TROVE STATUS REPORT
OCTOBER 2013

PURPOSE

1 This report highlights some milestones in content development, usage and service improvement that have been achieved since the last briefing to the Libraries Australia Advisory Committee in March 2013. The report also aims to give LAAC members a sense of the growing impact of Trove and the Trove API on research, and the continuing importance of user engagement for Trove’s development.

TROVE STAFFING

2 The past six months have seen a number of changes to the staffing of the Trove Support Team. Dr Tim Sherratt, former Harold White Fellow and digital humanities scholar, replaced Rose Holley as Trove Manager in May. The Library has provided additional funding for two new staff members for the next three financial years, and the team now has 5 fte staff.

EVALUATION OF CUSTOMER SATISFACTION WITH TROVE

3 In 2013 the Library commissioned a report from Gundabluey Research on customer satisfaction with Trove. Research was conducted in three stages and included an initial workshop with staff, a qualitative stage using online bulletin boards, and an online quantitative survey which was completed by 1087 respondents. The evaluation aimed to provide a demographic analysis of Trove users, information on usage patterns and preferences, and indications of whether Trove was meeting users’ information requirements. Opportunities in terms of functionality and how users expect to use Trove in the future were canvassed. Along with the analysis provided in the report, more than 200 pages of verbatim comments were captured in the research.

4 The report was delivered to the Library in July 2013 and analysis of its findings is continuing. The report offers us a rich information resource to draw upon for Trove’s future development. Tailored overviews which highlight findings of interest will be developed and shared with key stakeholders, including NSLA and CAUL. An overview of the report will be provided to the April 2014 LAAC meeting.

COLLECTIONS LARGE AND SMALL

5 Trove currently provides access to over 350 million resources. But the raw statistics sometimes obscure the diversity of Trove’s holdings and the exciting mix of
collections large and small, known and unknown. From voodoo dolls to research data sets, Trove’s focus on bringing uniquely Australian content to the surface is revealing all manner of riches.

A number of major collections have been added to Trove in the last six months. These include:

- Thirty journals from CSIRO Publishing, comprising more than 78,000 articles documenting the history of Australian science from the 1940s to the present.
- Information on more 67,000 Australian theatrical performances, harvested from AusStage, an online service developed by a number of Australian universities, and led by Flinders University.
- Nearly 60,000 objects from the National Museum of Australia.
- Broadcast details of an additional 14 Radio National programs including Background Briefing and Late Night Live.

Amongst the smaller contributions are collections like that of the Queensland Police Museum which includes the aforementioned voodoo doll as well as a stuffed dog called Peter, who was submitted as evidence in a murder trial. Peter has now become a regular feature in Trove presentations such as ‘Let us work with you to tell your story’, delivered at the Museums Australia national conference in May.¹

10 MILLION PAGES AND BEYOND

The Digitised Newspapers zone continues to be the most heavily used part of Trove and to attract the most devoted following. Thanks to the work of the Australian Newspaper Plan and the Newspaper Digitisation Program a number of major milestones have been achieved over the last few months:

- the 10 millionth page was digitised and made accessible through Trove²
- the 100 millionth article was digitised and made accessible through Trove
- our first handwritten newspaper, the Renmark Pioneer, became available


We are also looking forward to welcoming our first Chinese language newspapers sometime in the current financial year.

An analysis of the current contents of the newspaper zone by state reveals the impact of recent investments. This graph shows the number of available newspaper articles per year for each state or territory. The recent focus on the World War One period in NSW and Victoria is clearly evident.³

³ An interactive version of this graph is available at: http://wraggelabs.com/shed/trove/states/
Our dedicated band of volunteer text correctors has continued their labours, reaching an astonishing milestone of their own – over 100 million lines of text corrected. Perhaps even more remarkably, our top corrector recently passed the 2 million mark.

The Trove team monitors with interest the effects that access to digitised newspapers is having on a range of practices and professions. The importance to writers seeking period detail for their works has been highlighted recently by a number of comments and acknowledgements. Matthew Condon, a speaker at the Writing the Australian Landscape Conference, noted:

*In recent years I have been enmeshed in works of creative non-fiction, namely my monograph/memoir Brisbane and my two true crime books about the life and times of former police commissioner Terence Lewis and crime and corruption in Queensland over half a century – Three Crooked Kings and All Fall Down. Trove was a priceless resource in the way it allowed me to bring colour and fact to the social history elements of these books. It allowed me to explore the childhoods and teenage years of various true life characters in the crime books by virtue that they were mentioned, often obscurely, in regional newspapers in the middle of the 20th century. Over the three years of research for the crime books, a day didn’t go past where I didn’t dig around in Trove. I’m not sure I could have given the books the depth they deserved without this extraordinary digitisation program. I see it as a great gift to the nation.*

**NEWSPAPERS REDEVELOPMENT PROJECT**

The software that delivers digitized newspapers through Trove is currently being redeveloped. Although the project is primarily focused on delivering improvements in performance and stability, there will also be some significant changes to the user interface.

One important feature will be better support for mobile devices, recognizing the rapidly increasing number of users accessing Trove through their phones and tablets. An enhanced image viewer, a cleaner page design and a completely new browse interface are some of the other planned improvements.

As well as undertaking detailed internal testing, the Trove team is being supported by a team of eager volunteers, enlisted from the ranks of our public users, who are providing detailed feedback on each new prototype.

**FAREWELL MUSIC AUSTRALIA**

The decommissioning of the Library’s stand-alone discovery services – Picture Australia, Music Australia and Dance Australia – is almost complete, with only a small number of technical tasks remaining. As part of the decommissioning process, specifications for a new Performing Arts view of Trove were completed.

---

This selected subject view will allow easy filtering of search results through facets or the advanced search form. Indexing to fully support this view has commenced, with completion expected before the end of the calendar year. A new Biographical Tool was also created for managing party entries migrated from the decommissioned services.

17 A pilot project was undertaken to explore how selected subject views could be enhanced by targeted tagging. Musicologist Graeme Skinner was contracted to tag items in the Newspapers, Pictures, People and Lists zones that relating to music or the performing arts. More than 6,000 items were tagged in this way, far beyond the initial target of 2,200 items. Assessment of the full value of this work cannot be made until it is indexed as part of the Performing Arts view.

HIGHLIGHTING OPEN ACCESS

18 Discussions with the Australian Open Access Support Group, CAUL and the NHMRC have continued about enhancing access to Australian research within Trove. The Trove Support Team is investigating metadata and indexing changes that would improve discoverability, re-use, and options for reporting and analysis.

19 Separate NUC identifiers have been created and assigned to the institutional repositories of universities to distinguish them from their parent organizations, and their contents have been re-harvested. This means search results in Trove can now be limited to specific institutional repositories, and an accurate count of their holdings harvested into Trove can be found on the statistics page.

NEW RESEARCH USES

20 Researchers are starting to analyse Trove and its user community in new ways. Katherine Bode (Australian National University) obtained ARC funding to track the publication of serialised fiction through the newspaper database.

21 Sultana Lubna Alam (University of Canberra) and Annie Talve (Project Sisu) have both been examining the motivations of Trove users. Such studies are important as cultural institutions grapple with the possibilities and problems of crowdsourcing. Some of this research was featured at the ‘Inspired by Trove’ seminar held in February, and further events of this type are planned to feature innovative uses of Trove.

22 The Smart Services Cooperative Research Centre, based at the Queensland University of Technology, has been exploring possibilities for extracting geospatial data from newspaper articles in the development of their Paper Miner prototype. The PaperMiner prototype uses the Trove API to build a map view of search results in the newspaper database, giving a visual representation of the density of matches based on placenames extracted from the articles.

5 http://www.paperminer.org.au/
GOVHACK 2013

Trove featured prominently in GovHack 2013, a series of events around Australia aimed at encouraging the re-use of government data. Over 48 hours, from 31 May to 2 June, teams of developers in most of the capital cities worked to create applications, concepts and prototypes that demonstrated some of the possibilities of open data.

The teams were competing for a series of local and national prizes. NSLA sponsored two prizes in the Digital Humanities category – the Trove Prize ($1000) for best use of the Trove API, and the Trove Developer Prize ($3000) to enable a team to continue development of their entry using the API.

The winners of the NSLA prizes were:

- Trove Prize -- Piptory, a mobile app that collects and exposes geo-tagged historic photos from the State Library of WA, harvested via the Trove API.6

- Trove Developer Prize -- AdVintage, an iPad app that provides a way of exploring historic advertisements drawn from Trove.7

Significantly, a number of other prize winners also made use of Trove. For example:

- BrisBert, ‘an experiment in narrative-led federation’ won the Best of Digital Humanities Prize. The two runners-up in this category also used the Trove API.8

- Armchair Antarctica, a mashup of Antarctic-related data including historic newspaper articles from Trove, won the Best of Science Prize.9

The complete list of prize winners is available on the GovHack website.10 Approximately twenty teams submitted entries for the Trove prizes.

As required under the conditions of the Trove Developer Prize, the Advintage team has continued to work on improving and enhancing their app. James White described the process in an article entitled ‘Advintage. Ask for it by name.’ on the GovHack blog:

“At the conclusion of GovHack, we’d built a fully functional app. It connected with the Trove database and filtered the advertising content we were interested in. Users could select categories such as food, travel and clothing, could save adverts as favourites, and share them on Facebook or Twitter. However, like any weekend hack, there was a significant amount of work required to get the app ready for public release. Advintage has a very simple interface, but like many simple interfaces, it’s deceptively difficult to get it just right.

6 http://hackerspace.govhack.org/?q=groups/pixtory
7 http://hackerspace.govhack.org/?q=groups/advintage
8 http://hackerspace.govhack.org/?q=groups/brisbert
9 http://hackerspace.govhack.org/?q=groups/armchair-antarctica
10 http://www.govhack.org/2013-winners/
And of course any app drawing on a database of millions of images with associated metadata, will require a lot of optimisation to ensure a good user experience.11

THE POWER OF SERENDIPITY

Serendip-o-matic is a tool that connects your texts and sources to digital materials located in libraries, museums and archives around the world.12 It was developed in a week by a group of digital humanists as part of the ‘One Week | One Tool’ program organised by the Centre for History and Media at George Mason University. At its launch Serendip-o-matic included support for collections from the Digital Public Library of America (DPLA), Europeana and Flickr, however, within a few hours Tim Sherratt, Trove Manager, was able to contribute code to add in the Trove API. As a result, Trove was mentioned in a number of blog posts and articles describing the new tool, including ‘One Week | One Tool: Introducing Serendip-o-matic’ in The Chronicle of Higher Education.13

Also exploring the power of serendipity is @TroveNewsBot -- a simple Twitter bot that replies to user queries with articles from the newspaper database. TroveNewsBot also responds with historical newspaper articles to recent news headlines from the ABC and to random links from the Digital Public Library of America posted by @DPLABot.14 Most recently @TroveNewsBot joined the blogging site Tumblr, posting random newspaper illustrations at two hourly intervals.15

11 http://www.govhack.org/2013/08/27/advintage/
12 http://serendipomatic.org/
14 For documentation see: https://github.com/wragge/trovenewsbot
15 http://trovenewsbot.tumblr.com/
A generic version of the TroveNewsBot code has been made available so that Trove contributors can easily create their own collection bots.16 The National Museum of Australia’s @KasparBot and the John Curtin Prime Ministerial Library’s @CurtinLibBot have recently joined @TroveNewsBot on Twitter.

While these activities are not officially supported by Trove, they do demonstrate how the Trove API can be used to deliver Trove content in a variety of settings and to multiple audiences. The API can be used to take Trove into the online spaces where people already congregate.

SUPPORTING AND ENGAGING USERS

The Trove Support Team works to support our users through a range of activities and initiatives. A new Trove blog is currently under development and will be used to highlight recent contributions and new uses of the service.

We use Twitter to share interesting collection items and highlight the different ways that people engage with the service and are rapidly approaching 5,000 followers.

A series of nine new screencasts have been developed to guide users through a number of common tasks. These are currently available through the Trove YouTube

16 [https://github.com/wragge/trovebuildabot](https://github.com/wragge/trovebuildabot)
channel. The screencasts will be embedded within the new help system and linked from relevant points within Trove.

36 The team always enjoys the opportunity to talk to our users in person and have delivered a number of presentations in the last six months.

MAKING A LIST

37 Trove users continue to enrich resources by adding tags and comments. Indeed, the total number of tags now tops 2 million. However, the annotation feature that has shown strongest growth in recent months has been Lists. Lists are collections of resources created by users to organise and share their research. Lists can include anything in Trove, except for archived websites, and can be made public or private.

38 Some basic analysis gives a feeling for the extent and character of list-building activity. As of 18 September, there were 27,911 public lists created by 5,796 users, containing 449,624 items. By extracting the titles of lists and computing word frequencies we can create a picture of our users’ research interests. From this we see how features such as Lists support niche interests – the ‘long tail’ of online activity. An example is the dedicated Trove user who has created 107 lists related to lawn mowers.

39 Lists can also be used to foster collaborative activity. As part of its ‘Pitch In!’ project, the State Library of Queensland created lists to point volunteers to selected newspaper articles in need of text correction. The contents of public lists are also accessible through the Trove API, making it possible to create collections on Trove and then display the contents in other websites or applications.

SERVICE USAGE

40 Visitation to Trove grew strongly over the previous year, across all measures. The growth in use of mobile devices is particularly notable; growth in this measure has also been very strong in the first two months of 2013-2014. This growth reinforces the Library’s decision to redevelop Trove Newspapers to provide higher content and search capacity, and to provide a satisfying experience for those using mobile devices.

<table>
<thead>
<tr>
<th>Visitors</th>
<th>2011-12</th>
<th>2012-13</th>
<th>% Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique visits</td>
<td>7,934,622</td>
<td>12,219,940</td>
<td>54.01%</td>
</tr>
<tr>
<td>Total visits</td>
<td>14,804,081</td>
<td>20,832,047</td>
<td>40.72%</td>
</tr>
<tr>
<td>Highest # daily visits</td>
<td>57,445</td>
<td>78,738</td>
<td>37.07%</td>
</tr>
<tr>
<td>Mobile device visits</td>
<td>856,027</td>
<td>2,707,564</td>
<td>216.29%</td>
</tr>
<tr>
<td>Mobile device as % of total visits</td>
<td>5.78%</td>
<td>13.00%</td>
<td>124.82%</td>
</tr>
</tbody>
</table>

User registration continues to grow strongly; the 100,000 registered users’ milestone came at the about the same time that the 10 million digitised newspaper pages was achieved.

Use of Trove’s user engagement features has grown at almost exactly the same pace as overall growth in use, suggesting that new users find our engagement features as attractive as do longer term users.

Google Analytics data on use by state shows some variations compared to 2012-2013. The most marked of these is that Victorian users are now over-represented, NSW has almost achieved parity, and the over-representation of ACT and Tasmanian users has reduced. The Trove Evaluation Report found that response by state, and by metropolitan/regional/rural/remote categories almost exactly matched population distribution, with the exception of over representation of the ACT. We can now feel quite confident that Trove has achieved truly national reach.

Trove user locations remained fairly consistent over the year, with around two-thirds of users in Australia, followed by the US and UK.
RECOMMENDATION

The Committee to note the report.

Contact: Deirdre Kiorgaard

Director, Collaborative Services
National Library of Australia
dkiorgaa@nla.gov.au

4 October 2013