

---

**National Library of New Zealand**

**Institutional Repositories  
for the Research Sector**

*Feasibility Study*

---

Affinity

**John Rankin**

**27 July 2005**

This report was prepared on behalf of the expert working party for the Institutional Repositories Feasibility Study.

Public

Quality review: CARMEL JOE

Release: FINAL

Created: 26 MAY 2005

Modified: 27 JULY 2005 2:12 PM

Prepared for National Library of New Zealand by Affinity Limited. All rights reserved.



# Foreword

---

“...the intellectual life and scholarship of our [research communities] will increasingly be represented, documented, and shared in digital form, and ... a primary responsibility of our [institutions] is to exercise stewardship over these riches: both to make them available and to preserve them”

—Clifford Lynch

---

i

The world wide web is so much part of our lives that we sometimes forget that the first modern browser, Mosaic, only appeared in 1993. The Web was invented to make it easier for scientists at CERN (European Nuclear Research Centre) to collaborate and share research findings, so it's no surprise that the research sector continues to be at the forefront of Web thinking and innovation. This report starts with a simple and compelling proposal—that the output of publicly-funded research, the new knowledge that such research creates, is a public good.

The big idea which this report examines is that New Zealand research institutions join with the global research community to establish a network of *Institutional Repositories*, into which authors deposit copies of their research outputs. This gives authors a way to make their research results available to anyone with internet access. Anyone can search the full texts of deposited outputs in their field of research interest, wherever these are held. By putting research outputs in a repository, authors will enhance the visibility and impact not just of their own research, but that of the whole New Zealand research sector.

Institutions face some tough issues on the way to making this idea a reality. What kinds of works do we want our repository to contain? How do we encourage authors to contribute their works? What are the copyright implications? What kind of software will we need? What if we want to restrict access to some works? Each institution needs to find its own answers to these and other questions. This report explores the feasibility of setting up a New Zealand network of institutional repositories, as essential research infrastructure for the digital age.

**Penny Carnaby**  
National Librarian





# Summary

---

“What is the case for establishing a national institutional repository framework for New Zealand and how can this best be achieved?”

An institutional repository (IR) is a set of services for storing and making available digital research materials created by an institution and its community. The National Library of New Zealand set up a steering group and expert working party with representatives from across the research sector to carry out a feasibility study and recommend actions for a national institutional repositories framework for New Zealand’s research outputs.

---

iii

## Why institutional repositories matter

---

The Digital Strategy aims to create “a digital future for all New Zealanders, using the power of information and communications technology to enhance all aspects of our lives”. It states that “developing uniquely New Zealand content is critical”. Institutional repositories contribute to this by facilitating greater access to the research outputs from universities, Crown research institutes, government agencies, other institutions, and individual researchers. They provide a window on New Zealand’s research knowledge and are a key part of the national content strategy to preserve, share and manage digital objects.

Repositories support the open access goal of transforming scholarly communication by making it easier for researchers to find and share the results of research, through free and unrestricted on-line availability. Various factors have led policy-makers and research institutions in other countries to establish repositories; for example:

1. They want to increase research impact and consider that publicly funded research should be publicly accessible. Placing copies of research outputs in an open access institutional repository lowers the barriers to, and cost of, access.



2. They see a way to improve efficiency by linking the acquisition of research outputs with existing research management processes, enabling them to capture research output data once and re-use these for multiple purposes.
3. They want to preserve digital research outputs and see a repository as a digital treasure chest for which they have a custodial responsibility.

Institutional research repositories can create positive outcomes for all interested groups. Authors gain visibility; information seekers find research more easily; institutions raise their research profile; and funders see wider research dissemination. Achieving an active and sustainable network of institutional research repositories in New Zealand requires suitable repository software installed in, or available to, all New Zealand research institutions, supporting on-line availability of most published and unpublished research outputs.

## Make institutional repositories happen

---

The proposed New Zealand institutional research repositories federation aims to *enhance the impact of New Zealand research* by providing open access to research outputs. Research outputs are the published records of research activities and the data that support the results and conclusions in such published records.

The most widely used solutions are based on a “data plus services” model with 2 key design characteristics: *unbundling* and *interoperability*. The model consists of:

1. basic institutional facilities—the data acquisition and retention level
2. added value capabilities for specific groups of users—the services level

In establishing research repositories, institutions face a number of cultural change issues. Sustaining the cultural change programme is their greatest challenge. In particular:

- ❖ Support a national framework. The distributed repository architecture enables local institutional decision-making while facilitating the emergence of national services which join up the local collections.
- ❖ Create incentives for authors to deposit their works. This appears to be the biggest issue for an institution to consider, as the value of a research repository depends on the number of authors contributing to it.

- ❖ Link to the PBRF (performance-based research funding) process. There are 2 distinct linkages to consider between institutional research repositories and PBRF—at a policy level and at an operational level.
- ❖ Build repository management capability. All projects investigated have highlighted resource constraints as a major issue. Implementation involves policy makers, library staff, and IT staff, as well as academic staff.
- ❖ Copyright must allow open access deposit. Authors are properly worried about infringing copyright arrangements with publishers and institutions need to help them comply with publishers' agreements while retaining relevant rights over their works.

The initial focus is to establish and populate research repositories in a number of institutions as rapidly as possible, starting with demonstrator projects. A research portal will enable discovery of and access to research outputs held in these demonstrators.

---

## Choose suitable repository software

---

A range of repository software options exists. The Open Society Institute Guide to Institutional Repository Software describes 9 packages meeting its criteria:

- ❖ they are available via an open source licence
- ❖ they comply with the latest version of the Open Archives Initiative metadata harvesting protocols
- ❖ they are currently released and publicly available

As with any software acquisition, the first step is for an institution to define its requirements and priorities, and mix of open and restricted access works. In choosing a solution best suited to its requirements, an institution may wish to consider and evaluate the following options.

1. A hosted service such as ProQuest Digital Commons offers the lowest barrier to entry.
2. Eprints is a low cost option for a repository aimed at open access to article pre-prints and post-prints, including digital theses.
3. DSpace is functionally richer and supports a wide range of object types, including text, sound, images and video.
4. An institution may wish to consider ARNO as an alternative to DSpace.
5. For an institution seeking a repository suitable for very large collections and having the technical expertise to support it, Fedora is a possible solution.



6. CDSware may be a suitable alternative to Fedora for large and complex collections.
7. For an institution considering multiple discipline-based repositories with a resource discovery service, i-Tor may be suitable.

All of the software options are based on the principle of *mobile code* and run in any Unix or Linux environment.

## Recommended repository actions

---

The feasibility study has found that institutional research repositories are viable and practical for the New Zealand research sector and recommends a number of actions to:

- ❖ establish a national federation
- ❖ support institutional initiatives
- ❖ adopt a common road map

# Contents

---

## 1 Introduction 1

- 1.1 What is an Institutional Repository 1
- 1.2 A repository framework for New Zealand 1
- 1.3 Report structure 3
- 1.4 Acknowledgements 4

## 2 Why institutional repositories matter 7

- 2.1 Context 7
- 2.2 Benefits of institutional repositories 12
- 2.3 Business satisfaction criteria 14
- 2.4 Strategic intent 17

## 3 Make institutional repositories happen 19

- 3.1 Programme definition 19
- 3.2 Business issues 24
- 3.3 Outline project plan 29

## 4 Choose suitable repository software 31

- 4.1 Technical solution options 31
- 4.2 Technical issues 35

## 5 Recommended repository actions 41

- 5.1 Establish a national federation 41
- 5.2 Support institutional initiatives 42
- 5.3 Adopt a common road map 43



## Appendices

APPENDIX A *Write-up from the Australian site visits*



# Introduction

---

In May 2005, the National Library of New Zealand set up an expert working party with representatives from across the research sector to investigate the feasibility of establishing a national institutional repositories framework for New Zealand's research outputs. This document sets out the working party's findings.

## 1.1 What is an Institutional Repository

---

An institutional repository (IR) is a set of services for storing and making available digital research materials created by an institution and its community—a digital collection of the community's research output<sup>1</sup>. IRs are becoming an essential tool for scholarly communication in the digital age. They can form part of a larger national, regional and global system of repositories, indexed in a standardised way and searchable using one access interface. They can also provide a foundation for new models of scholarly publishing and can be linked to related services such as eLearning, on-line journals and print-on-demand publishing.

## 1.2 A repository framework for New Zealand

---

The National Library is fostering the creation and launch of a work programme to improve access to New Zealand's research outputs, by collaborating with institutions to stimulate the setup of research repositories. The initial investigation stage answers the following question:

“What is the case for establishing a national institutional repository framework for New Zealand and how can this best be achieved?”

---

<sup>1</sup> <http://www.arl.org/newsltr/226/ir.html>



The Library set up a steering group with representatives from across the research sector. This steering group gave an expert working party the task of carrying out a feasibility study to explore issues and options, and recommend a way forward. The working party's terms of reference were to:

- ❖ identify the issues, including challenges to be resolved and opportunities to be exploited
- ❖ consider institutional repository architectures and models available internationally
- ❖ make recommendations on architectures and models most suitable for New Zealand
- ❖ make recommendations for developing a national framework for New Zealand institutional research repositories

The feasibility study seeks to reach agreement among stake-holders on the repository programme's strategic context, definition, objectives, and scope. It considers options for repository architectures and for a national repository framework. It explores a range of topics, including:

- ❖ what is an institutional repository
- ❖ what might an institutional repository contain
- ❖ what might an institutional repository *not* contain
- ❖ what advantage is there to a university or other research institution in setting up an institutional repository
- ❖ what advantage is there to an author in depositing academic work in an institution's repository
- ❖ what options are available to independent researchers or those in non-research organisations
- ❖ how much does it cost an institution to set up and operate an institutional repository
- ❖ what off-the-shelf software tools can be used to develop an institutional repository
- ❖ what are the main issues that institutions and consortia need to address in implementing an institutional repository
  - ◆ organisational, administrative and cultural issues
  - ◆ content policies and accession, retention and preservation
  - ◆ researcher engagement and participation
  - ◆ technical options and infrastructure issues
  - ◆ intellectual property issues and access rights
- ❖ where do we go from here

To explore these and other topics, in June 2005 the working party went on a fact-finding visit to Australia to see the repository projects at the following institutions:

- ❖ University of Southern Queensland

- ❖ University of Queensland
- ❖ Macquarie University
- ❖ Australian National University
- ❖ National Library of Australia
- ❖ University of Tasmania
- ❖ Tasmania State Library
- ❖ Monash University

Following these visits, the working party held a number of workshops to formulate the proposals set out in the report. It presented the report in draft form to the steering group at the end of June, incorporated their comments and suggestions, and issued the final version in July. Appendix A contains the *Write-up from the Australian site visits*.

The investigation used a fully open process. All working party discussions and the findings from the Australian visits are documented on a public Web site<sup>2</sup>. All participants had the opportunity to read, comment on, add to, or edit the contents of all pages, or to add their own pages. Everyone can see, and respond to, everyone else's comments. The Web site is a wiki-based open editing system, hosted by the Ministry of Education. All contributions are licensed under a Creative Commons ShareAlike licence<sup>3</sup>.

### 1.3 Report structure

The report uses the following structure.

It first positions institutional research repositories in the wider national and global environments, and presents the rationale and business case for investment—“Why institutional repositories matter” (page 7).

It then defines the programme of work currently envisaged and examines the cultural changes required to establish a repository framework—“Make institutional repositories happen” (page 19).

It describes the kinds of technology changes that will be required and examines the software options available—“Choose suitable repository software” (page 31).

Finally, the report proposes actions to take the programme forward to the next stage—“Recommended repository actions” (page 41).

<sup>2</sup> <http://wiki.tertiary.govt.nz/~InstitutionalRepositories>

<sup>3</sup> <http://creativecommons.org/licenses/by-sa/2.0/>



The report structure reflects the approach used. It recognises that different institutions will have different repository needs and priorities. It therefore takes the approach of recording relevant facts, issues and options as the working party sees them, allowing each institution to draw conclusions appropriate to its own situation. Where there is a need for sector-wide agreement, such as in the area of standards, this is identified.

## 1.4 Acknowledgements

---

The convenor and chair of the steering group is Penny Carnaby, Chief Executive and National Librarian, National Library of New Zealand. The group has the following representatives.

### *Representatives from the New Zealand Vice-Chancellors' Committee*

On behalf of Research Directors

Jonathan Hughes, Senior Policy Analyst (Research), New Zealand Vice-Chancellors' Committee

On behalf of CONZUL (Council of New Zealand University Librarians)

Brian Flaherty, Digital Services Manager, University of Auckland

Gail Pattie, University Librarian, University of Canterbury

Keith Webster, University Librarian, Victoria University of Wellington

On behalf of IT Directors

Alan Dempster, Director of Information Technology Services, Victoria University of Wellington

### *Representatives from the research communities*

On behalf of the Sciences

Dr Hamish Campbell, Senior Scientist, Institute of Geological & Nuclear Science

On behalf of the Humanities

Dr Brian Opie, Senior Lecturer, Victoria University of Wellington

On behalf of Institutes of Technology and Polytechnics of New Zealand

Trish Brimblecombe, Associate Dean, Whitireia Polytechnic

On behalf of the Crown Research Institutes

Marlene Wilkinson, Information Services Manager, Industrial Research Limited, currently the CRI library managers convenor

*Representatives from government*

Ministry of Education

Murray Leach, Senior Manager, Tertiary Information Systems and Sector Liaison

Ministry of Research, Science and Technology

Louise Munro, Senior Advisor

Tertiary Education Commission

Clare Ward, General Manager Steering and Investment

National Library

Steve Knight, Programme Architect, National Digital Heritage Archive, National Library of New Zealand

*Working party*

Brian Flaherty, Digital Services Manager, University of Auckland

Steve Knight, Programme Architect, National Digital Heritage Archive, National Library of New Zealand

Murray Leach, Senior Manager, Tertiary Information Systems and Sector Liaison, Ministry of Education

John Rankin, Project Consultant, Affinity Limited

Keith Webster, University Librarian, Victoria University of Wellington





## Why institutional repositories matter

---

Research is essential to New Zealand's social, environmental and economic wellbeing. Knowledge-based innovation and transformation have enhanced and enriched all our lives. The innovation cycle requires a stream of new ideas and new knowledge which are absorbed and combined with existing knowledge to create innovation. This relies on a vibrant research community that can effectively disseminate research outputs to a wide range of interested groups. It requires mechanisms that facilitate knowledge sharing and collaboration. It gives young people more options for taking up a research career in New Zealand.

Open access institutional research repositories are a way to disseminate research results quickly and efficiently, in a research environment in which the world wide web plays an increasingly important role.

### 2.1 Context

---

This section places institutional research repositories in their strategic context, both nationally and globally. Institutions establishing research repositories do so as part of their wider information management framework. Repositories are becoming a fundamental and essential part of the research infrastructure.

#### 2.1.1 Repositories are part of the Digital Strategy

The *Digital Strategy: Creating a Digital Future for all New Zealanders*<sup>1</sup> states:

The Digital Strategy is about creating a digital future for all New Zealanders, using the power of information and communications technology (ICT) to enhance all aspects of our lives. It is a strategy

---

<sup>1</sup> <http://www.digitalstrategy.govt.nz/>



for ensuring New Zealand is a world leader in using information and technology to realise our economic, environmental, social and cultural goals. The Digital Strategy isn't just about technology ... it's about people and their ability to connect to the things that matter to them.

The strategy rests on 3 pillars: *content* (information we can access that can enrich the quality of our lives); *confidence* (the skills to use ICT and a secure environment in which to do so); and *connection* (getting access to and using ICT). For *content*, the strategy states (**emphasis added**):

Developing uniquely New Zealand content is critical to New Zealand's digital future.

*Content* is information which is made available through digital networks. *Information* is a broad concept that encompasses national heritage collections, government information, Maori language resources, **research databases**, literature, **history** and entertainment, as well as relationships that can be conducted through on-line facilities (e.g. **e-learning**, on-line GST returns or Internet banking).

*Goal:* To unlock New Zealand's stock of content and provide all New Zealanders with seamless, easy access to the information that is important to their lives, businesses, and cultural identity.

Institutional repositories contribute to New Zealand's digital strategy and in particular to the *Content Goal* by facilitating greater access to the research outputs from universities, Crown research institutes, government agencies, other institutions, and individual researchers. The aim is to provide a window on New Zealand's research knowledge that will:

- ❖ support open access to the results of New Zealand research
- ❖ give authors and institutions control over access to their works
- ❖ preserve digital research outputs as an essential record of scholarship
- ❖ lay a foundation for exploring new scholarly publishing options
- ❖ enhance and enrich scholarly communication
- ❖ join and contribute to the global IR community

The momentum for these types of digital repositories is building in universities and other research institutions on every continent. They represent an important new role for academic libraries in the 21<sup>st</sup> century. They are key part of New Zealand's national content strategy to preserve, share and manage digital objects.

### 2.1.2 Repositories are a global success story

A number of countries are actively pursuing digital repository initiatives, at both the institutional and national levels. The United States is the most active with 124 known repositories. Other examples include:

*UK Select Committee on Science and Technology*

In its tenth report for session 2003–04, the Select Committee “recommends that all UK higher education institutions establish institutional repositories on which their published output can be stored and from which it can be read, free of charge, on-line”<sup>2</sup>. It further recommends that “funders mandate their funded researchers to deposit a copy of all of their articles”. It notes that “[s]et-up and running costs are relatively low, making institutional repositories a cost-effective way of improving access to scientific publications”. The UK has 55 known repositories.

*Canada ARL Institutional Repository Pilot Project*

This is a Canadian initiative aimed at implementing institutional repositories initially at 9 Canadian universities. Its theme is *innovation in scholarly publishing*. The intent is that “Institutional Repositories centralize, preserve, and make accessible the knowledge generated by academic institutions. IRs also form part of a larger global system of repositories, which are indexed in a standardized way, and searchable using one interface, providing the foundation for a new model of scholarly publishing.”<sup>3</sup> Canada has 27 known repositories.

*Australia Research Information Infrastructure*

The Australia Department of Education, Science and Training has provided project-based funding “to enhance the creation and management of information, improve access to information resources, and facilitate the discovery and dissemination of new information to researchers and institutions”.<sup>4</sup> This covers 4 projects collectively called FRODO (federated repositories of digital objects): Meta Access Management System (MAMS); Towards an Australian Partnership for Sustainable Repositories (APSR); The Australian Research Repositories On-line to the World (ARROW); and Australian Digital Theses Program Expansion and Redevelopment (ADT). Australia has 17 known repositories.

*DAREnet: Worldwide access to Dutch academic research results*

DAREnet gives digital access to academic research output in the Netherlands. Its *Cream of Science* website “contains over 25,000 publications by leading Dutch scientists and researchers, available full-text for fellow academics, educational institutions and interested members of the public. Within just a day of its launch on May 10<sup>th</sup>, it

<sup>2</sup> <http://www.publications.parliament.uk/pa/cm200304/cmselect/cmsctech/399/39903.htm>

<sup>3</sup> <http://www.uottawa.ca/library/carl/projects/ir/>

<sup>4</sup> <http://www.dest.gov.au/Ministers/Media/McGauran/2003/10/mcg002221003.asp>



registered half a million hits”<sup>5</sup>. As of May 2005, DAREnet had harvested over 40,000 digital records from the Institutional Repositories of 16 institutions. In addition: “DAREnet guarantees free and open access to all content for everyone. No restrictions.” Holland has 16 known repositories.

#### *Wellcome Trust insists on open access*

After 1 October 2005, all new Wellcome Trust grant recipients will be required to post an electronic version of peer-reviewed research articles arising from the funded research in an open-access repository. The Trust notes that: “If you want to read an exciting new research paper on the internet, chances are you can’t—unless you or your work place have paid a large fee. The research may have been funded by taxpayers’ money, and the scientists who wrote the paper will have done so for free, but the paper’s publisher holds all the aces: the copyright, the access and the chance to charge you to read the text. Contrast this to the human genome sequence. Anyone, anywhere in the world can search, browse and explore the sequence ... at free websites such as Ensembl. If the access to research papers were similarly unfettered, ... it is not just scientists who would benefit: doctors, teachers, students and the public could find and read about the latest scientific and medical discoveries.”<sup>6</sup>

There is an opportunity for New Zealand institutions to consider and take advantage of relevant experiences gained from these and many other overseas initiatives, some large and active, others less so. A common theme in all successful overseas models is the co-evolution of local, institution-based repositories and national repository policy frameworks. The links between repositories, such as resource discovery or alerting services, and the services connecting repositories to information seekers, are as important as the repositories themselves.

### **2.1.3 Repositories enable open access**

The open access movement is part of a global trend to transform scholarly communication by making it easier for researchers to find and share the results of research, through free and unrestricted on-line availability. Open access contributions must satisfy two conditions:<sup>7</sup>

1. The author(s) and right holder(s) of such contributions grant(s) to all users a free, irrevocable, worldwide, right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any dig-

<sup>5</sup> <http://www.darenet.nl/en/page/language.view/home>

<sup>6</sup> [http://www.wellcome.ac.uk/doc\\_WTX024858.html](http://www.wellcome.ac.uk/doc_WTX024858.html)

<sup>7</sup> <http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html>

ital medium for any responsible purpose, subject to proper attribution of authorship..., as well as the right to make small numbers of printed copies for their personal use.

2. A complete version of the work and all supplemental materials, including a copy of the permission as stated above, in an appropriate standard electronic format is deposited (and thus published) in at least one on-line repository using suitable technical standards (such as the Open Archive definitions) that is supported and maintained by an academic institution, scholarly society, government agency, or other well-established organisation that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.

Open access asserts that the knowledge embodied in deposited works is a public good. A global network of connected research repositories and resource discovery services are essential pieces of open access infrastructure. The University of Southampton's worldwide registry of OAI-compliant open access archives<sup>8</sup> lists 434 repositories containing some 3.5 million records in 38 countries, as at 17 June 2005. Currently, New Zealand has *no* OAI-compliant repositories, although New Zealand universities are intending to make metadata about PhD theses available to the Australasian (formerly Australian) Digital Theses project.

Open access records cover the full text of pre-prints, post-prints of peer-reviewed journal articles, and dissertations. Repositories are not just for open access works; some materials may have access restrictions placed over them for legal, commercial or other reasons. Depending on the institution's policy, a research repository may contain any mix of open access records, metadata-only records, private records, and non-textual materials. Many overseas institutions see open access as a key to enhancing scholarly communication. Some institutions and scholars have also used the open access approach to develop new models for scholarly publishing. Consideration of the implications of open access for scholarly publishing (as distinct from communication) is beyond the scope of the current study. Scholarly publishing is one of many services that can be offered over the research repository data layer.

On 28 June 2005, the 8 UK Research Councils, under the umbrella of Research Councils UK, announced their proposed position<sup>9</sup> on access to research outputs. They have proposed to make it mandatory for research papers arising from Council-funded work to be deposited in openly available repositories at the earliest opportunity. They describe the proposals as a starting point and are seeking the views of all parties involved in the debate. JISC (joint information systems committee) has published some common questions and answers about open access<sup>10</sup>.

---

<sup>8</sup> <http://archives.eprints.org/eprints.php?action=browse>

<sup>9</sup> <http://www.rcuk.ac.uk/access/index.asp>

<sup>10</sup> [http://www.jisc.ac.uk/issue\\_qaopen.html](http://www.jisc.ac.uk/issue_qaopen.html)



## 2.2 Benefits of institutional repositories

This section sets out the main factors leading policy-makers and institutions to establish research repositories. These factors all centre on enhancing the research culture by strengthening the research infrastructure. Different institutions give these factors different priorities. The section focuses on the benefits to institutions of establishing and supporting repositories of research outputs. It gives an institutional view, rather than an individual view.

### 2.2.1 Increase research impact

The core argument is that publicly funded research should be publicly accessible, thereby increasing its potential impact. The object of open access is to maximise research *impact* by maximising research *access*. Placing copies of research outputs in an open access institutional repository lowers the barriers to, and cost of, access. In his article *Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age*, Clifford Lynch identifies a number of benefits, including<sup>11</sup>:

- make it easier for scholars to disseminate their work, both in traditional forms such as electronic journal articles and new forms such as virtual scholarly monographs
- encourage exploration of new forms of scholarly communication that take advantage of the new capabilities offered by the digital medium for the exposure, discovery and re-use of distributed digital collections, especially across disciplines
- relieve researchers of the time-consuming burdens of exercising stewardship over digital research content, maintaining personal web sites to distribute that content, and ensuring the long term preservation of content held only in digital form

Evidence is accumulating that self-archived research articles are cited more often than those published in traditional journals<sup>12</sup>. An institutional repository makes research outputs more visible to more people, thereby potentially increasing the impact of that research. A combination of full text indexing and metadata that describe the content in a standardised way ensures research outputs (and any associated digital objects) will be readily found, even as the volume of material grows. To do nothing means falling further behind. As the volume of published research grows, research that is hard to access becomes less visible. So unless institutions take active steps to make their research outputs easier to discover, the impact of their research will steadily erode.

<sup>11</sup> <http://www.arl.org/newsltr/226/ir.html>

<sup>12</sup> <http://www.dlib.org/dlib/june04/harnad/06harnad.html>

### 2.2.2 Preserve digital research outputs

A repository is an institution's digital treasure chest. Digital research outputs become digital research *assets*<sup>13</sup> when the institutional custodian has processes in place to support ongoing content management across the entire asset life-cycle:

- ❖ establish—what to do to set up and run the repository
- ❖ manage—processes for supporting content providers and information seekers
- ❖ co-ordinate—mechanisms for federating with other local and global repositories
- ❖ sustain—principles for the long term future, including
  - serving the needs of all its user communities
  - ensuring the assets are discoverable and available
  - ensuring the content endures (e.g. file format migration)
  - retention and disposal policies
  - institutional commitment

Institutions have recognised that to achieve this means treating research repositories as an essential component of the research infrastructure, rather than relying on the goodwill of individual departments and scholars. They have also recognised the importance of ongoing stewardship over the datasets associated with the research outputs. In many disciplines, the most valuable part of older research is the data collected, which are often subject to re-interpretation in light of advances in the field of study.

Thus many institutions with active repository programmes encourage scholars to deposit their older works as well as new works; they help scholars who are about to retire to preserve their most valuable research outputs and datasets; and every time someone asks for a thesis currently held on paper, they digitise it (subject to copyright agreement). Institutions also have policies and practices in place to define and manage access rights over research works.

### 2.2.3 Harness institutional efficiency

Institutions have recognised the potential to link the acquisition of research outputs and their descriptive metadata with existing research management processes, enabling the institutions to capture research output data once and re-use these for multiple purposes. Many governments have a reporting process analogous to that for the New Zealand performance-based research funding or the Australian research quality

---

<sup>13</sup> an asset earns its keep by doing useful work



framework. Many of the research output descriptive data elements are required for both performance reporting and resource discovery.

Several approaches have been used, depending on institutional scale:

- ❖ create one research output data capture process that feeds both an institutional repository and a research management database
- ❖ use data harvested from either a research management database or an institutional repository to populate the other system
- ❖ create a reporting service that draws data from an institutional research repository to meet government reporting needs

Once an institution chooses to expose certain data items in an open form over the Web, anyone may harvest from these. We envisage that as open standards compliant, open access repositories become more widely installed and populated, central government bodies will be able to harvest a wide range of data needed for policy analysis automatically. This has the potential to lower the compliance costs for all parties involved. Similarly, researchers can harvest data from their institution's repository to support a range of currently-inefficient activities, such as:

- ❖ automatically maintain an up-to-date ePortfolio or Web résumé
- ❖ generate information to support research funding applications
- ❖ fulfil requests for reprints of journal articles and other publications

## 2.3 Business satisfaction criteria

---

This section sets out the expectations of the various groups with an interest in the services of an institutional research repository. It provides a view of how each community sees the benefits set out in the business case—"what's in it for me". It shows that institutional research repositories can create positive outcomes for all interested groups. In essence, anyone with internet access will be able to read authors' research outputs, without restriction. Anyone will be able to search the full text of all outputs in their field of research interest.

### 2.3.1 Authors gain visibility

Communicating research results to their peers is the main reason scholars publish their work. In other words, the goal of publishing research is to make an impact on the field. Authors expect an institutional research repository will allow them to broadcast the results of their research more widely, more quickly. Once an author deposits a

research output in an institutional research repository, it becomes possible to deliver usage information at minimal cost. This can include:

- ❖ number of times the research output is accessed, e.g. to view the abstract
- ❖ number of times the output is downloaded
- ❖ visitor country of origin analysis

The presentation at the University of Tasmania noted that papers published in open access journals are cited 2–5 times more often than those published in traditional journals, depending on the discipline. Papers that are downloaded are also more likely to be cited. The implication for repositories is that authors will expect deposited works to be more visible and have more impact. In fast-moving fields, early and wide exposure of article pre-prints can be especially important.

While technically astute authors may know how to ensure papers placed on their personal web sites achieve high search rankings, many do not. Authors lacking the knowledge or interest in this topic will expect the institutional repository to take care of the necessary details.

Abstracts have always been easier to access than the full text of articles, initially through abstracting services, such as ChemAbs, and now through journal publishers' Web sites. In many cases, access to the online abstract is free, but retrieving the article incurs a fee. By eliminating these access fees, repositories may reduce the incidence of researchers citing articles based on their reading the abstract only.

### 2.3.2 Information seekers find research more easily

As the volume of published research grows, so the amount of effort information seekers are willing to expend to locate relevant publications diminishes. Placing properly-described copies of research outputs in an institutional repository creates new discovery pathways, including search engines, resource discovery services, and discipline portals. Information seekers are becoming less willing to spend time looking for research outputs that are hard to find. The searcher's assumption is often that "Google knows"—*search the Web* has replaced *read the manual*. The corollary of the assumption that Google (or any other resource discovery service) knows the answer is that any information not part of the answer will probably stay undiscovered.

Once a federation of linked research repositories is established, with content properly described, information push services become possible in the form of automatic alerts. An information seeker tells the discovery service subjects of interest and that service delivers regular bulletins to the seeker's news reader or browser software. The discovery service can draw content from local, national or global repository federations. The information seeker saves time by not having to go looking in several places and by only seeing works matching the specified criteria.



Specialist public search services, such as Scirus, index articles deposited in open access repositories, whereas most commercial indexing services, such as Biological Abstracts, cover scholarly journals only and (up until now) do not index repositories.

### 2.3.3 Institutions raise their research profile

One of the measures used in international research rankings is a publication and citation rating—using citation as a convenient surrogate for research impact and quality. A number of converging trends are leading to the institution's research repository as an increasingly important way of raising research profile:

- ❖ search engines such as Google give a higher ranking to the open-access copy of an article
- ❖ the more often a paper is downloaded, the more likely it is to be cited
- ❖ measures such as the Thomson ISI *Web Citation Index* are emerging, which report citations to and from institutional repositories

In some fields, such as medicine, deposit in an open access repository is becoming either the accepted norm or a mandatory requirement set by the funders. The Web's ability to reveal otherwise hidden content means that financially challenged institutions in small remote countries can, in their areas of research specialisation, aspire to appear on an equal footing with any institution in the world.

Institutions without repositories may see their research profile erode. Institutions with well-managed and sustainable research repositories will expect to find new ways of using the content to enhance their local and global profiles, and strengthen the links between teaching and research. Examples include:

- ❖ honours or graduate students doing research projects being able to view and cite student research papers from past years
- ❖ prospective staff, graduate students or post-graduate researchers being able to examine the kinds of research being produced in their fields of study
- ❖ revealing new connections between disciplines and institutions, potentially leading to new opportunities for collaboration

### 2.3.4 Funders see wider research dissemination

Researchers have always built on the work of those who have gone before them (“standing on the shoulders of giants”), so rely on access to past outputs. Funders faced with choosing which research projects to fund want to be confident that the outputs will have the widest possible accessibility and hence impact. For example, when funding software

research, the Mellon Foundation encourages recipients to release any programs created under an open source licence, meaning that anyone can use, modify, improve or redistribute the source code without incurring a licence fee. Funders expect an institutional research repository will ensure the outputs of funded research receive wide circulation, at least cost, thereby making the funding process more transparent and the research more visible.

Funders, whether public or private, are taking a range of approaches to encouraging institutions to establish and maintain research repositories. These include various combinations of:

- ❖ funding repository establishment and initial operation
- ❖ mandating that grant recipients deposit copies of outputs on an open access repository
- ❖ hosting discipline-based repositories on behalf of institutions
- ❖ encouraging (and providing funding for) authors to deposit
- ❖ noting the potential while providing no active financial support

---

## 2.4 Strategic intent

---

This section sets goals for what the repository environment may be expected to look like in future to achieve these benefits and recognises that building repository capability is critical to a successful outcome.

### 2.4.1 Beyond 2007: repositories are business as usual

To achieve an active and sustainable federated network of institutional research repositories in New Zealand, the longer term goals are:

1. OAI-compliant (Open Archives initiative) research repository software is installed in, or available to, all New Zealand research institutions, supporting on-line availability of most published and unpublished research outputs
2. a national research resource discovery service harvests metadata from institutions' research repositories, enabling anyone with an internet connection to find and view the full text of most of New Zealand's publicly funded research outputs
3. data collected as part of the output deposit process are also used to meet the institutions' requirements for reporting to government on their research outputs
4. federated authentication and authorisation grants access to protected works held in research repositories, within and between institutions, including trans-Tasman access



5. institutions see research repositories as essential and valued parts of the research infrastructure and are confident that their repositories are sustainable long term; deposit has become routine

The first step to achieving these goals is to establish demonstrator repositories in a small number of research institutions.

### 2.4.2 Create centres of repository expertise

Establishing successful, well-populated research repository demonstrators is only the first step. Building on these initial demonstrators means addressing issues such as:

- ❖ drawing on the content store to support other services, such as e-journals, collaboratories, virtual monographs, e-learning, or subject portals
- ❖ adding federated authorisation and authentication services to control access to restricted works
- ❖ addressing the need for long term content preservation, format migration, and access

To realise these and other opportunities requires systematically building capability and experience in repository development, management, and operation. The Australian model of funding projects in a number of institutions, both to deliver tangible results and build capability, is achieving its desired outcomes. A particular challenge for the New Zealand research sector is how to build and maintain the capability needed to make wise decisions in what is still an immature and evolving area, and then to follow through with implementation. This capability needs to be in the institutions, close to the places where research outputs are created and used. However, this makes smaller institutions particularly vulnerable to loss of key staff. We envisage a number of potential solutions to this problem:

1. provide incentives for larger institutions to develop repository centres of excellence, with the ability to offer repository services to others on a fee-for-service basis
2. provide incentives for smaller institutions to collaborate in establishing shared repository services
3. investigate the possibility of providing a sustainable and cost-effective New Zealand research repository management service
4. outsource to a specialist overseas repository hosting service and focus solely on the research outputs deposit process
5. collaborate more closely with Australia to build trans-Tasman support teams that can help with planning, set-up and training

Implementing these and other solutions successfully and sustainably will require some form of project-based funding similar to the Australian model.

## Make institutional repositories happen

---

Establishing an effective and active repository community will require significant effort, both locally within institutions and nationally to develop complementary shared services. Institutions will face a number of cultural change management challenges and information policy questions. Effectively managing the cultural change presents institutions with a major challenge and will require their ongoing commitment.

### 3.1 Programme definition

---

This section sets out a *strategic roadmap* for establishing a research repository framework in New Zealand. It defines an over-arching context within which individual institutions can set priorities that meet their own local needs *and* reflect national and international needs.

#### 3.1.1 Purpose

The proposed New Zealand institutional research repositories federation aims to *enhance the impact of New Zealand research* by providing open access to research outputs. It has the following objectives:

1. Develop a national research repository infrastructure by setting up, populating and linking individual repositories
2. Stimulate development of services that draw on research information made available through the repository infrastructure
3. Promote submission of research outputs to, and use of content from, the repositories
4. Grow a community of practice to sustain ongoing development of the repository infrastructure and services
5. Provide a window that gives open access to New Zealand's digital research outputs



### 3.1.2 Principles

*Digital content* Support and promote creation and preservation of digital research outputs.

*Discovery* Support and promote discovery of related digital research outputs across content collections and disciplines.

*Collaboration* Support and promote national and international collaboration to enhance management and interoperability of digital content.

*Connection* Support and promote connections between digital research, learning and administration services.

*Standards* Support and promote use of relevant open standards, such as the OAI (Open Archives Initiative) Protocol for Metadata Harvesting<sup>1</sup>.

*Software* Support and promote use of existing systems as proven tools for creating OAI-compliant repositories.

*Diversity* Support and promote software diversity within an OAI-compliant framework—one solution does not fit all.

*Evolution* Support and promote “learn by doing” and use of demonstrators to build shared experience.

*Authors* Support and promote processes that make it easy and desirable for authors to submit their research outputs.

*Cost* Support and promote approaches that minimise the costs to institutions of participation and compliance.

*Sustainability* Support and promote solutions which can be readily looked after, such as the ability to scale up as usage grows.

*Preservation* Support and promote choices for digital object formats that enable long term preservation of and access to content.

### 3.1.3 Scope aspires to include all research outputs

Research outputs are the published records of research activities (for example, monographs or articles published in academic journals or conference proceedings) and the data that support the results and conclusions in such published records.

---

<sup>1</sup> <http://www.openarchives.org/>

An Institutional Repository may contain any type of scholarly digital research output, including but not limited to:

- ❖ journal articles (some or all with supporting data)
- ❖ article pre-prints
- ❖ theses and dissertations
- ❖ books and monographs
- ❖ conference papers and public lecture series
- ❖ working papers and committee reports

Research outputs will become increasingly hypertextual and contain multiple media—rich and complex collections of linked digital objects.

An IR may contain a wide range of digital objects associated with a research output, such as image, video and sound files, as well as text. Text files may include special mark-up to represent equations, chemical formulæ, music, or rare language sounds.

An institution may choose to associate supporting materials (such as datasets, cartographic work, models, and simulations) with some research outputs, or may set an articles-only policy.

IR content is mostly open access and toll-free, viewable by anyone on the Web, but may also contain content to which access controls apply.

Any institution carrying out or commissioning research may become part of the repositories federation.

An institution may operate one institution-wide repository or several discipline-based or other IRs. It may join with others to operate a community IR.

It contains electronic copies of research outputs, and may also include pointers (links) to copies (or representations) held elsewhere.

It aspires to hold all an institution's research outputs; it may also contain research outputs of non-aligned scholars (those not based at an institution).

It sets rules for supported file formats and standards, such as XHTML, other XML-based standards such as TEI, and PDF. It may support specialist formats such as L<sup>A</sup>T<sub>E</sub>X.

It conforms to the Open Archives Initiative Protocols for Metadata Harvesting. Each output has a corresponding metadata record that an external process can harvest. Such harvesting may be general (e.g. all open access outputs) or specific (e.g. theses only).

It can be searched using any Web browser on any operating system. Some documents may be viewed in any web browser; others may require use of special viewer software, such as a PDF reader or a special mark-up translator.

It holds an institution's digital research outputs in perpetuity.



### 3.1.4 Scope is not...

An institutional repository is not a repository for digital learning objects, although deposited research outputs may be used as learning objects. Learning objects require additional services and metadata, not provided by repositories geared to public discovery of research outputs.

It is not solely an archive of journal articles. Supporting research datasets and analysis algorithms may be deposited along with published articles, or dataset metadata may be provided. Potential future scope extensions may affect initial planning decisions.

It is not limited to the outputs of publicly-funded research, although the initial focus is likely to be on publicly-funded research.

It is not limited to the outputs of academic researchers; private scholars and others may also deposit their research outputs perhaps to a national repository.

It is not structured around document types; not just theses or just journal articles. There will not be a “New Zealand Digital Theses” repository; there may be a Digital Theses portal to several repositories.

It usually does not contain documents in proprietary file formats such as doc.

It is not just a repository of open access outputs; some outputs may have access restrictions placed over them.

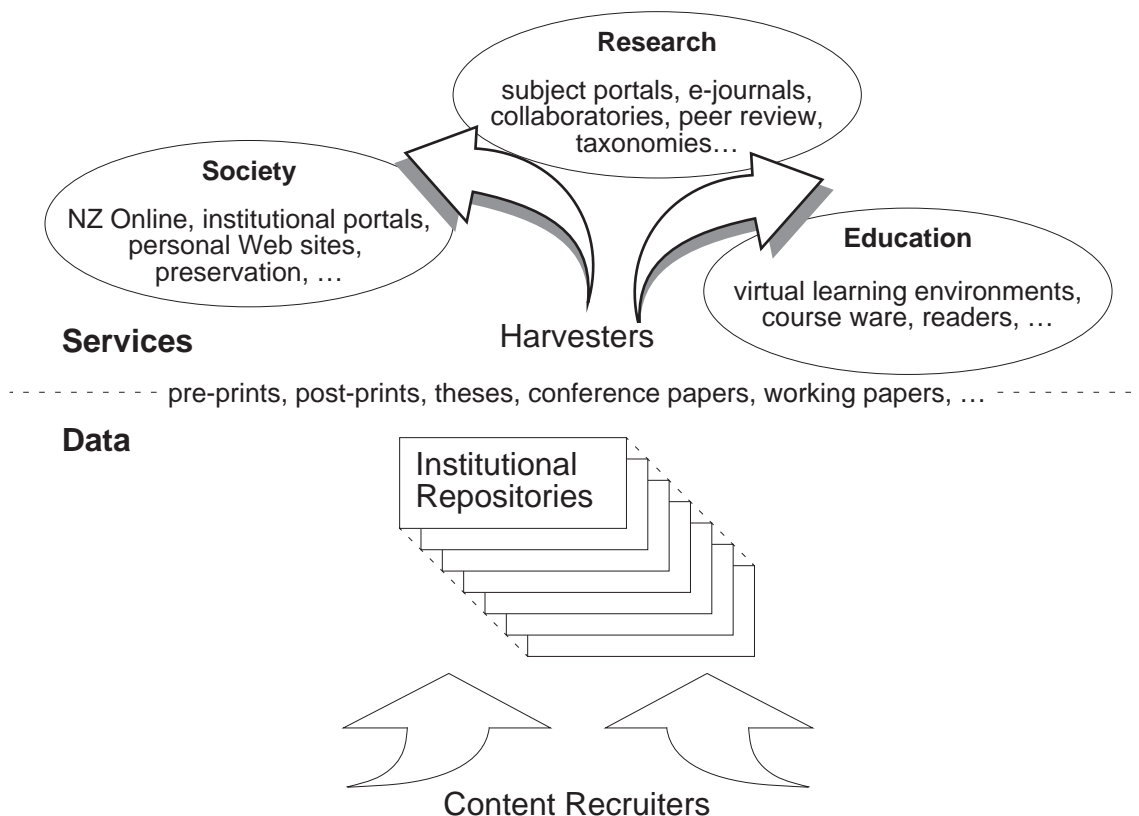
It is not a discrete, institution-specific undertaking; the full potential is achieved when “the network infrastructure” enables services spanning multiple repositories, such as discipline portals.

### 3.1.5 Architecture

The proposed architecture uses the “data plus services” model that repositories commonly use internationally. This architecture has 2 key design characteristics: *unbundling* and *interoperability*. It consists of:

1. basic institutional facilities—the data acquisition and retention level
2. added value capabilities for specific groups of users—the services level

The following diagram shows the unbundling of the overall architecture into discrete, independent parts, and interoperability between parts through use of open standards. The power of this architecture is that parts can be added, changed, or removed independently of one another (for example, to exchange one repository storage system for another or to offer a new service); and the distributed parts can be combined to create one grid-like data resource. The services shown are intended as examples of the possibilities that may be offered.



The following edited extract from the DARE Data and Services Model description, on which the above diagram is based, summarises the rationale for this widely-used approach<sup>2</sup>.

*Data Level*

The data level is where the infrastructure is set up and maintained. Every ... institution establishes a repository that stores research information from that institution and keeps it available for use (or reuse) according to uniform international standards: working papers [and] pre-prints, dissertations, research reports, datasets, conference reports, multimedia material etc., including the corresponding metadata.

*Services Level*

Basic material from the data level can be used to develop services providing added value for [researchers], students, universities, [funders] and other interested parties. The possibilities are numerous, including current or new services (management information or updating of résumés respectively); services the institutions themselves wish to offer, either singly or in association with others (subject portals); or services provided by third parties such as publishers (e-journals). Supply and development can take place at the

<sup>2</sup> [http://www.surf.nl/en/download/DARE\\_data\\_services\\_model.pdf](http://www.surf.nl/en/download/DARE_data_services_model.pdf)



local level (digital display) as well as at national level (national research information service) or even internationally (virtual communities). The most appropriate level of co-operation can be decided upon depending on the situation in question: most important is that institutions retain control of their own information.

This model also allows the repositories federation to take advantage of services offered by other standards-based service architectures, such as authentication and authorisation services, eLearning frameworks, long term preservation, or research management and reporting systems. It allows different repositories to focus on different types of objects, and allows new object types to be added from time to time as needed. The model supports any kind of digital object and its associated metadata.

---

## 3.2 Business issues

---

This section identifies the business issues for institutions to consider when setting up and managing a research repository. Addressing these cultural issues is critical to the long term sustainability of a research repository. This means setting appropriate expectations, identifying business champions, gaining commitment of resources (time and money), and choosing the appropriate policies. Sustaining the cultural change programme is the greatest challenge in implementing a repository successfully. Many repository projects have struggled to move beyond the demonstrator stage (a few hundred records and static) into “business as usual” (many thousand records and growing).

### 3.2.1 Support a national framework

The distributed repository architecture enables local institutional decision-making while facilitating the emergence of national services which join up the local collections. There are a number of repository services which have a national component, including:

- a resource discovery and alerting service
- a research repository in which non-aligned scholars may choose to deposit their research outputs
- a digital object preservation service (although this may also be done at an institution level)
- federated authentication and authorisation (although this has wider application than just repository access)
- co-ordinating standards across the sector, for example so that the same terms are used to describe similar objects

At a minimum, institutions need to ensure that the software they choose to run their research repositories is OAI-PMH compliant, to enable harvesting of exposed metadata. They also need to ensure that their identity (directory) services become SAML-2 (security assertion mark-up language) compliant, for future support of federated access to restricted works. See “Access control over protected works” on page 33.

The national services complement and add value to, rather than compete with, other services. For example, national resource discovery can offer services such as federated alerting (alert an information seeker to new materials across several repositories). They can also deliver services which would otherwise be uneconomic or are not part of any other institution’s mission, such as a repository for non-aligned scholars (independent researchers). Establishing effective national services relies on effective funding regimes for institutional initiatives.

In supporting such national services, institutions automatically ensure their repositories will fit the wider international repository framework. An author whose field of study is global (which means most researchers) will expect that depositing copies of research outputs in an institutional repository will make it more accessible to peers and colleagues everywhere.

### 3.2.2 Create incentives for authors to deposit

This appears to be far and away the biggest issue for an institution to consider, once it has decided to establish and fund a repository. The value of a research repository depends on the number of authors contributing their works to it. The current thinking can be summarised as follows:

1. most researchers support the idea of self-archiving to an open access research repository in principle
2. voluntary self-deposit has limited success; institutions can struggle to reach 10% of authors depositing after several years
3. voluntary deposit works best in technical disciplines and amongst those publishing large numbers of papers
4. about 80% of authors will willingly comply if deposit is mandatory<sup>3</sup>; this could be funder or institution mandated
5. for some research output types, such as theses, changing policy to mandate deposit could be accomplished readily, given the will

The trend appears to be moving towards mandatory deposit, but each institution needs to set the policy that best meets its own local conditions. Regardless of the institution’s or funder’s deposit policy, the research repository service needs to show authors tangible and direct

<sup>3</sup> [http://www.jisc.ac.uk/index.cfm?name=jcie\\_scg](http://www.jisc.ac.uk/index.cfm?name=jcie_scg)



benefits from depositing their works. Options that Australian practices have shown are effective include:

- make it as simple and fast as possible for an author to deposit a work; this could include an assisted deposit service if appropriate
- in return, the author gets a permanent URI (universal resource identifier), meaning links to the work will always work, guaranteed safe-keeping in a trusted location, and high search rankings
- make usage statistics available, such as most frequently downloaded papers and papers downloaded by author
- provide profile services, such as the ability to generate a list of publications dynamically on the author's home page
- identify champions in each discipline who actively promote the benefits of the repository to their colleagues *and* listen to their feedback

Authors at different stages of their careers have different motivations. In the early stages, presence and visibility are most important, so the enhanced exposure a repository offers is a major benefit. In later stages, preservation of works that may otherwise be lost takes on a greater priority.

### 3.2.3 Link to PBRF reporting process

There are 2 distinct linkages to consider between the institutional research repositories framework and PBRF (performance-based research funding)—at the policy level and at the operational level. At the policy level, what weight, if any, should be given to the deposit of a research output in an open access repository? Recognition of the value of research repositories within the PBRF framework would help to put additional weight behind those charged with implementing and supporting repositories for their institutions. Any such policy needs to cover all research outputs, including books and monographs, accepting metadata-only records where full text is unavailable or not appropriate.

At the operational level, there are opportunities for efficiency gains. To what extent would it be possible or desirable for the Tertiary Education Commission to harvest selected research metadata in which it has an interest automatically from the data elements exposed to the Web by an OAI-compliant institutional repository?<sup>4</sup> Similarly, a number of New Zealand universities are planning to implement the Research Master system. What is the degree of overlap between the data elements held in Research Master and those held in a research repository? Further data and process analysis is needed to ensure that data of interest

<sup>4</sup> Institutions decide and control which, if any, data elements of their repositories they expose for others to harvest.

are collected once only, using the most efficient process, validated, and made available to the systems that need them, at least cost and effort.

Smaller institutions may wish to explore options for using the data in their research repositories (potentially augmented with additional elements) to support their research reporting requirements. This option may be relevant to Crown research institutes, as well as to universities and polytechnics.

### 3.2.4 Build repository management capability

All the projects visited in Australia highlighted resource constraints as an issue. This includes gaining the attention of institutional policy makers to engage them in discussions on priorities; availability of academic library staff to develop and secure adoption of information management policies around the institutional research repository; access to IT staff with the skills and experience necessary to set up and configure repository hardware and software; and staff to promote the repository service to academics and support them during the deposit process if needed. In addition, there are 2 distinct models that institutions have used when establishing research repositories, and each institution will need the capability to make an informed judgement on which model will work better. The models are:

1. one large institution-wide research repository, crossing all disciplines, potentially with an institutional portal giving different views of the content
2. several smaller discipline-based research repositories within one institution, potentially with an institution-wide resource harvesting and discovery service

Some institutions that started with discipline-based research repositories have since moved, or are moving, to an institution-wide repository. In a few cases, research funders have led the creation of cross-institution, discipline-based repositories, but this model is less common. The current generally-accepted thinking appears to be:

- ❖ an institution-wide research repository is more likely to be sustainable and better reflects the strategic value of a repository to an institution, but this approach requires significant commitment from the highest levels
- ❖ a discipline-based research repository has lower cost of entry and is easier to launch, but often relies on the enthusiasm and commitment of local champions so may not be sustainable in the longer term
- ❖ it is better for an institution to start small with discipline-based research repositories than to do nothing, but plan to migrate to an institution-wide repository in the medium term once the concept is accepted



Institutions may wish to consider what mechanisms they might use to build research repository capability and learn from one another's experiences. What is the role for inter-institutional partnerships where several institutions share a common repository purpose? How might such partnerships be structured and funded?

### 3.2.5 Copyright must allow open access deposit

Authors are properly worried about infringing agreed copyright arrangements with publishers. Such infringements can sometimes be inadvertent. The University of Queensland presentation described a common situation whereby publishers explicitly grant authors the right to deposit copies of post-print articles on open access repositories, provided that these are new instances of the works. Some authors have been depositing PDF copies of their published articles, not realising that this breaches the agreement. In the case of monographs, authors may reasonably expect to deposit the table of contents, perhaps selected chapters, and ideally (but not necessarily) the entire work.

Institutions can take a few simple steps to help authors comply with publishers' copyright agreements, while retaining the relevant rights over their works. Examples of Australian practices are:

- draw authors' attention to the SHERPA RoMEO service<sup>5</sup> which lists publishers' permission policies with respect to self-archiving
- encourage authors to consider publishers' self-archiving policies when choosing where to publish their research and to favour those placing fewest restrictions (RoMEO green publishers)
- give authors a standard clause they can use when submitting articles for publication, reserving the right to deposit pre-prints and post-prints in the institution's research repository
- include in the article deposit (file upload) screen a description of what steps an author needs to take to ensure the copy being deposited does not infringe copyright arrangements
- provide guidance to authors to help them manage the copyright in their research articles and ensure they do not assign it to a third party in a way that would prevent deposition of a copy in the repository

Journal publishers are increasingly moving to accept such provisions. Institutions and funding agencies around the world are considering these and other similar positions. If implemented, the arrangements for self-archiving suggested above will become part of the normal research practice in every discipline. It is not clear how long this will take to happen or how widely these practices will be adopted. One option would be for all New Zealand research institutions and funding

<sup>5</sup> <http://www.sherpa.ac.uk/romeo.php>

agencies to adopt the same set of guidelines to authors. Such guidelines would need to be developed in consultation with all stakeholders.

### 3.3 Outline project plan

The initial focus is on establishing and populating research repositories in a number of institutions as rapidly as possible. This section lays out a way forward for the next 12 months, with an outlook beyond that. The pace at which the repository movement is evolving makes detailed long term planning very difficult.

#### 3.3.1 Start with demonstrator projects

A demonstrator is a limited implementation with clear business goals, real users, and representative content, that can be extended or modified subsequently. The initial goal is to establish by the end of 2006 a demonstration national digital research repository federation, consisting of:

- OAI-compliant digital research repositories in universities and other research institutions, and for independent scholars
- a New Zealand research portal that enables discovery of and access to research outputs held in these repositories

The intent is that suitable demonstrators will be presented at the National Digital Forum<sup>6</sup> in November 2005. The forum is an opportunity to further the discussion, debate the recommendations and demonstrate overseas initiatives. We envisage that institutions will choose solutions based on local needs and the diversity of needs will lead to a number of different approaches. We thus expect that the demonstrators will use 2–3 different repository software packages, reflecting the different needs and creating an opportunity to gain direct practical experience of the different options' strengths and weaknesses.

The first step is to set up and recruit content for research repositories in at least 3 institutions. For the purposes of demonstration, the most effective approach will be for these to focus on a small number of willing and committed disciplines. The demonstrators are an opportunity to investigate and address the policy, people, process and promotion issues, as well as the technology. Choosing software with the lowest barrier to entry, provided that it meets the institution's needs, will allow the institution to focus most of its efforts on managing the cultural change.

<sup>6</sup> <http://ndf.natlib.govt.nz/>



Institutions will need to consider whether to adopt a strategy of one institution-wide research repository or several discipline repositories with an institution-wide discovery service. For the purposes of the demonstration, the most important goal is to populate the repositories with sufficient representative content that people can see and judge the value. However, institutions also need to gain confidence that the solution will be sustainable going forward. Each institution will need to consider and decide the scope it wants its repository to address. See “Scope aspires to include all research outputs” on page 20.

### 3.3.2 Set up cross-functional teams

Building and recruiting content for a demonstrator research repository will require the engagement of the academic library, the information technology service, the research office, and the schools identified as early adopters. We envisage that at least the following core team roles will be required (one person may take more than one role and roles may be shared):

- a process change champion who embeds the new policies and procedures into existing processes
- an advocate who promotes the benefits at all levels of the institution and in particular to authors
- a cataloguer who works with (or on behalf of) authors to enhance the description of research outputs
- a solution architect who installs, configures and adjusts the chosen software and hardware
- a sponsor who ensures the project has visibility, priority and ongoing commitment

Each institution’s project team is likely to involve the equivalent of 1–3 people for about a year, depending on the scope of content and type of solution chosen. Once the repository is in place, ongoing support is likely to require less than 1 person-equivalent per year, depending on the level of activity.

At least in the early stages of work, it would be valuable for teams from different institutions to meet regularly (perhaps every 3 months) and exchange experiences, discuss common issues, or share problems and decisions. The National Library may wish to facilitate this collaboration on behalf of the research sector. Project teams may also find it valuable to continue to use the institutional repositories wiki to share their knowledge.

Use of demonstrators gives institutions a practical way to investigate the business issues and options they will face, in a way that suits their particular environment.

## Choose suitable repository software

---

Several viable technical solutions are available, but each institution needs to make its own decision on which solution will best meet its needs. To ensure interoperability and federated resource discovery, all institutions wishing to participate in the global repository network must follow certain standards.

### 4.1 Technical solution options

---

This section sets out illustrative options for each aspect of the overall repository framework. It does not recommend preferred options, as different institutions will have different needs and priorities. Choosing a suitable technical solution is as important as managing the cultural issues, but requires less effort to implement.

#### 4.1.1 A range of repository software options exists

The Open Society Institute Guide to Institutional Repository Software<sup>1</sup> describes 9 repository packages that meet the following criteria:

1. They are available via an Open Source licence—that is, they are available for free and can be freely modified, upgraded, and redistributed.
2. They comply with the latest version of the Open Archives Initiative metadata harvesting protocols—this OAI compliance helps ensure that each implementation can participate in a global network of interoperable research repositories.
3. They are currently released and publicly available—several new systems are currently being developed. As these systems become available for public release, a revised guide will include them.

---

<sup>1</sup> <http://www.soros.org/openaccess/software>



The guide describes its approach as follows.

The systems presented in this guide—Archimede, ARNO, CDSware, DSpace, Eprints, Fedora, i-Tor, MyCoRe, and OPUS—meet these criteria and allow an institution to implement a complete framework for an OAI-compliant repository without resorting to in-house technical development. While this guide describes these solutions, it does not attempt to identify the “best” system or to recommend one system over another. In each institution’s case, the best software will be that which aligns well with the institution’s particular requirements.

The latest version of the guide was published in August 2004. Some proprietary or open source content management systems may also potentially be suitable, but they are generally not OAI-compliant and the working party has not investigated this option. Overseas experience suggests that it is better to choose a special purpose repository management system, rather than fitting a repository into a more general-purpose solution. This reflects the “service unbundling” principle of the repository architecture. See “Architecture” on page 22. The study has also not examined options for various value-added services, such as the MDID software for teaching and learning with digital images, the Almagest Media Repository, or the Greenstone Digital Library.

The most widely installed repository software solutions are Eprints and DSpace, with 161 and 90 registered repositories respectively on the global Institutional Archives Register<sup>2</sup>. While all the solutions offer repository content management services, i-Tor also provides a metadata harvesting and resource discovery component. Proprietary repository software is starting to become available from some library system vendors, such as DigiTool from ExLibris or Symposia from Innovative Interfaces. The study has not investigated these.

#### 4.1.2 Needs and priorities determine choice

As with any software acquisition, the first step is for an institution to define its requirements and priorities, taking into account the different needs of different user communities. Topics for an institution to consider include:

- will the institution run a single institution-wide repository, several discipline-based repositories, or form a consortium with others to create a cross-institution repository or repositories
- what is the scope of materials to be included; this covers both the type of object (such as articles, monographs, multimedia works, and images) and whether to include unpublished materials (grey literature) and supporting datasets as well as published materials

<sup>2</sup> <http://archives.eprints.org/eprints.php?action=browse>

- ❖ what are the ingest (content recruitment) requirements; this covers matters such as self-archiving or assisted-deposit, item-by-item loading or bulk ingest, deposit of supporting data, and whether deposit is mandated or voluntary
- ❖ what are the metadata management requirements; this covers matters such as use of subject thesauri, control lists for object types (such as whether to distinguish between PhD theses and research Masters theses), and control lists for access types (such as public, restricted to staff at this institution, and so on)
- ❖ what are digital object management requirements; this covers matters such as what file formats to support, quality control, persistent object identification, archiving and disposal, and long term preservation
- ❖ what are the access requirements; this covers matters such as aligning the public interface with the institution's standards, aligning the deposit interface with the desired work-flow process, exposure of metadata to resource discovery services, and exposure of content to full text search services
- ❖ what is the content acquisition plan; this includes matters such as identifying early adopters and champions, deciding where to start, setting up an initial demonstrator, repository promotion, and answering frequently asked questions

These topics show that an institution will want to sponsor its repository initiative at a senior level, to ensure there is strong and committed leadership.

### 4.1.3 Access control over protected works

Repositories allow institutions to manage both open *and* restricted access materials. The SPARC institutional repository user guide states<sup>3</sup>:

A variety of legitimate circumstances might require an institution to limit access to particular content to a specific set of users. These circumstances might include copyright restrictions, policies established by a particular research community (limiting access to departmental working papers to members of that department, for example), embargoes that an institution's Sponsored Programs Office might require to keep the institution in compliance with the terms of sponsor contracts, and even monetary access fees for certain data. Implementing these policy-based restrictions—which necessarily challenge the notion of 'pure' open access for legitimate reasons—requires robust access and rights management mechanisms to allow or restrict access to content—and, conceivably, to parts of digital objects—by a variety of criteria, including user type, institutional affiliation, user community, and others.

<sup>3</sup> [http://www.arl.org/sparc/IR/IR\\_Guide.html](http://www.arl.org/sparc/IR/IR_Guide.html)



There are broadly 3 levels of access control: the metadata and work are open; the metadata are open but the work is restricted wholly or in part; and both metadata and the work are restricted.

In a federation of OAI-compliant repositories, the additional need arises for an information seeker at one institution to gain access to protected works held at a different institution. This implies a system of *federated trust* among institutions, rather than a central authentication and authorisation authority. It requires the information seeker to be registered with an *identity provider*, able to pass an identity token to a *service provider*, in this case a repository, on request. Based on the identity token, the service provider decides whether to grant access to the requested work. The collection of service providers and identity providers together make up a federation. A federal agent mediates interactions between providers. The agent's role is to ask the information seeker "where are you from" and send a request for identity to that institution.

Shibboleth<sup>4</sup> has emerged as the leading and most widely-accepted standard for accessing multiple distributed resources using a single identity. It defines a set of protocols for securely passing identity information between institutions and service providers. It relies on the institution to establish identity and the service provider to determine access rights, based on the identity information provided. Shibboleth is designed to protect an information seeker's privacy. The information an institution passes to a service provider is about *status* (such as "this person is a graduate student at our institution"), *not* identity ("Holly Golightly is a Senior Lecturer in the Faculty of Media Studies"). Australia and a number of European countries are adopting Shibboleth. The MAMS (meta access management system) open source project<sup>5</sup> is a FRODO initiative aimed at providing a Shibboleth server on a CD, making it easier for an institution to join an existing federation.

Identity management is not a repository function. Rather, repositories need to be in place and well populated, with access rules around protected works, so that when identity management comes to fruition, richer services can be progressively added. Some institutions may decide to become early Shibboleth adopters; others may decide to maintain a watching brief.

#### 4.1.4 Other services

It is important also for a repository to distinguish between peer-reviewed research and non peer-reviewed research. Including both formal and informal scholarly communications in the same repository is

---

<sup>4</sup> <http://shibboleth.internet2.edu/>

<sup>5</sup> <http://www.melcoe.mq.edu.au/projects/MAMS/>

desirable so long as information seekers are made aware of what they are reading, for example an un-vetted pre-print or a peer-reviewed article. If a repository includes both peer-reviewed and non peer-reviewed works, these must be clearly labelled as such, enabling readers to discriminate readily between the different types of content.

Institutions may choose to offer services that complement the research repository, thus making the content more valuable and deposit more attractive. Scholarly publishing services in collaboration with university presses are one example. Possible services include:

- ❖ specialist electronic journals, including peer-reviewed e-journals
- ❖ virtual monographs—complementary on-line versions of print publications
- ❖ multimedia theses—rich and complex forms containing a variety of media objects, with internal and external linking
- ❖ print-on-demand publications, including an ability to assemble selected components and create a “print run of one” or mix and burn a CD from a dynamically-created play-list of audio material
- ❖ repository-based subject portals which give access to everything deposited on a particular subject, such as the Treaty of Waitangi or volcanoes—such portals are intrinsically global
- ❖ discipline slices which can act as a virtual journal, such as a New Zealand physics portal
- ❖ podcasts drawn from sound archives or picture collections to support distance learning

Traditionally, academic presses have small print runs and single editions. If the repository holds deposited copies of the content of monographs, this can create opportunities for such presses to print individual copies on demand or to publish revised editions at low cost.

In planning a research repository, an institution can choose to be a data provider, a service provider, or both.

## 4.2 Technical issues

This section identifies the technical issues for institutions to consider when setting up and managing a research repository.



### 4.2.1 Choose suitable software

In choosing the solution best suited to its requirements, an institution may wish to consider and evaluate the following options.

1. A hosted service offers the lowest barrier to entry—a third party service provider hosts the repository on behalf of an institution on a fee for service basis. One option is the ProQuest Digital Commons service<sup>6</sup>.
2. Eprints<sup>7</sup> is a low cost option for a repository primarily aimed at open access to article pre-prints and post-prints, including digital theses. A range of object types can be uploaded, including video, audio, images and zip files. There are 11 Eprints repositories in Australia and the University of Tasmania has offered a consulting service to help New Zealand institutions choosing this option. This option offers easy entry, at the price of limited functionality.
3. DSpace<sup>8</sup> is functionally richer and supports a wide range of object types, including text, sound, images and video. It provides detailed implementation guidelines. The institutional repository at the Australian National University uses DSpace. Institutions for which Eprints is not quite suitable may find DSpace more closely meets their needs, without being unnecessarily complex.
4. Acknowledging the success of the Dutch DAREnet repository system, an institution may wish to consider ARNO<sup>9</sup> as an alternative to DSpace.
5. For an institution seeking a powerful, well-architected repository suitable for very large (1,000,000 plus) collections, and that have the technical expertise to support it, Fedora<sup>10</sup> is a possible solution. Both Monash University and the University of Queensland are building services layers to improve Fedora's usability. Monash has a production repository containing about 70 papers.
6. CDSware<sup>11</sup> may be a suitable alternative for large and complex collections. While it does not appear to be widely-used, it is proven in the rigorous and demanding environment of CERN.
7. For an institution considering multiple discipline-based repositories with a metadata harvesting and resource discovery service, i-Tor<sup>12</sup> may be a suitable option.

<sup>6</sup> <http://www.umi.com/proquest/digitalcommons/>

<sup>7</sup> <http://www.eprints.org/>

<sup>8</sup> <http://www.dspace.org/> plus guidelines at <http://www.dspace.org/implement/>

<sup>9</sup> <http://arno.uvt.nl/~arno/site/index.html>

<sup>10</sup> <http://www.fedora.info/>

<sup>11</sup> <http://cdsware.cern.ch/>

<sup>12</sup> <http://www.i-tor.org/>

The above list shows that different options offer different capabilities. As in any software selection exercise, the preferred option depends on fit to requirements and priorities. The least risk, most widely installed and best proven solutions are currently Eprints and DSpace. Detailed identification and evaluation of all possible software options, considering both open source and proprietary solutions, was outside the scope of the feasibility study.

Institutions are encouraged to check references for their preferred option, choosing reference sites with requirements similar to their own.

#### 4.2.2 Adopt relevant global standards

Common standards facilitate interoperability and development of services such as resource discovery, authentication and authorisation, or federated searching. Institutions wishing their repositories to form part of the emerging global network of linked repositories need to ensure their repositories can interoperate with others. To achieve interoperability, the following areas (and potentially others) require discussion, elaboration and agreement across participating institutions:

- ❖ avoid unnecessary limitations on the types and content or format of digital objects
- ❖ adopt a common standard for the unique, persistent identification of digital objects
- ❖ digital objects must have accompanying descriptive metadata, conforming at least to some form of Dublin Core<sup>13</sup>
- ❖ repositories may expose additional metadata specific to their domain needs
- ❖ the format for metadata must be XML-based
- ❖ adopt OAI-PMH<sup>14</sup> as the standard for data exchange between networked repositories
- ❖ adopt HTTP as the protocol for linking, so that anyone using any browser can access the repository

The software packages described above support these general requirements, but implementation details may vary.

#### 4.2.3 Systems integration

Many institutions have, or are planning, systems for gathering, processing and reporting information about research carried out by groups and individual researchers. For example, several universities are imple-

---

<sup>13</sup> <http://dublincore.org/>

<sup>14</sup> <http://www.openarchives.org/>



menting Research Master<sup>15</sup>. Although research management systems and institutional repositories have different purposes, there is significant overlap of metadata. It is impractical to expect institutions to capture these data twice. Institutions need to determine how to transfer data from a repository to their research management system and vice versa.

Institutions unable to justify the cost of a research management system may wish to explore the option of generating research reports from a repository. This may require the capture of additional metadata.

Both approaches imply an ongoing commitment to populate the repository with data for the bulk of the institution's research outputs. Reporting requirements may also influence decisions on the scope of an institution's repository, for example whether or not to include research data as well as published articles.

#### 4.2.4 Prepare for the future

Institutional repositories are at an early stage of their evolution and there are many areas still under active development. A number of complex questions around long term repository sustainability remain unresolved. At the same time, by offering a repository deposit service, an institution is making an implied promise to look after deposited works either in perpetuity or until there is an informed decision to dispose of a work. Issues for institutions to consider include:

- how to ingest (recruit) large content and metadata collections; this requires some form of bulk import capability
- how to migrate content to a different repository software package if the need arises; this requires bulk data export of content and the associated metadata
- mechanisms for archiving and disposing of selected content that is no longer required, including the purge of associated metadata
- how to preserve repository content in an accessible form, for an indefinite period of time; this means sustaining data (matters such as format migration) *and* sustaining methods (such as the ability to run programs over datasets)
- protecting digital rights, which may be time-varying; this includes ownership, knowing what access rights are specified and whether these expire, and potentially privacy of some content

We expect institutions will focus first and foremost on recruiting and ingesting content: acquiring the digital assets. However, a repository implementation plan needs to consider the total content life-cycle:

1. content recruitment

---

<sup>15</sup> <http://www.researchmaster.com.au/>



2. asset disclosure and promotion
3. services that add value to data
4. data interoperability
5. preservation and disposal

At the start of the project, decision-makers may wish to ask the implementation team how active is the development community for the proposed solution and what is the exit strategy for that solution.

#### 4.2.5 Hardware is not an entry issue

All of the software options are based on the principle of *mobile code* and run in any Unix or Linux environment. Institutions can avoid potential problems by checking that their preferred software option is already running in their preferred hardware environment at another site or sites. Repository software is not an especially resource intensive application. Depending on the expected scale and rate of content ingestion, a dedicated server may cost from \$5,000 to \$15,000. It is likely to be several years before capacity constraints (large collections and frequent complex searches) arise.

All packages are designed to use an open source database management system, either MySQL or PostgreSQL. In addition, some solutions may support selected proprietary database management systems, such as Oracle and DB2. For institutions choosing a hosted service option, hardware is of course not an issue.





## Recommended repository actions

---

The feasibility study has found that institutional research repositories are viable and practical for the New Zealand research sector. They give authors enhanced exposure and can enable improved institutional efficiency. They are rapidly becoming an essential part of the global research infrastructure. The immediate mission is to create new and more efficient pathways to research output discovery, by establishing a network of federated repositories. The steering group recommends a number of actions to achieve this.

### 5.1 Establish a national federation

---

The steering group recommends that government agencies, in consultation with research institutions, take the following actions to ensure a consistent national framework:

1. develop a business case to government for funding institutional research repositories as an essential enabler of enriched scholarly communication in the digital age
2. adopt an architectural strategy of diversity with interoperability, noting that a range of proven and emerging technical solutions exists and that one size does not fit all needs
3. seek opportunities to learn from, collaborate with, and contribute to Australian repository and interoperability initiatives
4. investigate the overlap between data collected for PBRF reporting and research repository deposit to ensure common data items are only collected once
5. ask the National Library to set up a national research resource discovery service that harvests descriptive data from the demonstrator repositories
6. investigate the case for a national research repository for outputs of non-aligned scholars and confirm that the National Library is the most suitable host



7. develop strategies to make institutional research repositories a permanent and sustainable part of the national and local research infrastructure

The National Library has a critical ongoing role to play in providing the direction and energy required to implement the national framework successfully.

## 5.2 Support institutional initiatives

---

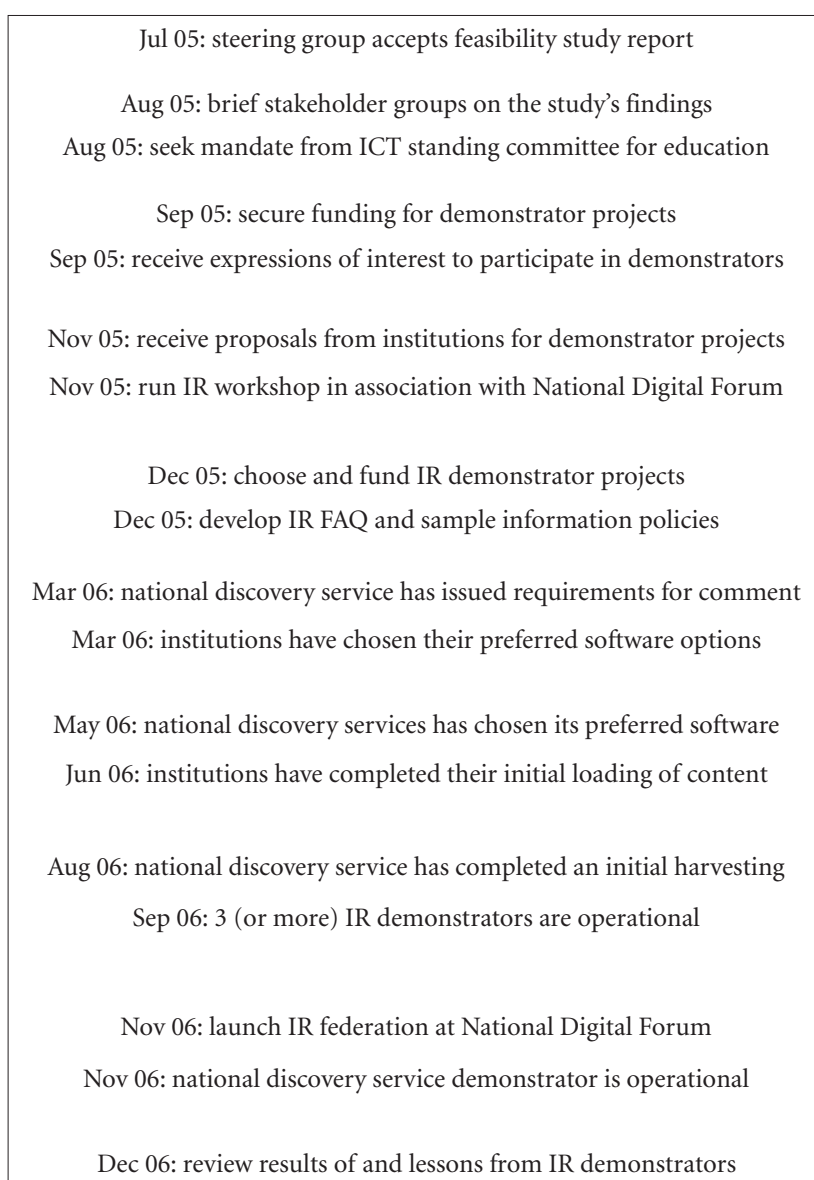
The steering group recommends that government agencies, in consultation with research institutions, take the following actions to support local repository initiatives and ensure these are consistent with the national framework:

1. seek commitment from institutions to fund demonstrator research repository projects and give them support to build their repository capability
2. promote the academic library as the natural home of research repositories and note that it needs the mandate and resources to fulfil this responsibility
3. consider what guidance to offer on the matter of voluntary or mandated deposit and propose to institutions that they mandate deposit of theses and dissertations
4. promote New Zealand institutions' presence in the redeveloped Australasian Digital Theses initiative
5. participate in regular colloquia to discuss common issues and exchange knowledge about repository implementation and support
6. promote use of software products that comply with the OAI-PMH standard, to enable development of shared repository services
7. consider the case for providing project-based funding to bring repository initiatives to fruition quickly

Repository technology has matured to the point where creation of a nationally federated community of IRs is feasible, and the benefits to scholars, researchers and the public are significant.

### 5.3 Adopt a common road map

The steering group recommends that all participants work to the same road map, and continue to co-operate and collaborate through the demonstrator phase. The following diagram shows the overall plan to the end of 2007. The dates shown are *completion-by* dates—different institutions will advance at different rates. The plan assumes funding will be available; for example, a possible funding formula would be for central government to meet 50% of the cost of demonstrator projects, on the basis that institutions publish and share their experiences.



The National Library will continue to act as the co-ordinator and facilitator, until such time as a repositories federation is in place with its own governance structures.





---

Appendix A

**Write-up from the Australian site visits**

