to suit their own requirements while still staying within an agreed framework

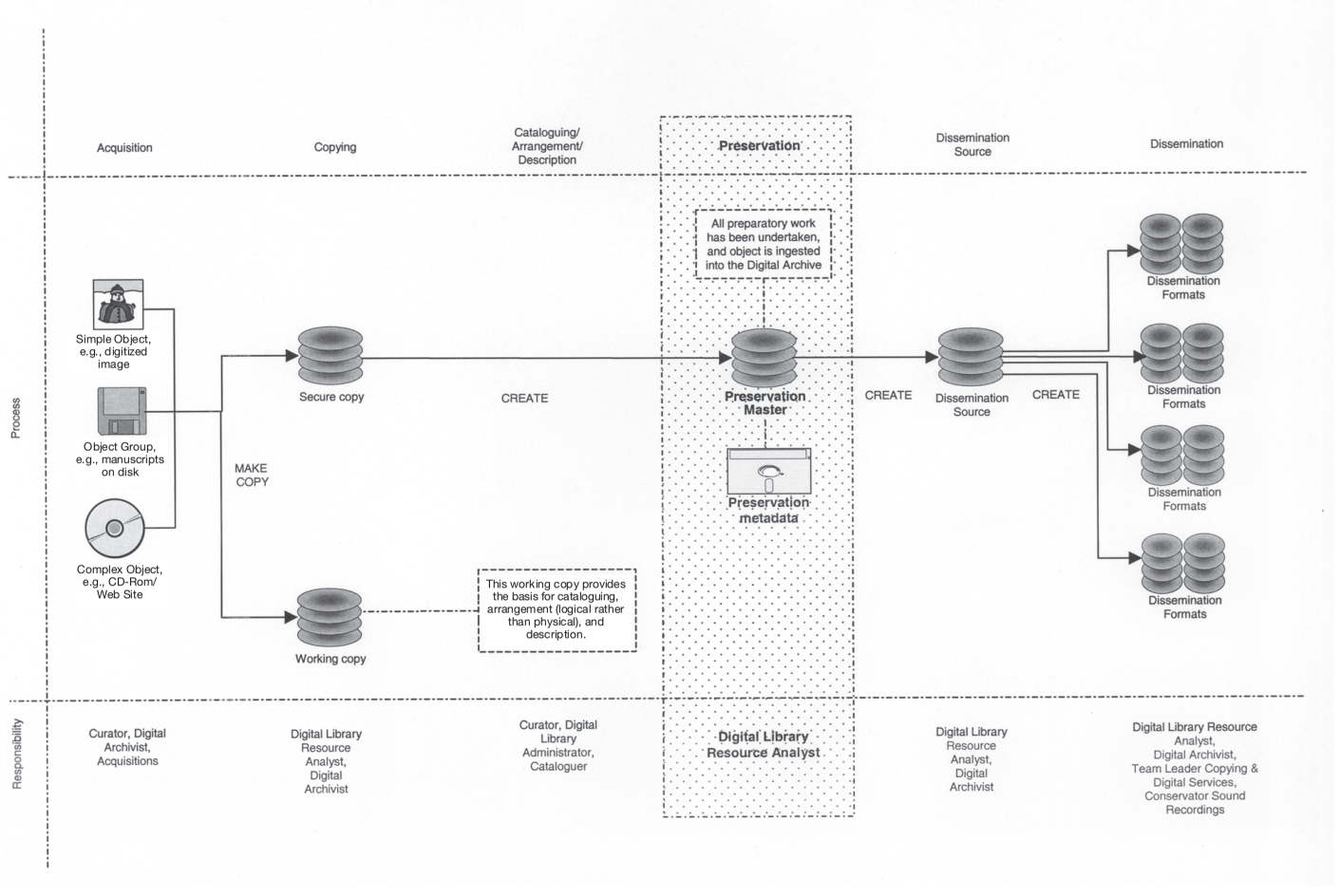

Figure 3. The Place of Preservation in the Management of Digital Objects

- Further development of an implementable data model turning the logical schema into something application-ready
- Further refinement of tools for the automatic extraction of metadata (for example, from file headers) to minimize handcrafting
- Further development of the XML schema version of the data model
- Research on the relationship between preservation metadata and METS implementations of structural metadata (in particular the use of File Groups and Digital Provenance)
- Conversations with the vendor community regarding support for preservation metadata and perhaps other aspects of digital preservation

The Wider Context of Digital Preservation

It is important to remember when looking at preservation metadata that this is only one component of our digital preservation activities and the drive to accommodate digital materials within our organizations. Figure 3 shows the place of digital preservation in the wider context of how digital materials are incorporated into our business as usual processes.

The Library is adopting a holistic approach to the long-term management of its digital assets, and it could easily be argued that, without successful resolution of all these activities, we will be unable to say with certainty that we are providing an environment conducive to digital preservation. Other areas of consideration for digital preservation include the following:

- Business process workflows; for example, selection, acquisition, and handling of digital objects
- Infrastructure for digital material; for example, storage, access, file naming (important where multiple objects have the same name, such as "Annual Report 2003"), role definitions (how do we know when a digital object is a preservation master, dissemination copy, preview copy, thumbnail, etc.), data authentication, the notion of a trusted repository (licensed/registered, not self-assigned), scalability, and sustainability (the potential to leverage a national infrastructure)
- Associated digital library activities; for example, metadata (resource discovery, structural) and persistent identifiers
- Web archiving for the capture and preservation of New Zealand Web sites
- Researching the potential of migration and emulation (especially for complex objects)
- Generic interface—one of the key elements in delivering digital material is to make its discovery layer seamless with our usual bibliographic searching tools
- Rights—online delivery does not abrogate us of our obligations to respect the rights of the owners of that material; for example, copyright, and moral rights; the impact is at both a business and a technology level

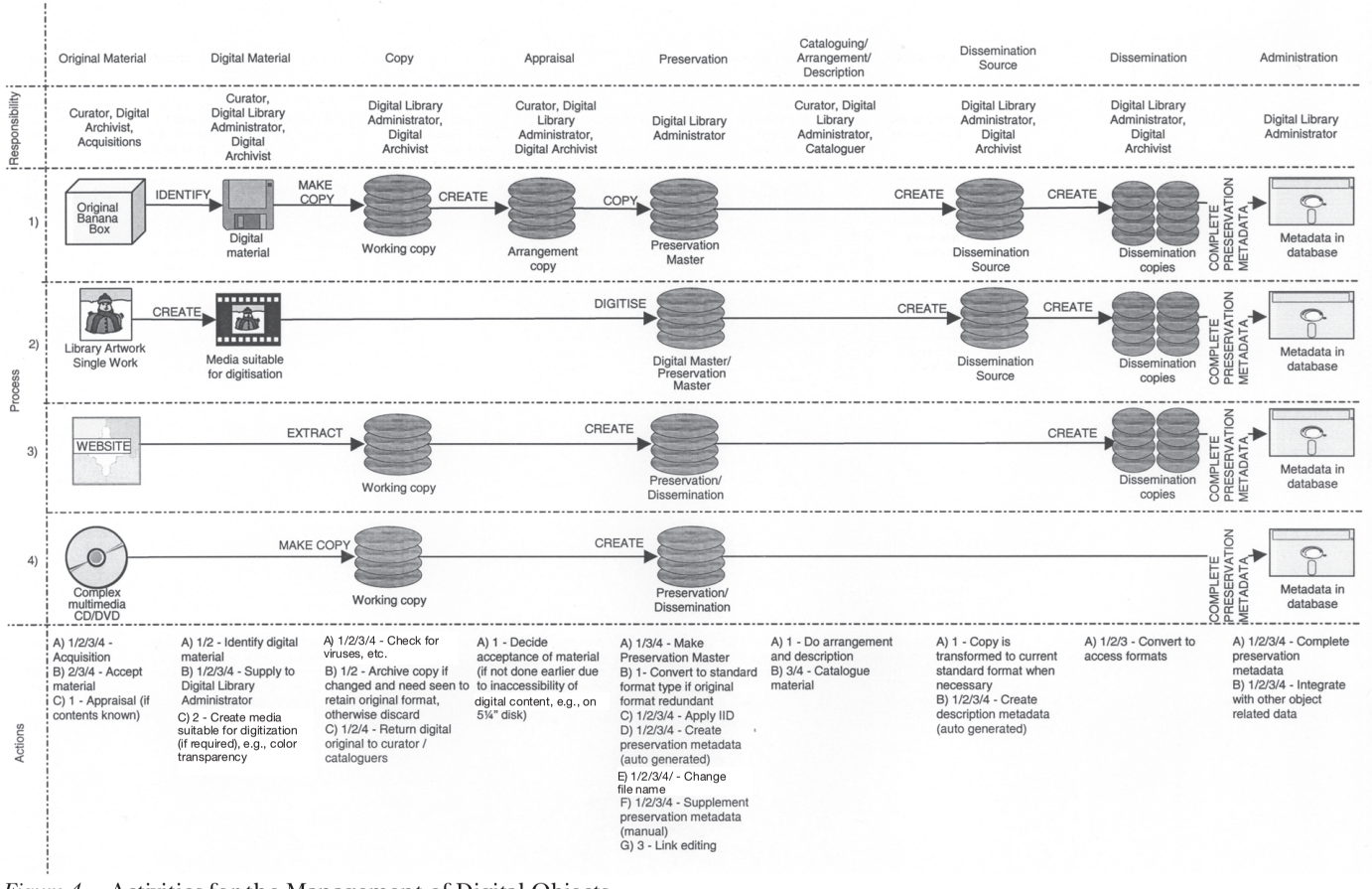

Figure 4. Activities for the Management of Digital Objects

Figure 4 is a slightly more complex version of figure 3 and shows more clearly the continuum of activities—from selection and acquisition through preservation and on to end user access—that need to be undertaken in order to incorporate digital materials into our processes as business as usual.

Organizational Impact

Digital preservation, in all its aspects, is going to require some form of organizational transformation, and it is likely that

> in addition to redefining responsibilities of organisations, it may be necessary to redefine roles within organisations to ensure long-term access to digital information. For example, responsibility for maintaining long-term access to digital records may be shared between business managers, records management and information technology personnel, and individual creators. (National Library of Australia, n.d.)

While it still remains unclear how this will manifest itself within the Library, it is clear that a mix of curatorial and technical responsibilities are already evolving around the management of digital preservation. Complicating this scenario, however, is the morphing of a number of our traditional disciplines. For example, most of our organizations have a cataloguing or arrangement and description component, but now we are having to capture and/or create preservation metadata, structural metadata, rights metadata, etc. It is clear that we will not be given the equivalent numbers of staff for these activities and that automation must be the answer for these new types of description. But how will this impact our traditional lines of responsibility within the Library, and where will the skilled staff to undertake these tasks come from?

Peter Graham and Paul Conway probably described it best when they noted that "nothing makes clearer that a library is an organization, rather than a building or a collection, than the requirement for institutional commitment if electronic information is to have more than a fleeting existence" (Graham, 1995) and that "the real challenge is creating appropriate organizational contexts for action" (Conway, 1996).

Implications for the Library include the following:

- How do we make resource allocation decisions to digital preservation when other aspects of our response to the digital world—digitization, national site licensing, etc.—are competing for the same funds in an often static funding environment?
- What are the legal implications of the various strategies for digital preservation, almost all of which require copying of some form or other?
- Where do we find the staff needed to implement digital preservation strategies?
- With technological change proceeding as fast as it is, who is able to train people with the appropriate skills and ensure that those skills remain current?

There are also wider issues, including those discussed below.

National Capability

There is an increasing acceptance that digital preservation may not be the province of a single organization. However, there are probably few organizations, even internationally, who have the mandate, let alone the technical, staffing, and financial resources, to develop a sustainable "trusted digital repository." How might distributed responsibility for digital preservation work in a small nation of only four million people such as New Zealand? Is it even viable? If not, what are the statutory, social, and professional implications of a single, centralized approach to digital preservation?

Cost

It is difficult to say with any certainty what the costs of digital preservation are going to be. A recent report from the National Library of Australia notes that

> a surprising observation . . . 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. (Gatenby, 2002)

Similarly, a recent review of the National Library of New Zealand's digital archiving activities found that

> 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. (Ross, 2004, p. 43)

Redundancy

The objects we collect will increasingly be created on computers, collected from computers, stored in computers, preserved in computers, and made accessible from computers. As a consequence of this, the need for redundancy will increase. The review of the Library's digital archiving activities quoted above also noted that

> as the digital holdings of the Library 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 the record of the cultural and scientific heritage of New Zealand and the South Pacific . . . 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. (Ross, 2004, pp. 27–29)

A Brief Note on Trusted Digital Repositories

Possibly the greatest challenge facing us in relation to digital preservation is the notion of a "trusted digital repository" as articulated recently by RLG and OCLC. (Research Libraries Group, 2002). Garrett and Waters recognized in 1996 that

> for assuring the longevity of information, perhaps the most important role in the operation of a digital archive is managing the identity, integrity and quality of the archives itself as a trusted source of the cultural record. Users of archived information in electronic form and of archival services relating to that information need to have assurance that a digital archives is what it says that it is and that the information stored there is safe for the long term (Garrett & Waters, 1996, p. 23).

Implicit here is the notion of *provenance* (the relationship between records and the organizations or individuals that created, accumulated, and/or maintained and used them in the conduct of personal or corporate activity). Whenever digital preservation is discussed, issues of migration, encapsulation, emulation, etc. arise. What must be kept explicit in these discussions is the notion of the look and feel of the object, the intellectual content of the object, the need to minimize change to the object, and the need to fully document any change that has to be made to a digital object in order for it to be passed into the archive in a state ready for preservation. This includes every process undertaken against the Preservation Master (in the Library's model) along with information on who undertook the process, why, under who's authority, what the process was, how it was effected, any changes that were made to the object as a result of the process, etc. This record is in line with the notion that one aspect of provenance is the history of custody of the described materials since their creation, including any changes successive custodians made to them.

While the current work of the Library may enable it to resolve issues relating to the integration of digital resources into its normal business practices, it is clear that this does not automatically ensure that the Library fulfills the requirements of a trusted digital repository. Nor does it mean that the Library will not have to develop relationships with other organizations that might wish to achieve trusted repository status in a country with a small population base and few agencies of appropriate size, funding, and willingness to take on the role.

While trust is already a feature of the Library in its capacity as a national library, it is not a given that that trust will automatically be bestowed upon it in the digital arena. The Library's work with digital material needs to leverage off its status of trust in the analogue context, but it must now develop a reputation for trustworthiness over time in these new activities through transparency of process, accountability, and reliability.

This is not only about the Library either. In New Zealand we have the opportunity to develop a national program (incorporating archives, muse-

ums, galleries, libraries, etc.) that hopefully will fit into a global structure of trusted digital repositories. However, this will require a level of cooperation and collaboration beyond anything we have attempted to date. It may also require individual disciplines to look at and transcend community-specific paradigms that have developed over time but that may not be appropriate in the digital context.

Conclusion

In 1996 Garrett and Waters rightly stated that

> the problem of preserving digital information for the future is not only, or even primarily, a problem of fine tuning a narrow set of technical variables. It is not a clearly defined problem . . . rather, it is a grander problem of organizing ourselves over time and as a society to maneuver effectively in a digital landscape. It is a problem of building . . . the various systematic supports . . . that will enable us to tame anxieties and move our cultural records naturally and confidently into the future (Garrett & Waters, 1996, p. 7).

This article places the National Library of New Zealand's work on preservation metadata in the context of the Library's overall response to the management of digital material and in the wider evolving context of the notion of trusted repositories. I noted above the catch-22 inherent in the preservation metadata approach to digital preservation. However, we need to move forward now on what we believe to be the correct path at this moment. We must demonstrate that the cost of incorporating preservation metadata into our preservation program today will be minimal compared with the future cost of not doing so. For the National Library of New Zealand it has been particularly gratifying to be able to show that implementing an end-to-end process for preservation metadata (schema, data model, gathering/extraction, and storage) is viable and that a significant amount of the required metadata can be gathered programmatically.

I noted earlier that digital preservation is a new business need of great complexity. With digital material requiring new methods of storage, management, and presentation, the work described in this article has begun the process of effecting the changes required in our organization to ensure the preservation of our fragile, ephemeral digital material.

The challenge for the Library now is to move from a period of high conceptualizing to implementing that ideal state where the nation's digital cultural heritage is preserved in perpetuity (New Zealand Government, 2003, p. 8), a challenge that we are looking forward to with genuine excitement. It is a great time to be a librarian.

Acknowledgments

I would like to thank Seamus Ross (Director, Humanities Computing and Information Management, University of Glasgow) and Frank Bischoff

(Director, Archives School, Marburg) for their permission to use parts of a paper developed from a presentation to an ERPANET seminar on preservation metadata (http://www.erpanet.org/php/Marburg/seminar.htm) held at the Archivschule Marburg in September 2003 (http://www.uni-marburg.de/archivschule/) (Knight, 2003b).

I would also like to thank the Library & Information Association of New Zealand Aotearoa (http://www.lianza.org.nz) for their permission to use parts of a paper presented at the LIANZA National Conference 2003 (Knight, 2003a).

NOTE

1. *Taonga* is a Maori language term generally denoting precious cultural heritage or physical treasures. See http://www.learningmedia.co.nz/ngata/index.html for an idea of the complexity of this term.

REFERENCES

Cedars Project. (2002). *Cedars guide to preservation metadata.* Retrieved February 10, 2005, from http://www.leeds.ac.uk/cedars/guideto/metadata/guidetometadata.pdf.

Consultative Committee for Space Data Systems. (2002). *Reference model for an Open Archival Information System (OAIS).* Retrieved February 10, 2005, from http://www.ccsds.org/CCSDS/documents/650x0b1.pdf.

Conway, P. (1996). *Preservation in the digital world.* Retrieved February 10, 2005, from http://www.clir.org/pubs/reports/conway2/index.html.

Day, M. W. (1997). Extending metadata for digital preservation. *Ariadne, 9.* Retrieved February 10, 2005, from http://www.ariadne.ac.uk/issue9/metadata/.

———. (1998). *Metadata for preservation: Cedars project document AIW01.* Bath, UK: UKOLN. Retrieved February 10, 2005, from http://www.ukoln.ac.uk/metadata/cedars/AIW01.html.

Dempsey, L., & Heery, R. (1997). *Specification for resource description methods. Part 1: A review of metadata: A survey of current resource description formats.* Bath, UK: UKOLN. Retrieved February 10, 2005, from http://www.ukoln.ac.uk/metadata/desire/overview/overview.pdf.

Duff, W. (2004). Metadata in digital preservation: Foundations, functions and issues. In F. M. Bischoff, H. Hofman, & S. Ross (Eds.), *Metadata in preservation: Selected papers from an ERPANET seminar at the Archives School Marburg, 3–5 September 2003* (Veröffentlichungen der Archivschule Marburg, Institute für Archivwissenschaft, Nr. 40) (pp. 27–38). Marburg: Archivschule.

Garrett, J., & Waters, D. (Eds.). (1996). *Preserving digital information: Report of the Task Force on Archiving of Digital Information.* Washington, DC: Commission on Preservation and Access and The Research Libraries Group. Retrieved February 10, 2005, from ftp://ftp.rlg.org/pub/archtf/final-report.pdf.

Gatenby, P. (2002). *Report on senior executive fellowship to research digital archiving in national libraries.* Retrieved February 15, 2005, from http://www.nla.gov.au/nla/staffpaper/2002/elect.html.

Graham, P. S. (1995). *Requirements for the Digital Research Library.* Retrieved February 15, 2005, from http://www.ifla.org/documents/libraries/net/drc.htm.

JHOVE (Jstore/Harvard Object Validation Environment). (2005). *Format-specific digital object validation.* Retrieved February 15, 2005, from http://hul.harvard.edu/jhove/.

Kenney, A., & Reiger, O. (2000). *Moving theory into practice: Digital imaging for libraries and archives.* Mountain View, CA: Research Libraries Group.

Knight, S. (2003a). *A brief introduction to digital preservation.* Paper presented at Oceans of Opportunity Whakawhitihia te Moana: Annual Conference of the Library and Information Association of New Zealand Aotearoa. Retrieved February 15, 2005, from http://www.lianza.org.nz/conference/conference03/papers/knight.pdf.

Knight, S. (2003b). *Preservation: The urgency? A case study from the National Library of New Zealand.* In F. M. Bischoff, H. Hofman, & S. Ross (Eds.), *Metadata in preservation: Selected papers from*

an ERPANET seminar at the Archives School Marburg, 3–5 September 2003 (Veröffentlichungen der Archivschule Marburg, Institute für Archivwissenschaft, Nr. 40) (pp. 39–58). Marburg: Archivschule.

National Library of Australia. (n.d.). *PADI (Preserving Access to Digital Information) roles and responsibilities.* Retrieved February 15, 2005, from http://www.nla.gov.au/padi/topics/8.html.

———. (1999). *Preservation metadata for digital collections.* Retrieved February 15, 2005, from http://www.nla.gov.au/preserve/pmeta.html.

National Library of New Zealand. (2000). *Metadata standards framework for National Library of New Zealand.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/4initiatives_metafw.pdf.

———. (2002). *Metadata standards framework—Preservation metadata.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/4initiatives_metaschema.pdf.

———. (2003a). *Introduction to the metadata standards framework for National Library of New Zealand.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/en/whatsnew/4initiatives.html#meta.

———. (2003b). *Metadata standards framework—Metadata implementation schema.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/nlnz_data_model.pdf.

———. (2003c). *Metadata standards framework—Preservation metadata (revised).* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/4initiatives_metaschema_revised.pdf.

New Zealand Government. (2003). *National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/Act03-19.pdf.

OCLC. (2003a). *Preservation Metadata Framework Working Group.* Retrieved February 15, 2005, from http://www.oclc.org/research/projects/pmwg/wg1.htm.

———. (2003b). *PREMIS (Preservation Metadata: Implementation Strategies).* Retrieved February 15, 2005, from http://www.oclc.org/research/projects/pmwg/.

Research Libraries Group. (2002). *Trusted digital repositories: Attributes and responsibilities.* Mountain View, CA: Research Libraries Group. Retrieved February 15, 2005, from http://www.rlg.org/longterm/repositories.pdf.

Ross, S. (2004). *National Library of New Zealand (Te Puna Mātauranga o Aotearoa): Digital library development review.* Retrieved February 15, 2005, from http://www.natlib.govt.nz/files/ross_report.pdf.

Thompson, D., & Searle, S. (2003). Preservation metadata: Pragmatic first steps at the National Library of New Zealand. *D-Lib Magazine, 9*(4). Retrieved February 15, 2005, from http://www.dlib.org/dlib/april03/thompson/04thompson.html.

University of Heidelberg Institute for Chinese Studies. (2005). *Digital Archive for Chinese Studies: About DACHS.* Retrieved February 15, 2005, from http://www.sino.uni-heidelberg.de/dachs/intro.htm.

Steve Knight, Manager of Digital Strategy Implementation, Electronic Services, National Library of New Zealand, P.O. Box 1467, Wellington 6001, New Zealand, steve.knight@natlib.govt.nz. Steve Knight is the Manager of Digital Strategy Implementation (DSI) at the National Library of New Zealand. The DSI team was established to ensure the long-term storage, preservation, and provision of access to New Zealand's digital cultural heritage and to enhance access to New Zealand's cultural heritage online through an increased digitization program. In conjunction with other business units, the team researches and facilitates the implementation of the operational and technical infrastructure for the integration of digital materials into the collections of the National Library. The DSI team is currently leading the business side of a multiyear project to establish a National Digital Heritage Archive. From a library background Steve has had experience in a range of information management disciplines, including records management and document management. Much of this work has been in the design and implementation of electronic services.