

Myths, mega-projects and making

Author : Tim Sherratt

Tagged as : [Trove](#)

Date : February 14, 2015

Keynote presented by video at [EuropeanaTech 2015](#), 13 February 2015.

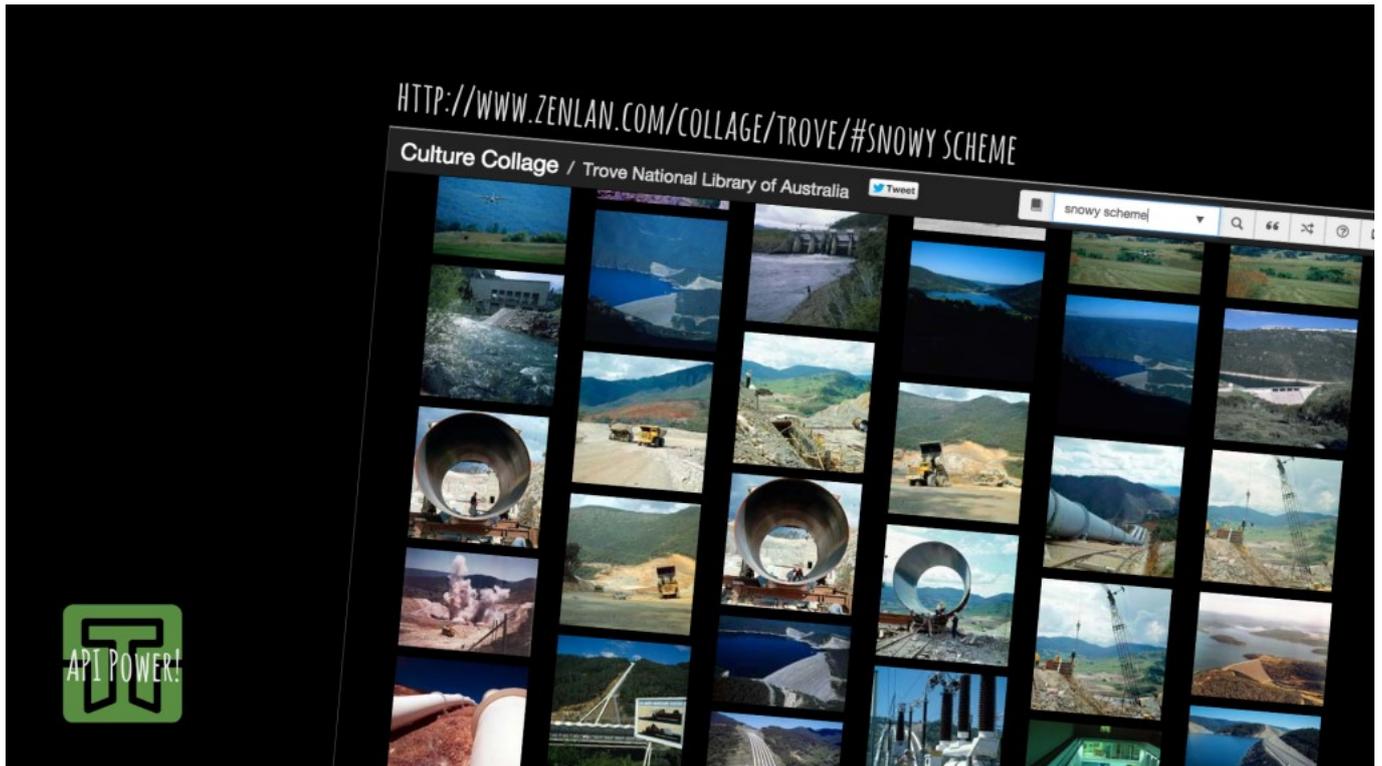
The [video](#) of this presentation is available on [Vimeo](#) and the [slides](#) are on [SlideShare](#).

In the aftermath of World War II, Australian hopes for a new era of national progress were expressed in a massive engineering project called the Snowy Mountains Scheme. The project promised new reserves of water and electricity to power development of Australia's inland.



'The Snowy Mountains Scheme', Canberra Times, 26 December 1975, p. 5.

Rivers were diverted, towns were relocated, and new reservoirs were created. Over 145km of tunnels were carved through the granite peaks of Australia's Great Dividing Range. Finally completed in the 1970s, the Snowy Mountains Scheme was an engineering marvel.



[http://www.zenlan.com/collage/trove/#snowy scheme](http://www.zenlan.com/collage/trove/#snowy%20scheme)

But this symbol of national pride would not have been possible without the labours of thousands of 'New Australians' drawn from across Europe. Some were recruited because of their skills, others were plucked from displaced persons camps and offered the chance of a new life — as long as they were prepared to work where the Australian government wanted them to.



‘AT WORK ON SNOWY MOUNTAINS SCHEME’, Sydney Morning Herald, 8 January 1954, p. 5.

The human and environmental costs of the project are still debated, but the Snowy Scheme is regularly invoked as the country’s prime nation-building project — an example of what can be achieved together through vision, leadership, and toil.

Why am I talking about this today?

Well, I suppose it’s a great chance to say ‘Hey Europe, thanks for all the people!’.

But it’s also because I wanted to highlight the mythic qualities of the mega-project — the cultural power that resides in the ‘big idea’ that promises to set us upon a path towards the future.

We are here today because we are embarked upon ambitious undertakings.

Our projects aim to reshape the cultural landscape. We are building pipelines and reservoirs — moving massive amounts of data across countries, across the world.

But as we’ve tried to show, these large scale efforts are only possible because of many smaller, local collaborations.

The Snowy Scheme was built by individuals fleeing the disruptions of war. They took a risk in the hope of something better. It’s important for us to reflect on the contributions and motivations of our communities, our partners, and our users. A big idea isn’t enough.

You probably all know that as well as metadata from libraries, museums, archives and universities, [Trove](#) provides access to almost 150 million full-text digitised newspaper articles. The OCR’d text of the articles is fully searchable, but suffers from the usual errors and inaccuracies.

Fortunately Trove users [have been eager to help](#). Anyone can jump in and correct the OCR output, and they do. More than 150 million lines of text have been corrected so far. Our top corrector (yes we have a scoreboard) has corrected more than 3 million lines of text.

Recently I’ve been [thinking about this work](#) and the limitations of language around online engagement. Our correctors are more than ‘users’ — ‘contributors’ perhaps, or ‘volunteers’?

But all of these words seem to place correctors on the other side of the interface — as clients rather than builders.

Each correction is a tweak of our search index. It changes the way the backend functions, increasing the efficiency of the system by getting people to the things they’re interested in more quickly.

Perhaps we should call our correctors ‘discovery engineers’?

The mythic mega-project maintains a sense of otherness — it is exceptional, an achievement above and beyond the realms of ordinary experience. But this obscures the many small acts of commitment and cooperation that make it possible. These are the expressions of ordinary lives, the routine and repetitive alongside moments of passion and meaning.

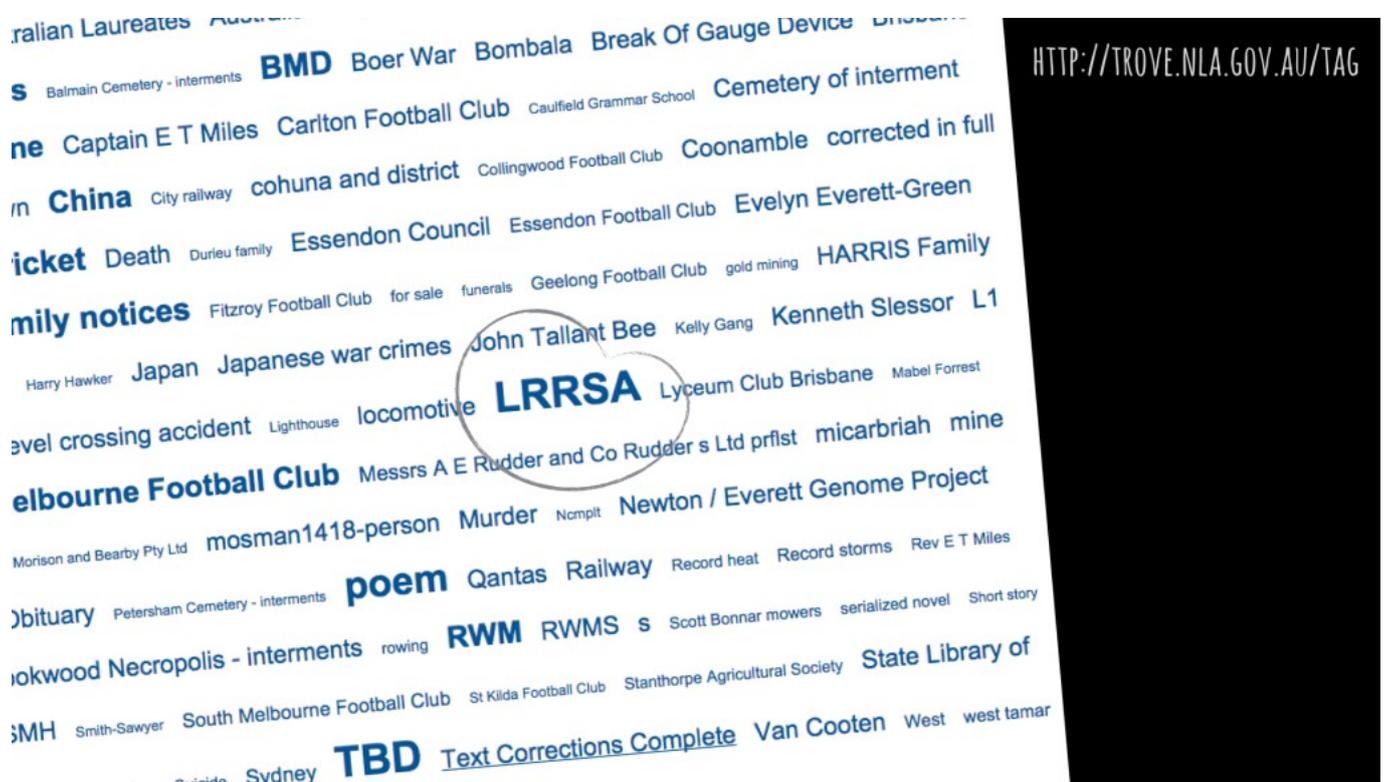
The success of our projects will ultimately depend not on the speed of our servers or the cleanliness of our code, but on the interactions that emerge as our aggregations become part of the simple business of living.

discontents

working for the triumph of content over form, ideas over control, people over systems
<http://discontents.com.au>

People correct our newspapers for many reasons, but few of these motivations are likely to align with our own strategic objectives.

It's not just corrections either. More than 80,000 comments and [3 million tags](#) have been added to resources in Trove. These are just plain text tags, we make no effort to control their content. This creates some interesting possibilities.



<http://trove.nla.gov.au/tag>

I wonder if you can guess the meaning of our most heavily-used tag? It's 'LRRSA' and it's attached to more than 16,000 items.

Any ideas? It's an acronym that stands for the Light Rail Research Society of Australia. Members of the society use the tag to share material of interest — it's become a means of collaboration.

Another popular tag is 'TBD' or 'To be done'. This one's used by text correctors to manage their own workflows.

The numerous guises of a simple tag illustrate the value of 'underspecified tools' — of leaving functionality open to ad hoc elaboration. The boundaries between systems and their use is fluid. Tagging behaviour can extend system functionality.

From machine to human and back again, the limits of what is possible are open to negotiation and change.

My favourite example of this is the work of [one man who has been identifying out-of-copyright sheet music in Trove](#). He's not a musician but he uses his computer to create performances of the pieces. He then uploads the performances to YouTube or SoundCloud and adds a link to them in a comment on Trove. People who find these works on Trove can now just click to hear them. The functionality of the system has been extended without a single line of code being written.

But the permeability of these boundaries means we can't take the roles of people and machines as given. Five years ago, crowdsourced text-correction was a cost-effective solution to the vagaries of OCR, but as the technology improves do we continue to ask humans to undertake tasks that a machine might do more easily? Do we continue to ask our volunteers to change every instance of 'tbe' to 'the'?

While an astonishing 150 million lines of text have been corrected, more than 96% of articles have no corrections at all. More articles are being added all the time and it seems the rate of corrections might be flattening out. The task seems beyond humans alone.

We're currently redeveloping our newspapers interface, making it more responsive, adding shiny new browse features, and improving the overall performance. We'll also be introducing some tools for 'advanced' text correction, allowing our users to modify not only the text, but some of the structural elements of the OCR — inserting new lines for example.

As we investigate opportunities for enrichment of our metadata, I think we'll also need to think about the work we offer our discovery engineers. Correction could extend to geocoded placenames; named entity extraction could be integrated with user-defined relationships.

This technosocial shift is also evident at the other end of our pipelines, when our aggregated data is [consumed and transformed](#).

Except APIs are not really pipelines are they? You don't just turn on a tap, you have to ask the API a question. Our questions interact with the content of the reservoir to shape and colour the flow of data.

An API is a tool for transformation.

New tools and interfaces explicitly change the nature of our aggregations by carrying their use into different realms, by shifting contexts, by asking new questions. Each new use changes how we see the whole.

This is not reuse or recycling — this is remaking. We can dig the tunnels and fill the reservoirs, but it's up to you — the coders, the builders, the developers and the makers — to show us what we've created.

The big challenge is to open up this transformative power to those who have no idea what an API is — people who have important and powerful questions to ask our APIs, but don't know the language.

We need to make sure that the myth of the mega-project doesn't blind us to the human dimensions of our undertaking. Let's foster interventions as well as innovations, activists as well as evangelists. Let's make sure our big ideas make space for other ideas to erupt and grow.

Share this:

- [Click to email this to a friend \(Opens in new window\)](#)
- [Click to print \(Opens in new window\)](#)
- [Click to share on Twitter \(Opens in new window\)](#)
- [Share on Facebook \(Opens in new window\)](#)

discontents

working for the triumph of content over form, ideas over control, people over systems
<http://discontents.com.au>

- [Click to share on Google+ \(Opens in new window\)](#)
-

This work is licensed under a [Creative Commons Attribution 4.0 International License](#).