Improving information infrastructure: recent activities of the National Library of Australia

Warwick Cathro
Assistant Director-General, Innovation
National Library of Australia
wcathro@nla.gov.au

Abstract:

The National Library of Australia is undertaking a number of initiatives aimed at improving the infrastructure that provides access to information resources by Australian libraries and their users. It is giving priority to developing infrastructure services which will benefit the general public by providing them with online access to a greater range of content and providing them with easy-to-use services to discover and access information content, including the collections of Australian libraries. The paper describes some of these activities, including the recently completed Kinetica Redevelopment Project.
Introduction

Our prime role as libraries is to provide our users with access to the information resources that they need. We exercise this role individually and collectively. Our library collections, web sites and catalogues provide a direct benefit to our own users, but they also form part of the whole information infrastructure.

The word “infrastructure” is frequently used within our community. We hear it in phrases like “research infrastructure”, “research information infrastructure” and “national information infrastructure”. For the purposes of this paper, “information infrastructure” is defined as “the set of services that supports access to information resources by the library community and its users, current and future”. Infrastructure services are often developed collaboratively and they are intended primarily to benefit the whole information access system.

Some examples of “information infrastructure” services in Australia are:

- Consortia arrangements such as those managed by the Council of Australian University Librarians (CAUL) and the Council of Australian State Libraries (CASL) for the purpose of acquiring access to commercial electronic information resources
- Libraries Australia and other discovery services managed by the National Library
- Informit, which provides search access to Australian databases, and the indexing services that create these databases
- AARLIN (Australian Academic & Research Library Network), as a shared portal service used by a number of universities
- The Australian Digital Theses Program
- Australian subject gateways, mostly funded by the Australian Research Council through its Linkage, Infrastructure, Equipment and Facilities Program.

Clearly there are many providers of such services. There is no organised process for achieving inter-operability between such services, or for identifying and dealing with overlaps and gaps.

Research infrastructure and public infrastructure

In 2001 the Department of Education, Science and Training launched the Systemic Infrastructure Initiative (DEST, 2001) as part of the Government’s “Backing Australia’s Ability” Program. The Initiative was about improving the effectiveness of Australian research by developing “research infrastructure” services. Essentially these services aim to allow researchers to collaborate more effectively, to discover resources (such as original source material, and papers written by other researchers) and to manage their own research outputs and make them accessible.
In 2003, the Minister for Education, Science and Training established the Australian Research Information Infrastructure Committee (ARIIC), to advise on the medium term needs for the development of services such as:

- collaborative support services (services needed by distributed research teams such as peer-to-peer data sharing, simulation and visualisation tools, and collaborative annotation tools);
- middleware services (such as federated authentication and authorisation, and digital rights management) which will support more streamlined access to the resources that each researcher is entitled to use; and
- services which allow institutions to manage, make accessible and preserve the outputs of research, such as research papers and data sets.

Many services, publicly funded to support research, have the potential to deliver benefits to the general public. In fact, the Internet itself is the most prominent and obvious example of “research infrastructure” which has evolved into public information infrastructure. The public, no less than the researcher, requires rapid and easy access to information resources that they are entitled to use. In other words, there is a strong overlap between “research infrastructure” and “national information infrastructure”.

In this paper I am principally concerned with the services that support information access. It is in this area that the requirements for research infrastructure and public information infrastructure strongly overlap. I will examine efforts, particularly those of the National Library, to improve the access infrastructure though provision of online content and through strengthening of discovery services.

**The National Library’s mandate**

The National Library has a longstanding record of support for the national information infrastructure. Its enabling legislation gives the Library a mandate to build a national collection, to make it available in the national interest, to provide a record of Australian publishing and other bibliographic services, and to “cooperate in library matters”. In acting on this mandate over the years, the Library has developed some key components of the national information infrastructure.

These activities support underlying objectives such as:

- to provide the public with online access to a greater range of content;
- to ensure that future users will be able to access online content;
- to provide the public with easy-to-use services to discover and access information content, including the collections of Australian libraries; and
- to give people efficient access to the online resources that they are entitled to use.
The Library is continuing to undertake projects to improve and renovate those components of the national information infrastructure which support the above objectives, where the Library is already playing a major role, or where it is best placed to take on this responsibility. In selecting such projects, the Library is guided by its Directions Statement and its Balanced Scorecard framework (National Library of Australia, 2005). After the Library has identified candidate projects which appear likely to make a significant contribution to achieving the objectives in its Directions Statement, the Library will often undertake a feasibility study to assess the benefits, costs and risks of these projects.

Examples of the National Library’s recent activities in pursuit of these objectives are set out below.

**Online access to a greater range of content**

The National Library aims to provide the public with online access to a greater range of content. While access to some content will require payment by the user at the point of access, the Library aims to extend the range of content that is available directly without payment and without authentication barriers. In some cases, this objective may be achieved by digitising content that is in the public domain. In other cases, it may involve a national licence arrangement, funded (for example) by the library community, that will give free access to all members of the public.

Additional Government funding would accelerate progress towards this goal. We have an example in the New Zealand National Content Strategy, which forms part of the New Zealand Digital Strategy (New Zealand, 2005).

During 2003 a Committee of the Australian Senate (the Environment, Communications, Information Technology and the Arts References Committee) conducted an inquiry into *Libraries in the online environment*. As part of this inquiry, the Committee investigated possible measures to increase the range and depth of online content that is available to the general public (Senate, 2003). The Committee recommended:

(a) that the National Library of Australia identify a number of key databases for which national site licensing might be desirable; and

(b) that additional Australian Government funding be extended to the National Library of Australia for this purpose (Recommendation 9).

The Government responded to the Senate Committee report in June 2004 (Department of Communications Information Technology and the Arts, 2004). The response did not commit the Government to any additional funding. Following that response, the Australian library community formed the National Licensing Reference Group, following a forum convened by the National Library in December 2004 (National Licensing Reference Group, 2005). The Group is developing a proposal for licensing access to a “core set” of online content, such as major Australian academic and popular journals. The content could be provided directly to library users (whether on site or remote from their library) including users of reference services such as AskNow.
This national licence arrangement would build on the existing collaboration between the state libraries and the National Library, which has established the CASL purchasing consortium. Under the Group’s proposal, funding will need to come from the various library sectors, and there is a risk that some Australians may not have access to the resulting national service. The National Library will be encouraging the broadest possible participation in this initiative by the Australian library community.

The National Library is also interested in expanding online access to Australian content by digitising public domain works in its own collection and the collections of partner institutions. There is a clear benefit in public funding of such an activity: by contrast, when public domain works are digitised by the commercial sector, the resulting licenses restrict use of the digitised versions to authorised, fee-paying users.

The Senate Committee on Libraries in the online environment recognised the public benefit in such activity, and it recommended the creation of a “new National Heritage Grants program for peak cultural institutions to assist in the digitisation of their collections” (Recommendation 3). While there has been no early action on this recommendation, the National Library is actively pursuing a range of mechanisms to fund this activity, including funding from the Australian Research Council and funding from the Library’s collection depreciation resources.

As a first step in this activity, the Library partnered with the Australian National University (ANU) and other institutions to prepare a funding bid to digitise a significant set of Australian newspapers (comprising major capital city daily newspapers published 1803-1954) and to make this set available to the public through a free online service. Funds were sought through a bid to the Australian Research Council, led by the ANU, in May 2005. Twelve other universities offered to contribute funds to support the project, and the state libraries offered to ensure the supply of source material of sufficient quality, in the form of microfilm versions of the newspapers concerned. The funding bid was not successful, but the Library remains convinced that such a service would be a valuable component of national information infrastructure. It is therefore continuing to explore mechanisms through which the project might be advanced.

An online searchable newspaper service provides a good example of infrastructure which could benefit both the research community and the general public, including historians, family historians and a wide range of other users.

**Access to online content by future users**

Our definition of “information infrastructure” recognises future users as well as current ones. We must ensure that future users (both researchers and the public) are able to access the online content which is being created and archived today.

The National Library is undertaking several activities to achieve this. For example, it continues to support the PANDORA archive as a shared infrastructure service with the state libraries and other partners, aimed at maintaining long term access to selected Australian web sites and online
During 2005, the Library undertook the complementary activity of commissioning and analysing a comprehensive capture of the Australian web domain.

Ensuring future access to such content will require a range of concerted measures beyond merely archiving it. These measures will depend in part on gaining a better understanding of the file formats which comprise the content, the obsolescence risks associated with each format, and the costs of dealing with obsolescence through format migration and other techniques.

The digital data sets created by researchers are as vulnerable to obsolescence as any other online content, and actions to address this challenge within the research sector are likely to deliver benefits for the wider public.

The Australian Partnership for Sustainable Repositories (APSR) has been funded by the Systemic Infrastructure Initiative to address this challenge in relation to the research outputs of Australian universities. The project, which is being led by the ANU, is attempting to build institutional repositories for long term sustainable access (Australian Partnership for Sustainable Repositories, 2005).

As a partner in APSR, the National Library will share the expertise on digital preservation issues that it gained through activities such as PANDORA. The Library will help the project to develop a sound approach to assessing the obsolescence risk of each file format represented in the APSR repositories, and it will trial a software tool that will alert repository managers of impending obsolescence. The Library will also advise on a strategy for including preservation metadata in the repositories, and will seek to influence the future development of open source repository software (such as DSpace and Fedora) to make use of preservation metadata. The aim of APSR is to make these tools and processes widely available to all those who are aiming to build sustainable collections of digital content.

Improve discovery and access services

The National Library aims to provide the public with easy-to-use services to discover and access information content, including the collections of Australian libraries. In its Directions Statement for 2003-2005, the Library defined its major undertaking as “to provide rapid and easy access to the wealth of information resources that reside in libraries and other cultural institutions, and to break down barriers that work against this”.

There is a significant challenge in improving the power, ease of use, and level of integration of the available discovery and access services, mainly because the discovery landscape is currently quite complex. Mechanisms for discovering information resources include:

- Google, including Google Scholar
- Institutional portals
Collaborative portals such as AARLIN (supported by twelve university library members, and initially developed with funding by the Australian Research Council and the Systemic Infrastructure Initiative)

Library catalogues

the Libraries Australia Search service (formerly the Kinetica Search Service)

Specialised discovery services such as PictureAustralia, MusicAustralia, the Register of Australian Archives and Manuscripts (RAAM) and the ARROW Discovery Service.

Indexes, databases and electronic journal aggregation services, subscribed to by university libraries, that support access to journal articles, conference papers, and similar resources

Subject gateways, including those funded by Australian Research Council grants (such as AustLit, Australian e-Humanities Gateway, Guide to Australian Literary Manuscripts, WebLaw).

The need to reduce the complexity and increase the effectiveness of the discovery services for the public has been actively discussed during the past year by CASL. In response to this need, CASL has developed an Information Access Plan (Council of Australian State Libraries, 2005).

The Plan was initiated by a CASL workshop held in Adelaide in late 2004. As one tool to aid its thinking, the workshop considered some hypothetical information seekers and examined the pathways they would need to follow to meet their information need. One of these, “Joan of Gilgandra”, was researching the Cooee Marches which were part of the public recruiting effort in the first World War. Information on this topic might be found, for example, through Google; through reference services such as AskNow; through the collections of Joan’s public library, state library, state archives, National Archives and the National Library; or through searching the Australian National Bibliographic Database, APAIS and other indexes and databases provided through the web sites of the state and national libraries.

The CASL Information Access Plan represents a start towards making the discovery process less confusing and more coherent. The Plan includes a number of initial tasks, such as:

- improving the design of the web sites of the CASL members;
- collaborating on Internet subject guides;
- exploring the feasibility of federated search across the collections of cultural institutions;
- exposing the Libraries Australia Search service to the users of the state and public libraries; and
- engaging with public libraries to streamline the access and interlending pathways to their collections.

The National Library, no less than the other members of CASL, is examining how its readers might obtain maximum benefit from the Libraries Australia Search service. The Library is currently reviewing the issues involved if it were to make the Australian National Bibliographic Database (ANBD) the primary target to be searched by its readers, as opposed to retaining its ILMS catalogue as the primary target.
Among the discovery services listed above are some (such as the AARLIN portal) that are not being fully exploited because they are provided on a cost recovery basis. External funding to subsidise such services, for example from research infrastructure programs, would almost certainly increase their use.

In this context, in conjunction with replacing Kinetica by Libraries Australia, the National Library has moved to develop a new business model for access to the Australian National Bibliographic Database. Since the ANBD was established in the early 1980s, access to it has been based on a cost recovery business model. As a result, that access has been available only to customer libraries and, to a limited extent, the end users of those libraries. The fact that charging was linked closely to usage provided a disincentive for libraries to make this service available to their end users.

In recent years, a number of overseas bibliographic services have moved to a more open business model. For example, the Research Libraries Group’s “RedLightGreen” service has made the RLG union catalogue freely available, with an emphasis on meeting the research needs of undergraduates. OCLC’s “Open WorldCat” program is making the data in the WorldCat database freely available to web users via popular Internet search, bibliographic, and bookselling sites.

With a similar motivation, the National Library has decided to provide free access by the Australian public to the Libraries Australia Search service from March 2006. The Library will introduce a basic free search service, while reserving certain advanced and valued-added services for subscribing libraries and their end users. The Library is also planning to seed metadata from the ANBD to Internet search engines such as Google and Yahoo, along similar lines to OCLC’s Open WorldCat service. This will help bring the content of Australian library collections to the attention of Internet users who otherwise might not have thought to search in library catalogues.

The new business model is one of the outcomes of the Kinetica Redevelopment Project, which was completed in December 2005. The Appendix provides a summary of the benefits of this Project and describes how it was managed. The Project itself represents a major investment by the National Library in improving the national information infrastructure. In addition to providing end users with a powerful but easy to use interface for searching the ANBD, the Project has delivered a more reliable and effective platform for Australian libraries to contribute to the database.

In its 2003 report, the Senate Committee on Libraries in the online environment recommended “that the Library receive additional funding to provide improved access to Kinetica for all Australian libraries and end users” (Recommendation 1). In the absence of any early action on this recommendation, the Library developed a strategy for making progress towards this objective. While the introduction of a free service represents a financial risk for the National Library, this risk is mitigated by the cost savings that will be achieved as a result of replacing Kinetica by Libraries Australia.

Effectively, the Library has chosen to share the benefits of its investment in Libraries Australia between the customer libraries and the Australian public. The libraries receive a benefit in the
form of improved services and a modest reduction in fees. At the same time a new subscription model, introduced in July 2005, has broken the link between usage and fees. As a result, the customer libraries now have no disincentive to make Libraries Australia available to their end users.

By introducing a basic free search service as part of Libraries Australia, the National Library has aligned the business model for that service more closely with the “free access” model of the Library’s other federated discovery services (PictureAustralia, MusicAustralia, Register of Australian Archives & Manuscripts, and the ARROW Discovery Service).

As part of its goal of providing the public with easy-to-use services to discover and access information content, the Library intends to consolidate and extend these other federated discovery services. For example, the Library is currently working to re-engineer the Register of Australian Archives & Manuscripts (RAAM) to make it a “view” of the National Bibliographic Database. It will retain its own branding and interface, but behind the scenes, a search on RAAM will be a search of the NBD with appropriate limits applied. This approach has a number of advantages: it will obviate the need for any special database update software for RAAM; it will reduce the Library’s IT maintenance overheads; and it will allow the users of RAAM to take advantage of the “Get this item” service which has been developed for Libraries Australia.

Once this task is completed the Library plans to re-engineer MusicAustralia along the same lines. (Currently MusicAustralia derives all its data from the ANBD, but holds it in a separate database and uses a separate set of software). It will be possible to build additional “views” of the ANBD in the future, to meet the needs of particular communities.

The Library also intends to extend RAAM to provide access to the content of finding aids, so that people will be able to search in more depth using these collection descriptions. To support this goal, the Library will invest in the conversion of finding aids, including (for example) the finding aids of the Australian Joint Copying Project, to the Encoded Archival Description standard.

Finally, the Library plans to develop a new federated service which will provide an entry point to Australian web-based biographical services. Provisionally branded People Australia, this service will construct a virtual web page for each person or organisation represented in the national authority file contained in the ANBD. Users of the service will be able to navigate from these web pages to entries in the ANBD for works by and about the person concerned, and will also be able to navigate to biographical information contained in other web-based services such as the Australian Dictionary of Biography Online.

**Improved access control services**

Access control services ensure that users gain efficient access to external information resources, where they are entitled to such access. They support requirements such as federated authentication and authorisation, and digital rights management. These services form another important component of the national information infrastructure. The major Australian project
aimed at developing improved access control services is the MAMS (Meta Access Management System) Project, funded by the Systemic Infrastructure Initiative (Macquarie University, 2005).

The National Library was invited to participate in the MAMS Project because of its work in the directory standards field. The Library will be modelling use cases that are relevant for library services, and will specify the mechanisms through which user attributes can be related to the service provider’s policies.

The Library has also been working to improve its own access control infrastructure. It has developed the Libraries Australia Administration service as a flexible, directory-based access control and administration service. The Library will examine the potential to link this service to the future MAMS infrastructure, in order to support federated authentication.

Under the federated authentication model that the MAMS Project is developing, a user seeking to gain access to any service that is part of the national information infrastructure would be referred back to a home institution for authentication there, and then redirected back to the service provider with a standard “security handle”. The service provider would then use this security handle to query the home institution about the user’s attributes. Based on these attributes, the service provider would then give the appropriate level of access rights to the user. All of these processes would occur through machine-to-machine transactions and would be managed by an “Access Control Federation” which will be established as a by-product of the MAMS Project.

During 2006, the National Library intends to pilot this approach with other institutions that elect to join the Access Control Federation.

The benefit of this approach is that the service provider will not need to set up and maintain its own user directory and authentication system. The service provider needs to know “is this person from an organisation that we trust?” and “does this person belong to a class that is entitled to use this service?” but does not need to maintain individual passwords or certificates, and does not need to know any more details about the user.

As with the other elements of information infrastructure, there is no reason why this approach could not be used beyond the research sector. In principle, it could be a future mechanism supporting access by public library users to a national licensed set of online content.

Conclusions

The information infrastructure has been defined as “the set of services that supports access to information resources by the library community and its users, current and future”. There are many providers of information infrastructure services in Australia, and there is no organised planning and funding mechanism.

A number of important infrastructure components are being funded through federal government programs to strengthen Australia’s research capability. Many of these components have the
potential to deliver benefits to the general public who, no less than the researcher, need rapid and easy access to the information resources that they are entitled to use.

In recent years, and with no additional financial support, the National Library has given particular priority to developing infrastructure services which will benefit the general public by providing them with online access to a greater range of content and providing them with easy-to-use services to discover and access information content, including the collections of Australian libraries.

The Library has been collaborating with other stakeholders, especially the state and university libraries, to improve the information infrastructure, both for the general public and the researcher. Examples include the Library’s involvement in the National Licensing Reference Group, its collaboration with university and state libraries on a plan to digitise Australian newspapers, its partnership with the state libraries in developing PANDORA, its partnership with universities to improve access by future users to research content, its participation in the CASL Information Access Plan, and its efforts to open up public access to the Australian National Bibliographic Database.
References


DEST see Department of Education, Science and Training.


Appendix: The Kinetica Redevelopment Project

Background

In October 2003, after consulting with the Kinetica Advisory Committee, the National Library took a decision to redevelop the Kinetica system and to develop what is now known as Libraries Australia. The motivations for that decision have been well documented, and may be summarised as:

- the need to improve services to the customer libraries;
- the need to provide end users with a powerful but easy to use interface for searching the ANBD;
- the need for an improved capacity to innovate and improve the service;
- the need to obtain the best possible value for money; and
- the need to position the service for the future by placing it on a sound financial and low-risk footing.

There were two aspects to reducing the future risk. One aspect was the need for a high degree of confidence that any software suppliers would be financially viable and reliable, so that Australian libraries would have a reliable and effective platform to contribute to the ANBD. The other was the need to minimise the risk of failure in any future redevelopment, by moving to a more modular architecture.

The Project was completed on 30 November 2005, in line with the timeframe that the Library announced in October 2003.

Project approach

From the time of the decision to undertake the redevelopment, it was envisaged that the future system would comprise four software modules, each with its own development strategy. These strategies involved a mixture of buying and building software, as follows.

The Bibliographic Sub-System (branded as Libraries Australia Cataloguing) was based on acquiring a product from the marketplace, having that product customised, installing and implementing it, migrating data to the new Sub-system, decommissioning the legacy systems (AMICUS and the National CJK System), and providing systems infrastructure support for the new Sub-system from the Library’s IT Division.

The Search & Products Sub-System (branded as Libraries Australia Search) was developed by the Library, using an existing XML database platform (TeraText) with which the Library was
familiar. The functional and performance suitability of this platform had been tested in a pilot development between April and August 2003.

The Customer Management Sub-system (later branded as Libraries Australia Administration) was developed by the Library using open source components and the Library’s applications development framework. This Sub-system was intended to help the National Library address its future needs for a corporate system to support user authentication, authorisation, logging, reporting and (where required) billing across a range of applications.

The fourth module of the new system was based on continuing to use the VDX software from Fretwell Downing Informatics, and re-branding it as Libraries Australia Document Delivery.

**Project management approach**

The Library decided to undertake this Project by taking upon itself the system integration role. The conduct of the Project was guided by the Library’s project management methodology, which it had adopted in 2001 and which was based on the “Prince 2” model.

However, the Library departed from the commonly accepted principle that every large project should have a single, full-time project manager. The Kinetica Redevelopment Project had a more complex structure, consisting of:

- a single Project Director (part time), who was responsible to the Library’s executive and the Council for the conduct of the Project;
- a Project Board of three people, chaired by the Project Director, with clearly defined Executive, Business Owner and Provider roles; and
- five small teams, each with a clearly defined role, led by three full-time and two part-time team leaders.

The Project Board met frequently, usually fortnightly. Typical Board tasks included signing off scope statements and specifications, monitoring the project budget and endorsing budget changes, accepting test reports, managing the project risk register and managing changes to the project plan.

The Library appointed a Quality Assurance consultant (from Ernst & Young) with a strong background in IT project management. The same consultant assisted the Library to develop the initial project plan and project risk register. The consultant would undertake an assessment of the project’s progress at regular intervals, averaging every four months. The consultant’s reports were transmitted to the Library’s executive as received, and a summary of the findings was included in regular reports to the Library’s Council.

The three full time team leaders were responsible for the development and/or acquisition of the software which was to support the three new modules of Libraries Australia. The other two teams (responsible for IT infrastructure and Implementation & Customer Migration) had part time team leaders corresponding to the appropriate ongoing responsibilities in the Library.
The architecture of the new system, and the interfaces between the four modules, was defined as one of the earliest project tasks, in meetings of the Team Leaders and other key project stakeholders.

Based on the initial project planning meetings, the Project Board took a decision to divide the project into two stages. The first stage would develop the Libraries Australia Search and Libraries Australia Administration components to the point where a new search service could be deployed, and would also complete the selection of a solution for the new Libraries Australia Cataloguing Service. The second stage would further develop the search service, implement the Libraries Australia Cataloguing Service, and decommission the legacy system components.

Each of the two development teams comprised a Team Leader, two developers and a business analyst. The two business analysts, while reporting to their team leaders, had a secondary reporting relationship to a senior business analyst who provided quality assurance on their work. The senior business analyst (who worked part time on the project) developed the skills of the business analysts in Use Case Modelling, and was also responsible for writing the Scope Statements for each of the two project stages. These scope statements were reviewed by meetings of all project stakeholders.

The five team leaders met every week during the Project, with the meetings chaired by the Project Director. The weekly Team Leaders Meeting shared information on project progress, discussed overlapping issues in the system architecture and the project plan, and managed and regularly reviewed the Project Issues Log.

The devolved structure described above worked well for this Project. Each of the two stages was delivered on time and on budget.