

Trove: Innovation in Access to Information in Australia

Keywords: *Trove, National Library of Australia, Australian Newspapers, Web 2.0, Social engagement, Data engagement, Digital libraries.*

Rose Holley describes a major development in the Australian national digital information infrastructure.

Introduction

In late 2009 the National Library of Australia released version 1 of Trove [1] to the public. Trove is a free search engine. It searches across a large aggregation of Australian content. The treasure is over 90 million items from over 1000 libraries, museums, archives and other organisations which can be found at the click of a button. Finding information just got easier for many Australians. Exploring a wealth of resources and digital content like never before, including full-text books, journals and newspaper articles, images, music, sound, video, maps, Web sites, diaries, letters, archives, people and organisations has been an exciting adventure for users and the service has been heavily used. Finding and retrieving instantly information in context; interacting with content and social engagement are core features of the service. This article describes Trove features, usage, content building, and its applications for contributors and users in the national context.

Opportunities for Libraries

I see tremendous opportunities for libraries this year because of advances in technology. The changes in technology mean that anyone can create, describe or recommend content, which means that many people and organisations are becoming librarians or libraries in their own way. Librarians should not be threatened or dismayed by this but rather encouraged, since it means that society is retaining its ongoing interest in the creation, organisation and dissemination of content, and we have an integral role to play in these developments. Libraries and librarians are relevant more than ever in this environment because we have vast amounts of data and information to share, a huge amount of information expertise, and an understanding of how technology can assist us in making information more accessible.

We need to have new ideas and re-examine our old ideas to see how technology can help us. What things have we always wanted to do that we couldn't before, like providing a single point of access to all Australian information? Is this still pie in the sky or can we now achieve it? Libraries need to think big. As Charles Leadbeater would say 'Libraries need to think they are leading a mass movement, not just serving a clientele.' [2] Librarians are often thought of as gatekeepers with the emphasis being on closed access, but technology enables gatekeepers to open doors as well as close them and this is the opportunity I see. However many institutions will need to change their strategic thinking from control/shut to free/open before they can make this transition, and take a large dose of courage as well. The American author

Harriet Rubin says, 'Freedom is actually a bigger game than power. Power is about what you can control. Freedom is about what you can unleash.' [3] The National Library of Australia already took this step forward in 2008 with the advent of the Australian Newspapers beta service, which opened up the raw text of digitised Australian newspapers to the public for improvement, without moderation on a mass scale [4]. With a long history of collaboration across the Australian cultural heritage sector [5] with regard to digitisation, storage, and service delivery, the National Library of Australia is well placed to take the lead with innovation in access to information.

Some people may say, 'But isn't Google doing that, so why do we still need libraries?' There is no question in my mind that libraries are fundamentally different from Google and other similar services. Libraries are different to Google for these reasons: they commit to provide long-term preservation, curation and access to their content; they have no commercial motives in the provision of information (deemed by various library acts); they aim for universal access to everyone in society; and they are 'free for all'. To summarise: libraries are always and forever. Who can say that of a search engine, or of any commercial organisation, regardless of size?

The National Library of Australia reviewed its strategic directions and thinking in light of changes in technology and society. The three strategic objectives for 2009 – 2011 [6] are now:

1. 'We will collect and make accessible the record of Australian life... We will explore *new models* for creating and sharing information and for collecting materials, including *supporting the creation of knowledge by our users*'
2. We will meet our *users' needs* for rapid and easy access to *our collections and other information resources*.
3. We will collaborate with a variety of other institutions to *improve the delivery of information resources to the Australian public*.

In addition the strategic directions for the Resource Sharing and Innovation Division [7] acknowledge 'The changing expectations of users that they will not be passive receivers of information, but rather *contributors and participants in information services*.'

The outcome is Trove.

What is Trove?

The library has redesigned its underlying infrastructure for all its discovery services, and developed a new discovery service which has many features for both data engagement and social engagement. Trove [1] provides access not just to the Library's collections but to any Australian content or collections. Warwick Cathro, advocate for collaboration and the strategic lead for the service states 'Collaboration requires effort, and it also requires a change of mindset. In particular it requires a willingness to examine services from a perspective which does not place one's own institution at the centre.' [8]

At present over 1000 libraries, archives, museums, galleries and other organisations have enabled their data to be shared in the service which currently provides metadata for over 90 million items. Trove is essentially a search engine. It harvests metadata thus aggregating it into one place for searching. Trove does not store the content, only the metadata, so users end up on the site which holds the source of the data. Results are relevancy-ranked, not returned in contributor order, or biased towards any one source. There is minimal work for organisations wishing to contribute their collections and all the extra user traffic goes direct to their own site. The difference between Trove and other search engines is that most of the

content discovered via Trove would not be found in other search engines because it is buried in the 'deep web', for example in collection databases; and Trove has an Australian focus especially on unique Australian materials. To date most of the contributors are Australian cultural heritage organisations that hold unique Australian data (the gold in the Treasure Trove). However, having said this, the Trove team considers it vital that the resources within Trove are more widely discoverable via tools in common public use (eg Wikipedia, Google, Yahoo). So part of the plan is to encourage other search engines to harvest Trove; to encourage use of Trove work ID's (persistent identifiers) in other sources so that linkages are created between sources; and to develop an API so that other sites can draw out information from Trove in different ways.

After the Trove infrastructure was developed the content of the nine separate collaborative services that the National Library of Australia has been managing for several years was integrated. These services were:

- Australian National Bibliographic Database (Libraries Australia),
- PANDORA (archived Web sites),
- Australian Research Online,
- Picture Australia,
- Register of Australian Archives and Manuscripts,
- Australian Newspapers,
- People Australia,
- Music Australia, and
- Australia Dancing.

Added to this were other international sources of digital content such as the Open Library, Hathi Trust and OAISTER. This immediately provided content of the order of 90 million items (books, journals, pictures, music, newspapers, sound, video, diaries, archives, and Web sites) which could be easily found in a single search. The search results are complemented by relevant information from other Web sites. Heavily used Web sites which have an open API and can therefore be added to Trove as a target are: Google Books, Amazon, Wikipedia, Flickr and Google Videos (includes YouTube). These results in context are differentiated from the main results by appearing to the left of them.

The Library has employed several methods of data collection. Although its preferred method is by using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH), very little data are actually collected this way. This is because many libraries and archives are still unable to implement OAI. Because of this the Library now recognises that more flexibility is required with data collection. Other methods may be using an API, crawling sitemaps, using FTP or HTTP. The strength of cultural heritage institutions is that we have all used common data description schemas, and that we agree that data should be open and shared. Trove is the realisation and result of many, many years of working towards common standards (such as MARC, Dublin Core and EAC (Encoded Archival Context)) in order to make information accessible and free.

Trove Development

Trove was planned and under development for a number of years [9] (previously referred to in staff papers as the Single Business Discovery Project). The Australian Newspapers beta

service was actually a test of the new technical infrastructure which was based on MySQL, and a Lucene search index. The Australian Newspaper service was also the test-bed for social and data engagement and had the library's first implementation of tagging in 2008. Because the Australian Newspapers service was very successful and scalable, it was decided to continue with the same infrastructure for Trove. Social and data engagement features were also to be implemented on all content the library would deliver. Trove was released in November 2009 at an early stage of development, which was 15 months after Australian Newspapers beta had been released, with key staff working on both projects.

During the first half of 2010 there have been new releases of Trove at least once a month. The development of Trove has largely been driven by public feedback. This mirrored the newspaper development process. Feedback from the public is actively encouraged and feedback is considered critical for the development of the service, so we are sure we are meeting user needs. The sentiments of Charles Leadbeater in his essay the 'Art of With' [10] are being carried out. Leadbeater advocates that cultural heritage institutions need to learn the art of doing things 'with' people rather than 'for' or 'to' them. This includes the development cycle as well as the end results. Collaboration with the users is key. This was a lesson the library had already learnt with the Australian Newspapers. How can we know if we are meeting the changing expectations and needs of our users if we are not in close contact with them, moving along new paths together?

Trove Features

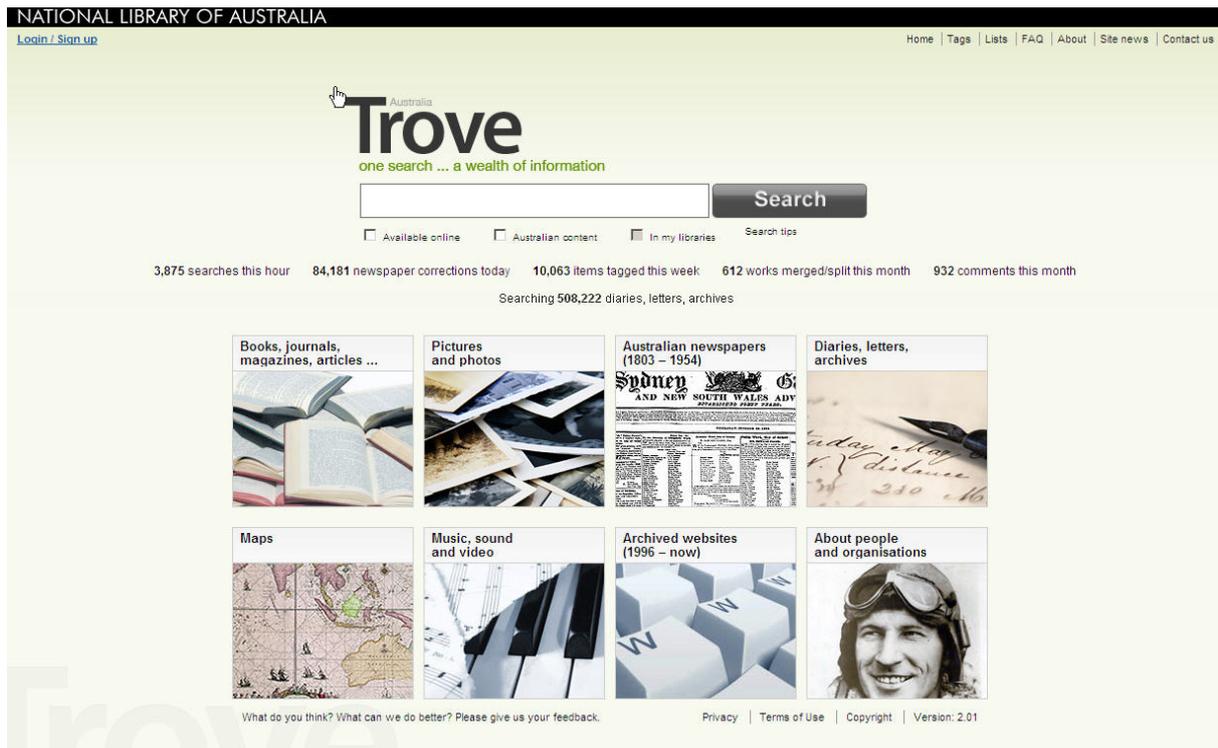


Figure 1: Trove Home Page

The key features of Trove are listed below and I would encourage readers to explore Trove for themselves by focusing on a topic of their own interest:

- Focus on Australian content
- Single search (search across everything at once in a few seconds)
- Find and Get (get options include digital, borrowing, buying, copying)

- Results returned ‘in context’ (the context reflects the eight groupings or zones (see [Table 3](#)) which were defined by users)
- Restrict search to Australian items, online items, or items in your preferred locations.
- Items that are online are obvious from the brief results display.
- Narrow down your search results by using ‘facets’ e.g. year, type, and language.
- Check copyright status.
- Cite the record using a persistent identifier and see different citation formats.
- View main record only or all the versions that are within it (relevant for books and images).
- Minimise some zones in result list.
- Browse items by zone.
- Registration and login are optional (required only to create a profile, search within ‘my libraries’, or create lists, private tags or private comments).
- View related items.
- ‘Did you mean?’ help with search terms.
- High-quality relevance ranking (National Library resources are not favoured over those from other sources).

Features that have been implemented in 2009/2010 as a response to changing expectations of users can be grouped by two categories: data engagement and social engagement. The difference between the two is that data engagement is normally for the benefit of the individual only, whereas the social engagement features encourage and nurture a virtual community to develop around the content or service.

Data engagement/creation features are as follows:

- Tag items.
- Rate items out of 5 stars.
- Add comments to items.
- Create contextual links to items outside Trove by adding hyperlinks in comments.
- Merge/split editions or versions of works which have not been automatically grouped correctly.
- Enhance/correct newspaper text.
- Create virtual lists (include items found within Trove or outside Trove, add own descriptions, annotations, context, order and make public or keep private).
- Add own images and descriptions via Flickr Picture Australia Group for inclusion in Trove (focus on Australian people, places, events, flora and fauna, etc).
- View history of your data interactions in your user profile.
- Be alerted to new newspaper content arriving (RSS feeds).
- View recent search terms.
- View items recently accessed.
- Identify your preferred libraries so that ‘my libraries’ search option and tab is active.

Social engagement features are as follows:

- View all public data interactions from home page (comments, tags, lists, text corrections, merges/splits).
- Other users can view your profile and your recent public tags, comments and text corrections.
- Crowdsourcing – text correction for newspapers, working towards a common goal.
- Newspaper text correction ‘hall of fame’ (monthly rankings of community), wall of fame (overall rankings).
- Newspaper articles – view all corrections on an article and users that made them.
- Comments – respond to comments.

- In newspaper zone your username appears in results list beside item you have corrected or commented on.
- Send persistent link for item to Twitter, Digg, Delicious, Facebook.
- User forum (add more information to your profile for others to see, eg interests, location; contact other users, arrange to work collaboratively, help other users by answering their questions).

Trove Usage

Trove has already established a user base of 1 million people in the first 6 months, which is comparable to the initial audience of the Australian Newspapers beta. Expected usage of Trove is for at least 10 million people (half the Australian population). This is based on the expectation that usage should exceed annual foot traffic through library doors. National, State and Territory Libraries of Australia (NSLA) had 7.7 million people through their doors in the year 2008- 2009 [11]. Figures from the Australian Bureau of Statistics show that in 2006 half of the population belonged to a library [12] and sought information on a regular basis, and that in December 2009 two-thirds of Australian households [13] had fast broadband access at home.

The National Library uses an extremely limited amount of its overall budget on marketing online services. Introducing a new service of this magnitude that is relevant and useful to a large section of the Australian population needs a targeted, well funded national marketing campaign. In the absence of such a campaign, it is expected that general awareness and usage of Trove will not grow rapidly in the first year, instead growing slowly over 2-3 years. Word will spread by mouth, in the online environment and through library connections initially.

Usage statistics gathered so far are as follows:

Users	Number
Unique users: cumulative for 2010	987,147
Number of registered users	12,858
Highest number of visits per day over period	19,084
Top location of users	Australia (72.8%)

Table 1: Trove Usage, 1 January 2010 – 31 May 2010

Content Type	Totals
Tags added	424,335
Comments added	9,192
Records merged	5,740
Records split	1,521
Lines of text corrected in newspapers	15.13 million
Lists created (introduced in May)	30
Images added by public via Flickr	69,131

Table 2: User-generated Content in Trove as at 31 May 2010

Trove Zone Name	Work Count in zone (millions)	Overall % of Trove Content	Overall % of Trove usage based on searches
All view	90.75	100%	16.5%
Books, journals, magazines, articles	30.40	33.50%	7.4%
Australian Newspapers	21.53	23.72%	63.5%
Pictures and photos	4.35	4.79%	7.9%
Archived Web sites	31.36	34.5%	1.3%
Music sound and video	1.37	1.50%	1.0%
Diaries, letters, archives	0.51	0.57%	0.8%
People and organisations	0.89	0.98%	0.7%
Maps	0.33	0.36%	0.5%

Table 3: Trove Content and Usage by Zone for 1-31 May 2010

Feedback from thousands of users tells us that creating connections and linkages between data is important, that finding related information and showing information in context is really helpful. Users think sharing, repurposing, mashing and adding to information is equally as important as finding it in the first place. Consistent messages received from users are ‘Give us.....’

- As much access to as much information as possible in one place.
- Tools to do stuff with the information we find.
- Freedom and choices with finding, getting and interacting with the information.
- Ways to work collaboratively together to achieve new things which have never before been possible so easily.

And in return our virtual community will give back their:

- Enthusiasm to do stuff and help us
- Expert subject knowledge
- Time
- Dedication

This has been clearly demonstrated by the degree of activity within the virtual community of Australian Newspaper text correctors [14].

Trove: Future Developments

In 2010 there are three main areas in which it is hoped to develop Trove. These are to encourage new contributors to provide their data to the service; to continue development of the service with key new features such as a Trove API and improved access to journal articles; and to raise awareness and usage of Trove. On the last point libraries can help by referring to the marketing page in Trove [15]. This shows you how to use a Trove logo on your Web site, to add a Trove search box to your site, to put Trove into your browser bar as a search box. Furthermore, some libraries are considering usage of Trove as their primary discovery service, or integrating Trove content into their own single discovery service when the Trove API becomes available later this year.

The success of Trove depends on having a large body of relevant content; the usability and functionality of the service; being able to migrate users successfully from the previous eight separate existing services to Trove; and raising awareness of the existence and usefulness of Trove in the community. Migration strategies are being developed for services such as Picture Australia, Australian Newspapers and Music Australia. Structured usability testing will be undertaken to address both the basic functions of Trove and difficult areas highlighted in user feedback later this year. A 'low-cost' marketing campaign has been developed for 2010 which mainly comprises distribution of bookmarks, attendance at conferences, speaking engagements by the Trove Team and publication of articles and adverts in library journals. Any organisation that has content in Trove is encouraged to undertake its own marketing. Trove did receive national media coverage in April 2010, but sustained ongoing marketing is important.

Conclusions

Trove would not have been possible without the long history of collaboration across cultural heritage institutions in Australia, the usage of common standards across this sector and the shared understanding that data should be open and accessible wherever possible. The National Library has taken a leadership role in demonstrating that a shift in strategic thinking and action must take place to respond to the changing expectations of users. Control of information is no longer the ultimate goal, but rather in giving users freedom and choices to interact with the data and each other, to create their own context within the information, and add their knowledge and content to it. In the eyes of the users this is just as important as finding the information in the first place. Libraries are well placed to respond to these needs having the technology, tools and information expertise, but more than that, not being driven by commercial gain. Their honest long-term goals are simply to make finding and getting information easier, now and forever and that is what Trove is all about.

References

1. Trove <http://trove.nla.gov.au>
2. Leadbeater, Charles (2009) The Internet and Society in the 21st Century. British Library Strategy Seminar September 23 2009, York.
<http://www.charlesleadbeater.net/cms/xstandard/The%20Internet%20and%20Society%20in%20the%2021st%20Century.pdf>
3. Fast Company (1998) The Fast Pack. Fast Company Magazine, Issue 13, January 31, 1998, page 5 <http://www.fastcompany.com/magazine/13/fastpack.html>
4. Holley, Rose. (2009) Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers, National Library of Australia, ISBN 9780642276940
http://www.nla.gov.au/ndp/project_details/documents/ANDP_ManyHands.pdf
5. Cathro, Warwick. (2009) Collaboration Strategies for Digital Collections: The Australian Experience. <http://www.nla.gov.au/openpublish/index.php/nlasp/article/view/1433/1738>
6. National Library of Australia Directions 2009 – 2011
http://www.nla.gov.au/library/NLA_Directions_2009-2011.pdf
7. National Library of Australia, Resource Sharing and Innovation Strategic Plan, January 2010 – December 2011 <http://www.nla.gov.au/librariesaustralia/documents/strategic-plan.pdf>
8. Cathro, Warwick (2010) Collaboration across the collecting sectors, Australian National Maritime Museum Lecture Series.

9. Cathro, Warwick (2008) Developing Trove: The policy and technical challenges. http://www.nla.gov.au/trove/marketing/TROVE_2010_02%20VALA2010_Cathro.doc
10. Leadbeater, Charles (2009) The Art of With. An original essay for Cornerhouse, Manchester
<http://www.charlesleadbeater.net/cms/xstandard/The%20Art%20of%20With%20PDF.pdf>
11. National and State Libraries of Australasia Web site, page accessed June 20, 2010.
<http://www.nsla.org.au/>
12. Australia Bureau of Stats (2006) Libraries – Library Facts and Figures
<http://www.abs.gov.au/websitedbs/a3121120.nsf/home/Client%20groups%20-%20Libraries%20-%20Library%20facts%20&%20figures>
13. Australian Bureau of Statistics (2009) 8146.0 - Household Use of Information Technology, Australia, 2008-09 December press release.
<http://www.abs.gov.au/ausstats/abs@.nsf/mediareleasesbytitle/180CCDDCB50AFA02CA257522001A3F4B?OpenDocument>
14. Holley, Rose (2009) A success story - Australian Newspapers Digitisation Program. Online Currents, 2009, vol. 23, n. 6, pp. 283-295 <http://eprints.rclis.org/17665>
15. Marketing materials for Trove <http://trove.nla.gov.au/general/marketing>

Author Details

Rose Holley
Manager -Trove
National Library of Australia
Parkes Place
Canberra
ACT 2600

Email: rholley@nla.gov.au
Web site: <http://trove.nla.gov.au/>
