



**NEW ZEALAND
LAW LIBRARIANS'
ASSOC.**

© Copyright New Zealand Law Librarians' Association 2006

New Zealand historical legislation: electronic capture, preservation and publication

Background paper and map for a scoping study for the New Zealand Law Librarians' Association

Michael Rubacki

9 March 2006

Table of Contents

Foreword		3
Part 1	Introduction	4
1.1	Purpose of document	4
1.2	Abbreviations and definitions etc	4
Part 2	Overall project objectives	5
2.1	New Zealand statutes: value, role and proposed future	5
2.2	Current position: access, preservation and policy framework	6
2.3	New Zealand PAL project	7
2.4	Proposed project objectives	7
2.5	Benefits	8
Part 3	Assumptions, interdependencies, broad requirements	9
3.1	Target material	9
3.2	Policy matters	10
3.3	Technology and format	10
3.4	Proposed approach	11
3.4.1	General	11
3.4.2	Data capture and accuracy	11
3.4.3	Source material	12
3.4.4	Scanning and PDF	12
3.4.5	Capturing older New Zealand material: challenges	12
3.4.6	Effective analysis, digital capture and preservation	13
3.4.7	Hypertext links	13
3.4.8	Metadata	13
3.5	PAL system: timing, integration and risk	13
Part 4	Technical scoping study: specific requirements	15
4.1	Analysis of content and PAL integration	15
4.1.1	Structural analysis and DTD development/modification	15
4.1.2	Problem areas	15
4.2	Metadata requirements	16
4.2.1	Minimum metadata needs	16
4.2.2	Metadata problem areas	16
4.3	Publication	17

	4.3.1 Integration with PAL database and website	17
	4.3.2 Formats	17
4.4	Identification and sizing of target material	17
4.5	Costing	18
4.6	Business case and project plan	19
Part 5	Scoping study: tender, process and budget	20
5.1	Tender	20
5.2	Budget	20

Foreword

For a number of years the New Zealand Law Librarians' Association has been concerned as the early volumes of the laws of our land deteriorated. In 2004 an article was published in the Auckland District Law Society's Law News, pointing out the problem of the "shattering statutes" – volumes from the late 19th and early 20th centuries which were printed on the first wood pulp based paper. The acid remaining in this paper has left us with the legacy of a whole section of our history which is now crumbling – literally, as the pages are turned. Although some volumes remain unaffected in such places as the Alexander Turnbull Library, many other libraries no longer have copies.

Although statutes are such an important part of our history, and are referenced more and more for Treaty research and by historians as well as by working lawyers, judges and law librarians, they seem to fall into an odd gap between government agencies. The current "statute book" is the business of the Parliamentary Counsel Office, which has for a number of years been developing the Public Access to Legislation (PAL) project, which will give online access to every citizen. PCO is not, however, responsible for preserving or providing access to historical statutes.

The government's Digital Strategy appears to offer the potential to digitise historical statutes. The NZLLA has sought assistance from the National Library of New Zealand in respect of a possible digitisation project, so far without success.

In December 2005 the NZLLA decided to take the next step itself, by commissioning this report. Michael Rubacki is a consultant who has worked with the PCO, and he has had wide experience with statutes digitisation in New South Wales. We commend his report to readers as guide to the broad requirements, technical scoping and process to be followed in working towards a solution. New Zealand, as Michael points out, is fortunate that its statutory history is not so long that a complete digital record would be an impossibly large project. If all statutes, current and historical, could be available to all of us through the Public Access to Legislation website, New Zealand would become a world leader in this regard. We seek your support for this important project.

Helga Arlington

President, New Zealand Law Librarians' Association

PO Box 25062

Wellington

New Zealand

<http://www.nzlla.org.nz>

Part 1 Introduction

1.1 Purpose of document

This report has been commissioned to define and map a project to capture, preserve and publish New Zealand historical legislation in a way that is compatible with the current Public Access to Legislation Project. In particular, the report is intended to describe the broad objectives, assumptions, and requirements for such a project and to provide specifications for a detailed technical scoping study that is essential to accurately measure and cost the project and test its feasibility.

This report, combined with the proposed technical scoping study, is intended to form the basis of a business case to secure funding and guide the conduct of a significant project. It could also form a blueprint for similar projects involving other important historical material such as Hansard and the Government Gazette.

The approach taken is one that enables a “capture and publish” project to be scalable and capable of being undertaken on an incremental basis, with the facility for later enhancement and linkages. The approach is also based upon compliance with government guidelines in the areas of web publishing and records management.

1.2 Abbreviations and definitions etc

In this report:

“DTD” means Document Type Definition, which is a set of rules that define the elements that can be used (and their order or combination) in a particular class of document. DTDs are used as the skeleton or framework for XML publishing, especially complex legislative and technical documents.

“HTML” means Hyper Text Markup Language, the authoring language most commonly used to create documents for publishing on the Internet. It defines the structure and layout of documents using tags and attributes.

“NZLLA” means New Zealand Law Librarians’ Association.

“OCR” means optical character recognition - the capturing of images into machine readable text usually by means of scanning.

“PAL” means Public Access to Legislation in the context of the current major IT project being undertaken by the New Zealand Parliamentary Counsel Office to develop an integrated system for drafting, storing and publishing New Zealand legislation.

“PCO” means New Zealand Parliamentary Counsel Office.

“PDF” means Portable Document Format. This is a file format created by Adobe Systems based on its page description language and delivers electronic files that appear exactly as they would on a printed page.

“XML” means eXtensible Markup Language and is a general-purpose markup language for creating, storing and publishing data. Its primary purpose is to facilitate the sharing of data across different systems, particularly systems connected via the Internet, and is commonly used to automatically generate HTML.

Part 2 Overall project objectives

2.1 New Zealand statutes: value, role and proposed future

The New Zealand statutes, commencing with the ordinances of 1841, provide the legislative backbone of the nation as well as serving as an important source of its history in annual volume form. The historical statutes provide revealing insights into the legal and economic life of the colony and emerging nation. They also represent the legislative output of the Parliament on a year-by-year basis and provide a valuable measure of the country's socio-political development.

Although most of the older statutes have been repealed and replaced with current laws, the New Zealand "statute book" consists of an indivisible body of law spanning its entire modern history, from the early colonial government to the present day. Some nineteenth century statutes remain in force or continue to be of active interest to the legal profession and the courts, quite apart from the entire body of work being of historical interest to the wider community. The collection is an essential resource for the purposes of legal precedent and research and as a means for tracking the development of laws and legislative concepts.

However, the annual volumes are becoming increasingly difficult to access and use, especially the nineteenth century material, because of paper degeneration and also as they have not been digitised or captured electronically by the commercial or public sector publishers.

The main objective behind the proposed project is to capture all of the statutes, starting with the nineteenth century material, and bring them into a single and comprehensive collection, within the same technology being developed for the PAL project described below. This would enable the entire New Zealand statute book to be fully accessible and searchable online, and maintained permanently as a publicly owned resource as part of the PAL system.

Such a resource would enable the ready identification and retrieval of statutes by the courts, legal profession, academics, students and parliamentarians, as well as members of the general public with special interests in New Zealand history. The proposed technology and comprehensive nature of the collection would allow detailed text searching and citation-linking across the entire statute book and this would yield major savings over time in terms of research costs and convenience, in addition to enabling forms of research that would be almost impossible to undertake using paper volumes.

The proposed project is conceived as an active legal publishing system with all the flexibility and potential for reuse and interactivity that this entails. It does not propose a purely archival or other static library/repository model because the opportunity for a far better solution is so compelling in the current circumstances.

The capture and online publication of the historical statutes to the requisite high standards will be a major and costly task but New Zealand has several advantages:

- its entire collection of statutes is relatively small (less than 18,000 pages of 19th century material) and recent so data capture costs will be far less than in most comparable jurisdictions
- the structure and layout of the nineteenth century statutes are relatively simple and consistent

- the current PAL project is well-advanced and provides an ideal technological platform and business model for the capture and integration of the target material, when its current stage is complete.

The capture of the New Zealand statute book and its integration with the PAL system would represent an outstanding achievement in international terms and provide a best practice model for preserving and providing public access to historical statutes and similar documents.

2.2 Current position: access, preservation and policy framework

As mentioned above, access to the bound volumes of historical statutes, especially those enacted and printed in the nineteenth century, is becoming increasingly restricted. The documents were printed on acid paper and many collections are increasingly frail. Pages tend to crumble away or shatter on handling.¹ Access to older volumes is limited for this reason and many important legal collections are now incomplete.

There is a growing expectation that this heritage material should be permanently preserved and also made more widely accessible using online technology. This is encouraged by government initiatives in New Zealand and Australia in the areas of record keeping, “e-Government” and the associated provision of free public access to current legislation and related documents.²

The capture of historical material on microfilm or by reproducing facsimile paper editions are not satisfactory solutions as these approaches provide neither permanent nor wide public access. There is a strong case to effectively digitise this material once and for all time, and to the highest standard.

It is understood that at least one commercial publisher has or is planning to provide online access to historical New Zealand legislation and the commercial publishers all have scanned or reprinted some legislative material for selected periods or products. Such products and services are unlikely to meet current standards required in terms of accuracy, comprehensiveness, accessibility, or permanence. The material should be official and authoritative, clearly owned and published by the Crown, and freely available online to browse, search and download.³

¹ Arlington, H. “Acid paper ‘time bomb’ threatens NZ Statutes”. *Law News*. Issue No 06, February 20, 2004. Auckland District Law Society.

<http://www.adls.org.nz/profession/lawnews/2004/issue06/march37.asp>

² In particular, the New Zealand Digital Strategy launched in 2004 focuses on the digital future and seeks to ensure that the nation is a world leader in terms of using information and communications technology to realise economic, social and cultural goals. See <http://www.digitalstrategy.govt.nz/>

³ A comprehensive account of the recent history of the tangled web of legal government publishing in New Zealand and the case for public ownership and free public access is provided in Lawn, G. “Improving Public Access to Legislation: the New Zealand experience”. Positioning for the Future. Legal Information Symposium 2004. University of Auckland, Auckland, New Zealand, 22-24 July 2004. <http://www.pco.parliament.govt.nz/pal/gla3.shtml>

2.3 New Zealand PAL project⁴

The PAL Project will revolutionise the way in which New Zealand legislation (Bills, Acts, Statutory Regulations, and Supplementary Order Papers) is made available to the public. The aim of the project is to radically improve public access to up-to-date official legislation in both printed and electronic form.

The PAL project is focused on providing access to legislation in force (that is current legislation maintained in an updated form) rather than historical versions of legislation as enacted, although all enacted versions from 2007 will be available from the website and there is the possibility that earlier electronic versions of statutes could be added. After the go-live date, superseded or repealed versions of legislation (and the Bills that are captured or created within the PAL system and are subsequently enacted as statutes) will all be permanently retained on the PAL website as historical versions. This will provide a permanent and effectively a point-in-time collection. Users will be able to trace the development and movement of items of legislation and readily navigate within and across items. They will also be able to search across whole collections for a particular word, expression or reference.

The PAL project involves the implementation of a new XML-based drafting and publishing system within the PCO that is integrated with its current functions of drafting new legislation and publishing new legislation and reprints of existing legislation. Implementation of the PAL system is planned for early 2007.

The philosophy for the PAL project is based upon an authoritative legislation database owned by the Crown, and a website with free public access. There is a most compelling case for this to be the platform and model for a project to capture, preserve and publish the historical statutes.

2.4 Proposed project objectives

This paper describes a project to capture, preserve and publish historical New Zealand legislation with two overarching key objectives:

- Develop a comprehensive and permanent collection of all New Zealand statutes, as enacted from 1854, including at least the Ordinances made in the period 1841-1853 and the 1908 consolidation, to the present.⁵
- Integrate, maintain and publish this material within the PAL system (an essential interdependency for the project).

The reasoning and details underlying these objectives are provided below.

⁴ Ibid. The PCO also maintains a collection of background information and updates about the PAL project on its corporate website: <http://www.pco.parliament.govt.nz/pal>

⁵ The entire statute book is the long-term aim but it is acknowledged that the older material should be the first priority and the one-off capture of the entire set may be unaffordable. The inclusion of the 1908 consolidation is highly desirable and needs to be assessed in more detail. The 1931 and 1957 reprint series and subsequent reprints are also mentioned but are less likely to be candidates at this stage.

2.5 Benefits

The benefits flowing from the proposed project are difficult to quantify in simple economic terms (for example those benefits flowing from savings in library costs and research time), but the capture of historical New Zealand legislation would fulfil the widest spectrum of preservation and publication purposes and needs:

- Public ownership and access
- Academic, professional and general research
- Legal certainty
- Archival permanence
- Inestimable heritage value
- Best practice technology and model for similar projects.

Part 3 Assumptions, interdependencies, broad requirements

3.1 Target material

Ideally, the entire New Zealand Statute Book, consisting of the annual volumes and the 1908 consolidation, major reprints of 1931 and 1957 and even subsequent individual reprints should be captured, preserved and published online. Taking this proposal a step further, it would also be highly desirable if the subordinate legislation (rules and regulations) made or reprinted could be similarly dealt with⁶. Parallel arguments could be mounted for other key government documents and records including Hansard, other parliamentary papers and, in particular, the Government Gazette.

However, it is necessary for reasons of sheer scale and cost to approach this work incrementally and on a priority basis. The most compelling material to be targeted is the nineteenth century statutes as enacted but with the aim of eventually adding the statutes made between then and the go-live date of the PAL system, so as to preserve and provide access to the entire New Zealand statute book. The 1908 consolidation also merits early inclusion.

It is recommended that the statutes be captured as originally enacted. The prospect of doing any substantial retrospective point-in-time consolidations or any textual amendments would be an impossible task, although some later back-capture of point-in-time consolidations of truly “landmark” items of legislation would be feasible.⁷

One of the key tasks of the proposed detailed scoping study is to quantify the extent of data in the historical statutes, in terms of number of titles, pages and an estimated character count. Apart from assisting with estimating costs, this task would aid the breakdown of the material into manageable tranches or sub-projects, especially if the analysis of structure and formats is considered at the same time. Specific details are provided in the following part.

As a related but separate task, the 1908 consolidation needs to be assessed. Apart from the scale and structure of the work, its relationship with the annual volumes in terms of presentation and searchability will need to be clarified.

It would also be worthwhile briefly assessing the 1931 and 1957 reprints, although these would not readily fit within the “as enacted” annual volume model and may need to be treated as separate online collections.

It is assumed that applied Imperial Acts will only be captured if they were reprinted within the collections described above.⁸

⁶ Subordinate legislation was not published separately in New Zealand until 1936 and before then the instruments were published within the Government Gazette. Location and extraction of the pre-1936 material would add a considerable layer of complexity to the task.

⁷ See for example the work being done in New South Wales with point-in-time versions of selected titles under “What’s new” at <http://www.legislation.nsw.gov.au/>

⁸ There may be some possibilities in the areas of hypertext linking references to applied Imperial Acts.

3.2 Policy matters

The New Zealand PAL project is based on a free, public online access service, operated and owned by the Government. This is intended to provide official versions of legislation from an authoritative source site. It will constitute a permanent record by retaining all versions of legislation (from the point that the site is complete and operational).

The website has been designed to conform to the New Zealand Government Web Guidelines⁹ and is clearly intended to be a best practice legislation website especially in its accessibility. It will be underpinned by state of the art technology in terms of data storage, management and presentation. In addition to providing free public access to legislation to the public, source data is to be provided free to legal publishers and others wishing to reuse or republish it.

It is noted that free online access to legislation is now the accepted policy in all Australasian jurisdictions and most legislative drafting offices are directly publishing this material and integrating the process with their traditional drafting functions. It is unlikely that commercial publishers could provide the stability, continuity, or level of accuracy associated with an authoritative, free, public access website that complied with best practice guidelines and contained no advertising or revenue earning potential.

Accordingly, a compelling case can be made for the historical statutes to be integrated with the PAL system and made available on a similar policy basis and from the same website <http://www.legislation.govt.nz> .

3.3 Technology and format

The PAL system is based upon XML technology. This requires the data content to be closely analysed and then defined, captured and marked up to a DTD or Schema. This technology is relatively costly to develop but the benefits associated with XML-based systems are significant:

- Portability, inter-usability, and longevity of data (it is non-proprietary and vendor neutral) – this facilitates a future-proof repository for documents such as legislation
- Publishing in paper and multiple electronic formats from a single data source
- Fast internet accessibility and the widest capacity to hyperlink and search
- Freedom from proprietary software (particularly the word processing packages commonly in use and the disruptive effects of version migration etc).

The capture, markup and storage of historical statutes in XML will ensure their longevity and ready publication and reuse, particularly when rendered as HTML documents on the Internet. The flexibility and effectiveness of this format for the electronic publication of legislation are well established and will mesh with the PAL system and comply with the New Zealand Government Web Guidelines. (The compatibility of the historical statutes with the PAL system DTDs and their overall integration with that system need to be established as an early part of the detailed scoping study).

⁹ See <http://www.e-government.govt.nz/docs/web-guidelines-2-1/index.html>

The use of PDF as the primary form of online publication is not recommended although it is strongly favoured by users wishing to print material that resembles the original paper documents, and is relatively easy to generate.¹⁰ PDF documents generated from scanning and OCR technology cannot be effectively searched or reused. They do not meet the accuracy standards for legislation (unless fully proofread), are not capable of integration within the PAL system or website and do not meet the NZ Government Web Guidelines. PDF is still a proprietary format that although widely used now may not be so usable in 20 years time. Data that has been captured as text, fully checked and marked up to retain its structure and metadata in a generic and vendor-neutral format such as SGML or XML should always be accessible and reusable.¹¹

3.4 Proposed approach

3.4.1 General

As indicated above, it is recommended that the historical statutes be captured and preserved as an as “passed” or “enacted” collection and in a fully marked up XML format. They should be published online in HTML and be fully searchable in the same way as the current and dynamic legislation on the PAL website (assuming that the statutes can be accommodated within the PAL DTDs and system). The significant benefits of this approach will be the:

- High level of data accuracy
- Ability to add hypertext links
- Ability to add significant metadata
- Overall reusability of the source data.

This level of flexibility is especially useful as the data can be developed and added to in future as the demand arises or resources become available.

It is noted that the online output in HTML format will not be a facsimile version of the original paper version but it is considered that this limitation will far outweigh the speed, size and accessibility limitations arising from the online use of graphic images and PDF documents.

The HTML material should be supplemented by a PDF download facility, as with the PAL system.

3.4.2 Data capture and accuracy

It is proposed that the material be captured (probably by double keying or by a combination of scanning and re-keying as a compare and quality control mechanism) and marked up to a suitable DTD by a specialist contractor with a proven track record

¹⁰ For more detail on the advantages and disadvantages of HTML and PDF see Chung, C, Austin, D and Mowbray, A. “In defence of plain HTML for law: AustLII’s approach to standards”. Australasian Legal Information Institute (AustLII) 1999: <http://beta.austlii.edu.au/au/other/CompLRes/1999/11/>

¹¹ There are a number of large-scale “digitisation” projects underway internationally. These tend to involve millions of pages of historical documents and are being conducted as collaborative projects by libraries and universities. They are necessarily archival and static, based on microfilm and scanning technology.

in performing this work. The technology is not as critical as the level of accuracy delivered and the cost involved. There are well-established methodologies and standards for this type of task and the output would be at a standard of accuracy suitable for the “officialisation” process envisaged by the PAL project for the legislation in force. Examples of similar projects already successfully undertaken in the region can be demonstrated in New South Wales and South Australia. It is noted that the work does have to be very well documented and closely managed in order to achieve the quality required.

3.4.3 Source material

Regrettably, it is likely that a set of volumes will need to be found and sent offshore for imaging and data capture. This is likely to involve the bindings being removed and the set sacrificed (the cost of imaging a bound collection would be prohibitive and likely to damage the older items).¹²

3.4.4 Scanning and PDF

The alternative strategy of simply scanning original documents and producing PDF documents is not recommended as it requires much laborious proofreading and manual intervention to achieve an acceptable standard of accuracy, as well as providing a less flexible, searchable, and reusable data set.

The NSW PCO scanned the annual volumes of NSW statutes 1990-1999 as an in-house project and these are now available on its legislation website as PDF documents. This exercise proved to be vastly more labour intensive than estimated just to achieve a reasonable standard of reliability (and manageably sized electronic documents). These versions are not of the same standard as the “in force” material and cannot be relied upon for official purposes. The material was inspected but not proofread. The process proved unsatisfactory as the OCR system was not sufficiently accurate, especially when dealing with type faces that were not very clear, for example italics with ligatures (*ff*), or when distinguishing between bold and Roman type. Nineteenth century monotype would probably pose similar or greater difficulties.

3.4.5 Capturing older New Zealand material: challenges

The New Zealand volumes from the 1850s and 60s held at the Mitchell Library, Sydney, were inspected and found to contain the following features that would create problems for any capture exercise but especially one based on simple scanning:

- Brown rust spots (See 1873 last 30 pp)
- Ink hand-written amendments and additions including blots (See 1851 No 5, section 4)
- Damaged and missing type (See 1867 No 24, p viii where the lower third is unreadable)
- Frail and feather-edged paper
- Blue paper with blue lines
- Multiple type faces on one line, including extra large dropped caps, all small caps, caps and small caps, and bold italic (especially in preambles and long titles)

¹² It would be interesting to pursue the possibility of preserving or reusing the cut pages to produce a facsimile edition. There may be ways of avoiding the complete loss of a set.

- 8pt sidenotes, some over 200 characters (see 1866 No 8)
- Multiple page, multiple column, landscape table (see 1867 No 24)
- Landscape tables with three tiers of headings, some straddling, and even inverted portrait headings (1862 No 22)
- Inconsistent or absent tables of contents (some Acts have none, others have a two-column “analysis” at the head, some have only an “arrangement” list of Part headings blended in with the short title and commencement provision (see 1860 generally), others have a contemporary style full table of contents (see 1867 No 24).

3.4.6 Effective analysis, digital capture and preservation

Proper analysis of the source material and quality control, leading to full digital capture and mark up will deal with these challenges and make the material more accessible than it has ever been, as well as freeing it from the vagaries and limitations of nineteenth century printing technology or twentieth century proprietary software.

However, the proposed technical methodology should be reassessed and confirmed as part of the scoping study. It is noted that technology in this field is rapidly evolving and that more automated and intelligent processes may be available.

3.4.7 Hypertext links

One of the benefits of working in the XML environment is the capacity to readily hypertext link the material. In the case of the historical statutes it would be a simple task to link to the other statutes in the collection that are referred to within the target Act. When combined with the addition of key metadata, this facility would enable the online user to jump to the repealing instrument or track changes in citation, call up categories of instruments (and there are a few categories in the mid-nineteenth century). It is anticipated that this work would be part of the initial capture exercise but further linking could be built in later on a selective or incremental basis. Part of the scoping study will be to assess the level of cross-referencing and the degree of hypertext linking that is feasible and compatible with the PAL system.

3.4.8 Metadata

A further advantage of the XML/PAL environments is the rich metadata that can be captured and used to display and order the presentation of types of legislation on a website. In the case of a collection of historical statutes, there is also a range of currency information about each Act that is essential to the user. Without this, the user may just as well have access only to the printed bound volumes. Traditional, paper-based research requires access to a series of secondary tables or guides in order to establish the status and history of a particular Act. Apart from the key information required, the metadata could be extended in future to provide more detailed information about the Act, perhaps encompassing its amendment history.

The scoping study will need to assess the question of metadata as described below.

3.5 PAL system: timing, integration and risk

The PAL system is not expected to “go live” before the end of 2006 and any developments with the proposed project dealing with historical statutes are likely to be grafted onto that system by way of enhancements. It is important that the proposed project not impede or deflect the implementation of the PAL project.

One of the key tasks of the proposed scoping study is to assess how best the historical statutes could be integrated with the PAL system and website. The first and prerequisite task will be to assess the historical statutes and their compatibility with the PAL system DTDs.

In terms of complexity, cost and risk, it is considered that the major issues with the proposed project are:

- The cost of the data capture
- The cost of integrating the data into the PAL system.

The data capture should be reasonably simple to measure and cost (and perhaps scale accordingly) but the complexity and cost of the integration with the PAL system will depend on the degree of modification required to the DTDs and publishing applications. This component will require the close collaboration of the PCO and the PAL technical developers.

Part 4 Technical scoping study: specific requirements

4.1 Analysis of content and PAL integration

4.1.1 *Structural analysis and DTD development/modification*

The target material needs to be analysed by a legislative DTD expert to establish how it could best fit into the existing PAL system DTDs and architecture. This will be a significant task and should be undertaken as an essential first step or prerequisite in the project.

The analysis will require turning most if not all of the pages in the entire collection and closely matching this with the PAL DTDs and publishing applications. Significant shifts in structure or formats over the decades will need to be identified.¹³

It is noted that the historical material is mainly static (except for possible additional metadata) and a relatively loose and simple DTD should be adequate. Also, a relatively simple XML publishing solution could be found. However, the key requirement is that the material be integrated with live New Zealand legislation and the PAL website and, ideally, fit within the PAL DTDs.

4.1.2 *Problem areas*

In addition to the structural matching of the material with the PAL DTDs, any complex or unusual structural or typographic features will need to be identified and quantified, including:

- Short titles
- Sidenotes
- Tables of contents
- Maps or diagrams
- Formulae and equations
- Tabular matter
- Documents included within documents (eg contracts, charters etc)
- Material incorporated by reference (the possibility of linking to it)

The possibility of modifying some material to conform to current style may need to be considered. For example, if what are now known as section headings take the form of sidenotes, they may need to be rendered online as section headings and appropriate annotations made. To be useful, tables of contents need to appear. These are missing from some older instruments or appear in different ways (see 3.4.5) and will need to be created in a useful and standardised way. Similarly, short titles need to be identified and standardised (perhaps by masking the definite article or introductory text) so that they can be searched and displayed for browsing purposes.

¹³ Any such changes might provide logical places to break the collection into manageable tranches or sub-projects, in the event of a staged data capture.

4.2 Metadata requirements

4.2.1 *Minimum metadata needs*

The following minimum metadata is likely to be essential:

- Title
- Year and number (possibly also regnal year, session numbers etc)
- Classification (ie Public Act/Ordinance, Local or Personal Act/Ordinance, Imperial Act, Ordinance of the Legislative Council, Ordinance of the Province of New Munster etc)
- Assent date
- Commencement date(s)
- Currency: ie in force or repealed etc
- Date of repeal etc
- Repealing instrument(s)
- Alert or warning note (to deal with any uncertainty about individual status etc)
- Any name change.

The scoping study will need to assess and confirm (probably adding to) these requirements. The sources of some of the information will need to be identified, as will the timing and method of metadata capture. This exercise will also need to be harmonised with the metadata already developed and documented for the PAL system, which includes the NZGLS metadata, and a subset of metadata requirements documented.

Note about a “Catalogue”

In the event that the project is unable to proceed or does not encompass the entire collection of enacted statutes, there is another smaller project that could be undertaken locally. This would be to list all of the statutes with sufficient status metadata to provide a useful catalogue covering the entire NZ statute book.¹⁴ Ideally this would mesh with the PAL system and website (or perhaps just be available via the website and could be relatively inexpensive to complete to a useful level, even without some of the trickier metadata such as commencement and repeal dates).

4.2.2 *Metadata problem areas*

As identified in 4.1.2, the older material will need to be more closely assessed for conformity with naming and numbering conventions (short titles and numbers) used in current New Zealand material and particularly as catered for in the PAL system. The possibility of modifying or adding some material to conform to current style may need to be considered.¹⁵ The scoping study should also identify the sources of

¹⁴ There are various tables in existence that could be used to construct some of the historical material. For example, the 1908 consolidation and 1931 reprint have detailed tables

¹⁵ A matter to further investigate with PCO will be the possibilities and limits for adding editorial type annotations to status information for online versions, perhaps to the effect that sidenotes appearing in the original are reproduced as headnotes, or “The” in a short title has been removed from contents listings unless it is part of a corporate name?

metadata such as commencement and repeal information and any ambiguity and risks in that field.

4.3 Publication

4.3.1 Integration with PAL database and website

As already stressed, the historical statutes will need to be integrated with the PAL system although it is possible that they will need to be “stored” in a separate collection for both technical and size reasons. It must be possible for PCO staff to be able to handle the historical documents in terms of being able to add or modify text or create or modify additional or missing links, and metadata etc.¹⁶ The relationship between the versions captured under the proposed project with versions generated within the PAL system and any issues that may arise in this area will need to be explored with PCO.

The captured historical material should be searchable on the PAL website and have a similar look and feel to the current material. Status information should resemble that used for other PAL material. It should be possible to mount a search over the historical material alone or over all material and ideally a range of searches by year or decade should be possible. It should also be possible to browse the historical statutes by type, title, and in alphabetical or chronological sequence, although website functionality will need to accord with the existing functional specifications developed for the PAL website unless enhancements are possible (technically and financially). The scoping study should deliver a subset of requirements for the project.

4.3.2 Formats

The source format will be XML. The source material will be used to generate HTML for public online access via the PAL website and will be made available for republishing. A PDF download facility, using the PAL publishing tools, should also be provided and the scoping study will need to address any issues in the generation and appearance of the material.

4.4 Identification and sizing of target material

The key component of the data capture task is to identify and measure the target statutes. A suggested breakdown in priority order by the NZLLA is as follows:

- 1841-1853 Ordinances of Legislative Council and Province of New Munster
- 1854-1908 Annual volumes of statutes as enacted
- 1909-1931 Annual volumes of statutes as enacted
- 1908 Consolidation
- 1931 Reprint
- 1957 Reprint¹⁷

However, for reasons of comprehensiveness and integration with the PAL system it is recommended that the following approach be taken in the scoping study:

¹⁶ Ideally PCO staff should be able to use the processes and tools set up for PAL consolidations etc.

¹⁷ Subsequent reprints were not identified in the list.

- Measure the entire collection of statutes as enacted, including the early Ordinances and 1908 consolidation¹⁸ in order to assess its size. This should also note any major shifts in structure and format that might pose difficulties or warrant some separate treatment.¹⁹
- This measurement should consist of the number of titles, number of pages, and an estimated number of characters per page.²⁰
- It will be necessary to find what the end date will be for the statutes not already captured as enacted as part of the PAL project.
- For the more recent statutes it will be necessary to assess the extent, quality and form of any existing electronic data²¹ that is available for reuse (although it may be preferable to use the same methodology used for the main body of historical statutes, rather than have to fully proof-read or heavily rework older existing data).
- Create logical batches of annual volumes to be dealt with in priority order.

Measure the 1931 and 1957 Reprints.²²

4.5 Costing

The scoping study should deliver an estimate of the full costs for capturing and publishing the enacted statutes, broken down into the core elements:

- Integration with PAL system and website: structural analysis and DTD development/modification²³
- Identification and sizing, with approximate costing for capturing and marking up each tranche (this is likely to be a bulk rate)

¹⁸ It will first be necessary to identify the most authoritative collection to use. For example are the individual volumes of early Ordinances to be used (with ink amendments) or the reprinted 1841-53 compilation of 1871? A clean set will be needed and possibly sacrificed for the project.

¹⁹ These structural shifts will also need to be identified for the initial DTD analysis and PAL integration exercise. A brief inspection of the early material detected a shift in numbering. Sections were Roman before changing to Arabic in 1867. Many short titles are given “The” or are not clear in some early ordinances but seem to have been arbitrarily set or abbreviated for titles or running heads by the printer of the day. Tables of Content issues are described above in 3.4.5. There are unstructured headings and slabs of text. Where direct amendment is used, some new text is inserted unnumbered. There are some forms but these look very similar to those of the next century! One set of financial estimates appears separately with its own title page (1851 No 15). The 1851 volume also includes a regulation. The structure is otherwise fairly constant and simple. However, it is possible that there will be a need to legislate to affect some changes to bring some features into current style or format.

²⁰ The physical counting exercise could perhaps be more economically undertaken locally using students.

²¹ There is no doubt some source data held from the earliest days of electronic publishing by the former NZ GPO and Legislation Direct, possibly starting in the late 1980s or early 1990s and this should be quantified and assessed for the initial DTD analysis and PAL integration exercise.

²² This should be a quick assessment in terms of volume and structure.

²³ It may be possible for the scoping consultant to provide a high level description of how the integration might be achieved but this component will require a highly collaborative process with PCO and Unisys. As PAL implementation partner, system architect and host, Unisys will need to undertake or project manage any integration work.

- Additional costs for manual work in problem areas (maps, graphics etc) and adding metadata (if appropriate)
- Managing capture and mark up process, including QA and acceptance testing.

4.6 Business case and project plan

It is anticipated that parts of this preliminary report and the detailed scoping study would form the core of a business case to attract funding and a sponsor for the project. However, the following additional document components need to be developed and should be part of the full scoping study deliverables:

- Analysis of alternative solutions
- Costs/Benefits
- Project plan (timing, testing and acceptance regime, costs and resources)
- Project governance, risk management and IQA arrangements.

Part 5 Scoping study: tender, process and budget

5.1 Tender

Subject to consultation with key New Zealand stakeholders (particularly with PCO) and agreement on the proposed direction of the project, it is recommended that the detailed scoping study be the subject of a public tendering process.

In order to minimise risk, it is also recommended that the following mandatory criteria be applied in advertising and selecting a suitable contractor:

- Extensive experience in analysing legislation and developing legislative DTDs
- Knowledge of the PAL project and New Zealand legislation
- Proven record in successfully delivering legislative publishing solutions including the capture or conversion of legacy data.

Although these criteria appear to be somewhat restrictive, there are a number of consultants and developers who could meet the criteria or draw together the necessary expertise.

This strategy would also enable the project to move to actual implementation (without the need for yet a further set of specifications or requirements), although this would have to be made explicit in the tender for the scoping study but with an obvious breakpoint if the scoping study does not meet expectations etc, or the project does not proceed as planned.

As indicated in 4.1.1, a further prerequisite or breakpoint would be the ability to accommodate the historical material in the PAL system. Accordingly it would be advisable to undertake the detailed scoping study in two stages and first establish the integration and technological path forward, before undertaking the work on the actual data capture.

5.2 Budget

Given the size of the analysis and documentation components of the task, an estimate of the budget for the scoping study is between \$NZ50,000 and \$70,000 for the entire task. However, if the proposed technological path and PAL integration are demonstrated to be unattainable and the full business case is unable to proceed, the initial analysis and PAL system integration and costing assessment work is estimated to cost \$20,000.
