## STAFF PAPER



# Libraries Australia Advisory Committee paper LAAC/2008/2/7

Single Business Discovery Project

**Dr Warwick Cathro** 

## **Background**

The Committee has been advised previously of the National Library's intention to integrate its discovery services by undertaking a "Single Business Discovery Project".

This Project commenced in late August 2008. This paper provides extracts from the latest draft of the Project Initiation Document. It clarifies the scope of the overall Project, and the proposed scope of Stage 1.

#### Introduction

During the period from 1997 to 2008 the National Library has developed the following national resource discovery services:

- Register of Australian Archives and Manuscripts
- Picture Australia
- Libraries Australia
- Music Australia
- Australia Dancing
- PANDORA search service
- ARROW Discovery Service
- Australian Newspapers Beta service.

During the past two years, the Library has contemplated the establishment of two additional services which might have been branded as:

- People Australia, and
- Journals Australia.

With the exception of Australia Dancing, the existing services are provided using three underlying search platforms (Blue Angel Technologies, TeraText and Lucene). In 2006 the Library identified Lucene as its preferred search platform for the medium term future.

In September 2006 the Library initiated a review of its IT architecture. That review reported in March 2007 <a href="www.nla.gov.au/dsp/documents/itag.pdf">www.nla.gov.au/dsp/documents/itag.pdf</a>>. The review made three major recommendations:

- That the Library implement a service oriented architecture, implying the use of software components loosely coupled through web services;
- That the Library adopt a Single Business approach, implying the development of a "single data corpus" which can be deployed in a range of contexts; and
- That the Library use existing open source solutions where these are robust and meet the Library's functional requirements.

In 2007 the Library established a Business Integration Taskforce, to give impetus to the above recommendations, with particular emphasis on the application of the second recommendation to the Library's discovery services. That Taskforce commissioned the development of a Single Business Discovery Prototype, which was delivered in December 2007. It is available for reviewing at <a href="http://ll02.nla.gov.au/">http://ll02.nla.gov.au/</a>>.

## Project objectives and scope

The Project has the following objectives.

- To improve the discovery experience for users by:
  - providing access from one place to the resources currently discoverable via the multiple discovery services that are to be decommissioned as part of this project;
  - o providing access to a significantly greater range of resources from a wider range of sources;
  - o providing searching of, and access to, full-text content;
  - enhancing ease of discovery by providing improved relevance ranking, refinement by facets, FRBR grouping of related items, exploitation of thesauri; and
  - o engaging with users through content and metadata annotation services.
- To reduce the number of separately maintained and branded discovery services offered by the Library.

The scope of the project will include:

- development of a discovery service that achieves the above objectives;
- decommissioning of the legacy services;
- migration of data contributors, customers, and other stakeholders to the new service;
- development of data collection policies;
- development of new business relationships to support the contribution of additional data sources and links to external target collections; and
- development of a new brand for the discovery service.

The Library will build the new system itself using Lucene and other appropriate open source tools.

The new service will be based on a virtual "national metadata store" which will be dissected into a number of "Collection views". These views will recognise the unique requirements of different types of information resource. They will have home pages and result sets tailored to the needs of users searching the type of material included in the view.

Included in the national metadata store will be descriptive metadata as well as full-text content such as archived web sites, digitised newspapers, books, journals and other texts, finding aids, content listings and transcripts.

In its analysis in 2007/08, the Business Integration Taskforce recognised three types of collection view:

- format-based views (eg. manuscripts);
- topic-based views (navigation points for information about people and organisations, places, events, and concepts); and
- "specialisations" (discipline or contextual approaches eg. music).

The following additional strategies are proposed in building the new service:

- To actively seek out additional Australian content, whether structured metadata or full text;
- To identify appropriate external targets and work proactively to include these in the new service;
- To expose significant and externally useful functionality as web services for reuse by other National Library applications, and possibly by non-Library systems;
- To learn from experience with previous prototypes (ANBD Prototype, Single Business Prototype, Newspapers Beta), and with the new National Library Catalogue and the new ARROW Discovery Service;
- To use open source tools in building the new service, and to the extent that resources permit, to contribute the new software to the open source community;
- To use techniques such as RDF or topic maps in order to manage complex data relationships; and
- To adopt a staged approach to the Project.

## **Benefits**

In broad terms there are three benefits to be gained from undertaking this Project:

- A better service for users;
- Greater efficiency for the Library in maintaining discovery services; and
- Improved ability for the Library to innovate.

A better service for users

The Project will deliver significant improvements for users by:

 making visible all resources related to the user's information need (by contrast, with a silo approach, users would frequently need to repeat searches in different services or else miss relevant information);

- modernising all of the National Library's discovery services to support relevance ranking, faceted search refinement, easy navigation of the data corpus and links to full text;
- creating a vehicle for bringing additional content to users, including full text of digitised books, access to Australian journal literature, and access to major external databases;
- providing a full text search capability based on this content;
- assisting the State and Territory libraries to improve their discovery services as part of the "Connecting and Discovering Content" Project of the NSLA Reimagining Library Services plan; and
- supporting reuse of the national metadata store by exposing subsets for use by other organisations.

## Greater maintenance efficiency

The value of the Single Business approach is that significant improvements can be made by leveraging off the discovery system and interfaces already developed.

The alternative to this strategy would be to develop new silo applications such as People Australia and Journals Australia, and attempt to maintain ten discovery services as separate applications. Without a significant increase in the National Library's IT resources, it is likely that each of these applications would have to wait at least three years between major releases.

#### In addition:

- moving the Library's discovery services from the legacy platforms onto Lucene will reduce the number of different platforms that the IT staff are required to support; and
- Lucene is much simpler for developers to learn than TeraText, and Lucene has a
  much broader user base than TeraText, which would reduce difficulties in
  obtaining the required skills.

## Improved ability to innovate

Many library-related software developments are now using Lucene. Completing the National Library's move to Lucene would give the Library the opportunity to benefit from the work of others and to contribute to the open source community.

In addition, the move to Lucene will enable the Library to provide the proposed collection views with acceptable performance.

## Stage 1

As the implementation of the new discovery service will be a major undertaking, there will be a staged migration to the new model. An initial stage will populate all of the format-based views (eg. Books, Maps. Manuscripts). Later stages will generally involve the further development of many of these views. In addition, some functional services may be less developed in the first stage, and may be enhanced in later stages.

**Stage 1** will replace the following services:

- the Libraries Australia free search service;
- the Australian Newspapers Beta service; and
- if possible, the RAAM search service

and will include the ability to search for people and organisations as topics.

At the conclusion of Stage 1 the following services will continue to exist as independent services:

- Picture Australia
- Libraries Australia subscription search service
- Music Australia
- Australia Dancing
- PANDORA search service
- ARROW Discovery Service
- the National Library catalogue.

Stage 1 will occur in two parts.

The first part (Stage 1A) is planned for delivery at the end of February 2009, and the second (Stage 1B) is planned for delivery at the end of June 2009.

**Stage 1A** will develop the technical framework to support the Single Business Discovery Service, and will verify that the selected technologies will provide acceptable performance, especially for record update.

By the end of Stage 1A, the national metadata store will contain:

- the data in the Australian National Bibliographic Database (ANBD);
- RAAM records not in the ANBD;
- biographical data provided by partners in the People Australia initiative;
- digitised newspaper content from the Newspaper Digitisation project;
- records in the ARROW Discovery Service;
- records in Picture Australia; and
- a limited amount of digitised book content (imported from existing mass digitisation databases and, if feasible, digitised books held by the Library.)

The following functions will be developed in Stage 1A:

• a home page for each collection view (search box and static content only);

- simple search;
- presentation of result sets and detailed record displays;
- a basic implementation of FRBR using OCLC Work Identifiers;
- relevance ranking for each collection view;
- matching of name permutations using name authorities (eg. "John David White" matching with "J.D. White");
- links to existing delivery systems (Newspapers, Picture Australia, PANDORA and the Library's Delivery system);
- harvest and full-text search of a limited number of digitised full-text books;
- full text search of digitised newspapers from the Newspaper Digitisation project;
   and
- access to a priority set of external targets, including Google Books and Amazon

To illustrate the features to be available at the end of Stage 1A, mock-ups based on initial thoughts about what these screens might look like are provided at Appendix 1.

**Stage 1B** will further develop the user interfaces and will add additional features. The Project will aim to decommission the current Libraries Australia Free search service following an appropriate period of parallel operation at the end of Stage 1B. If possible, the existing RAAM search service will also be decommissioned at the end of Stage 1B.

A significant separate piece of work included in Stage 1B is the completion of the development of the People Data Administration module. This module will allow Resource Sharing staff to manage people records and the identities they belong to, with the aim of improving their quality.

The functions that will be developed by the end of Stage 1B have been classified into 'Essential', 'Desirable' and 'Nice to Have' as listed below. The Project will aim to complete Stage 1B by 30 June 2009. If time does not permit all of the functions below to be implemented, the 'Nice to Haves' will be the first to be moved to a further stage, followed by the 'Desirables'.

It is possible that these priorities will change during the course of Stage 1B as the Project learns more about what is most important for users.

#### Essential

- My library(s)
- Did you mean?
- Searching external targets
- Getting options
- User download of records/citations
- Authentication/authorisation

- Tagging/comments
- Sitemaps for search engines
- A more sophisticated home page for each collection view
- Persistent Identifiers for People

#### Desirable

- Sort results by date
- Topic suggestions
- RSS feeds
- Full-text search of online RAAM finding aids

#### Nice to Have

- View-specific Advanced search
- Curated trails/themes
- Trails/themes/scratchpad created by users
- Links from the National Library's catalogue.

## **Beyond Stage 1**

At the end of Stage 1 and all future stages, a decision will be made by the Project Board as to whether the project should continue. Decisions about the scope of each future stage will also be made by the Board.

At this time it is envisaged that future stages may include:

- Development of the Journals View and of the Research Specialisation. Additional
  indexing data available to the Library would be added to the national metadata
  store, along with the full text of any journals digitised by the Library. More
  external targets would be added to provide greater access to journals metadata
  and full text;
- Further development of the Archives and Manuscripts View to include finding aids, file descriptions, and agency-level metadata made available by partners;
- Further development of the Pictures View as required to allow Picture Australia to be decommissioned as a separate service;
- Establishment of the Music Specialisation and the Dance Specialisation. Music Australia and Australia Dancing would be decommissioned as separate services;
- Expansion of the Web Archives View to include the Whole Domain Archive;
- Additional features to allow replacement of the Libraries Australia Subscription
   Search Service. The TeraText-based service would then be decommissioned; and

• Further development of the Maps View, including an expanded set of external targets which might include data sets from GeoScience Australia and the Bureau of Meteorology, and development of a Geospatial search interface.

## **Communication Plan**

Within the National Library, communication strategies will include Project Team meetings, liaison with appropriate collection managers, reports to Library management and Council.

Externally, there will be extensive communication with key data contributors such as Libraries Australia subscribers and Picture Australia contributors. A reference group from the State and Territory libraries will be formed, as part of the "Connecting and Discovering Content" Project of the NSLA "Reimagining Library Services" plan.

The Library will also liaise with external institutions that share an interest in the development of such systems. This will include the exchange of technical ideas for achieving the Project's objectives.

When the service is operational, the public will be informed through media releases and other communication strategies. These strategies will be developed in consultation with the Library's Communication and Marketing Branch.

#### Recommendation

The Committee to note the report.

Contact: Warwick Cathro

wcathro@nla.gov.au

02 6262 1403

7 October 2008

9 1 1 5

## **APPENDIX 1: SCREEN MOCK-UPS**

The attached five pages show the project team's ideas about what the user might see at various stages of the search process. These screen mock-ups were developed to clarify the team's thinking about what could be delivered to users in Stage 1.

Note that the groupings of format types are just the team's ideas at this early stage. The groupings to be actually used in the service will be based on such things as a card sort exercise, user testing, stakeholder discussion and the limitations of the data.

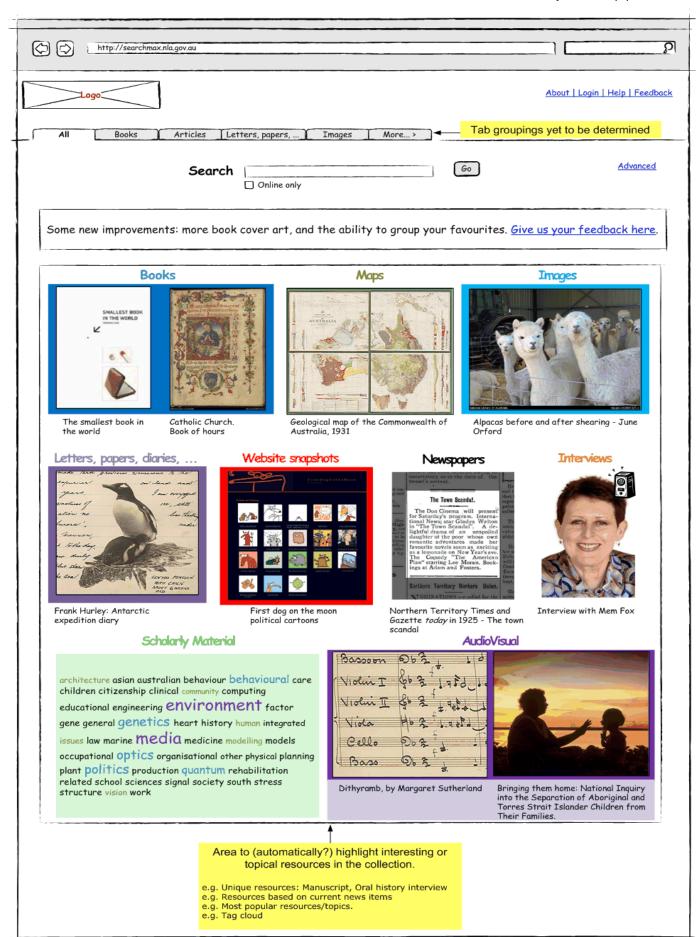
**Screen 1**: Home page for the service. Presents a set of images to illustrate the breadth of content that the user can find in the service. Allows the user to commence a simple search across all these collections simultaneously. Has tabs allowing the user to move straight to a particular collection view instead.

**Screen 2**: Shows the result of a search for "Peter Hollingworth" executed at Screen 1. Results sets are shown as a group of 10 "islands", plus an island for "External sources". A simple set of "Refine" facets is shown.

**Screen 3**: Shows the result of clicking on item 2 in the "Books, magazines, newspapers" island on Screen 2. The item is now displayed in the centre of the screen, grouped with related versions. "Get this item" options are shown. The other islands are still visible, but have moved to the right column of the screen.

**Screen 4**: Shows the result of clicking on the "Books" tab on Screen 1. Screen 4 would be developed into the Home page for the "Books, magazines, newspapers" View.

**Screen 5**: Shows the result of doing the search for "Peter Hollingworth" on Screen 4. A richer set of "Refine" facets is now shown.



Libraries Australia Advisory Committee paper LAAC/2008/2/7

