

# Australian Newspapers Digitisation Program



## Integration of Australian Newspapers Search and Delivery Service with Trove

Author: Rose Holley (Manager - ANDP)

Version: 1.0

Date: 10 November 2009

### Background

It was always intended that once the Australian Newspapers search and delivery system was completed and met the needs of users it would be integrated with a broader search and delivery system so that users would have the option of just searching newspaper content, or integrating their search to include other content such as books, journals and photographs. The broader search service (firstly called the Single Business Discovery Service prototype and now Trove) was developed in tandem with the Australian Newspapers service using the same infrastructure and software. Like the Australian Newspapers service it was released in a beta version to the public to gain public feedback. The June 2009 beta version of Trove contained a duplicate Australian Newspapers search service in order to test the system and public reaction. This was very successful. It is now intended that in December 2009 the Australian Newspapers search service will be fully integrated into Trove (i.e. the duplicate newspaper service within Trove replaced with the current live system). For the public this will mean:

- Changes to the look and colour of the Australian Newspapers home page and interface.
- An option to search across newspapers only, or newspapers with other content.
- Enhancement of functionality including implementation of some/all of the enhancement suggestions received from the public in 2008 (from Australian Newspaper beta testing).
- A new url for the Australian Newspapers service (a redirect will be permanently in place for old url's).

The existing user logins, records and rankings of text correction, tagging, and comments for Australian Newspapers users will be carried over into Trove, so users will notice no changes. The persistent identifiers (pi's, "cite this") already in use for newspaper titles and articles remain stable, constant and persistent as their name suggests.

The url's that will have a redirect permanently in place are:

<http://ndpbeta.nla.gov.au> and <http://newspapers.nla.gov.au/>

The new urls for Trove are:

Home page of Trove (search newspapers and other content) <http://trove.nla.gov.au>

Home page of Australian Newspapers <http://trove.nla.gov.au/ndp/del/home>

## About Trove (<http://trove.nla.gov.au>)

Trove is a new discovery experience focused on Australia and Australians. It supplements what search engines provide with reliable information from Australia's memory institutions.

If you are researching in the fields of the social sciences, literature, local or family history, or need inspiration for your school assignment, then this is the tool for you.

Trove was designed to:

- Provide a single point of access to the resources of the deep web
- Facilitate access to a significantly greater range of resources from major sources, including selected digitised material freely available online
- Support searching of, and access to, full-text content
- Enhance ease of discovery by providing improved relevance ranking, refinement by facets, FRBR-like grouping of related items (Functional Requirements for Bibliographic Records) and exploitation of thesauri
- Engage with communities and individuals through annotation services
- Ensure that relevant information is not missed in a search by reducing the need to search material-specific discovery services separately
- Provide a platform for niche services to query a vast resource of Australian metadata and adapt if for their own needs.

For example if researching images relating to Edmund Barton, our first Prime Minister, results will include descriptions such as people, book, manuscript, map and newspaper articles. A researcher searching for information on Nellie Melba will be presented with a range of results including biographies, pictures, music, newspapers, books etc.

### What can I find in Trove?

You can find information about resources collected by Australia's memory institutions, such as:

- Books, theses, reports, research articles, raw data sets, book chapters, conference proceedings as well as papers and records
- Maps and audio-visual materials including photographs, artworks, postcards, videos, musical sound, sheet music and sound recordings of interview
- The full-text of major Australian newspapers, from 1803 to 1954
- Copies of selected Australian websites which may no longer be available online or in their original form
- People and organisations

When you search Trove, each result may actually be a group of records. For example, all versions of "Possum Magic" by Mem Fox should be grouped together, so that only one appears in your results. When you click on this result, you will see a list of all the versions and editions in the group. This group can be referred to as a "work".

### What am I searching?

Trove provides integrated access to over 45 million items from a range of the National Library of Australia's collaborative services and from elsewhere.

Metadata (information about items) is pooled from Australian sources including:

- The [Australian National Bibliographic Database](#) - containing location information from more than 1,200 Australian libraries
- The [Picture Australia](#) program

- The [Australian Research Online](#) program
- The [Music Australia](#) program
- The Register of Australian Archives and Manuscripts
- The People Australia program including many biographies and relationships from
  - [Australia Dancing](#)
  - [Australian Dictionary of Biography Online](#)
  - [Australian Mines Atlas](#)
  - [Australian Women's Register](#)
  - [Collections Australia Network \(CAN\)](#)
  - [Libraries Australia](#)
  - [Music Australia](#)

and overseas sources including:

- [OAlster](#)
  - [PubMed Central](#)
  - [Project MUSE](#)
  - [NASA Technical Report Server](#)
  - [CiteBase](#)
  - [Nature Publishing Group](#)
  - and more than 1000 others...
- [Open Library](#) - online public domain books
- [Hathi Trust](#) - online public domain books
- [Wikipedia](#) - keywords (tags) associated with books

Text (the item itself) of

- [Australian Newspapers](#) - articles from historic Australian newspapers, 1803 to 1954.
- [PANDORA](#) - archived copies of Australian websites.
- The National Library of Australia's [manuscript finding aids](#)
- The [Library of Congress](#) tables of contents and sample book chapters
- The [Internet Archive's](#) full text public domain books

### **How does the relevance ranking work?**

The relevance ranking is complex and takes many things into account.

- A record with a match in the title, subject or author field is considered more relevant than one with a match anywhere else. (But this is complicated by the number of matches found, and other rules which can push records higher or lower)
- Exact matches are ranked higher than stemmed matches.
- Search terms close together are ranked higher than search terms far apart.
- If two records have very similar titles, but the second record also has the search terms found in the notes or full text of the record, the second record will be considered to be slightly more relevant than the first.
- A record which has a higher "density" of matches with the search terms is ranked higher than a record with a low density. (That is, a short sentence with one occurrence of your search terms will rank higher than 100 pages of text with one occurrence of your search terms)