A national approach to the sharing of data and content

Andy Neale

and

Bill Macnaught (Presenter)
National Library of New Zealand
Wellington, New Zealand

Meeting: **181 — *National libraries and open data: new discovery and access services* — National Libraries**

Abstract:

In 2008 the National Library of New Zealand launched an ambitious programme to help people find, share, and use New Zealand (NZ) digital material. Over the past four years the Library has systematically built a national infrastructure to help open up NZ content and data. The work is based on the premise that the National Library has a role in providing access to the wider collections and knowledge that are outside traditional library networks. This paper will introduce the new tools, partnerships, and approaches that have been needed for opening up broader access to content and data. Topics include the building of national aggregation and API data sharing systems, open licensing, and data mashup and remix initiatives.

Background

In 2008 the National Library of New Zealand became the lead agency for an ambitious programme called DigitalNZ, which sought to help people find, share, and use New Zealand digital content. Born out of the New Zealand Digital Content Strategy, it was one of a number of initiatives that aimed to "chart a course for a content-rich digital New Zealand, where New Zealanders are actively engaged in creating, discovering, sharing and using content in a digital form".

Four years on, the DigitalNZ services are still growing. Not only underpinning many of the public digital access services provided by the National Library of New Zealand, but also representing a new digital collection of New Zealand materials that span organisations and countries. It now sits alongside other national initiatives that are helping to share New Zealand digital content and data.

The DigitalNZ website in 2012

Creating a new New Zealand collection

The creation of a virtual collection of NZ digital content was the main focus in the original setup of the DigitalNZ services. There was a growing collection of digital content that sat outside of New Zealand libraries, and these materials were just as valuable as the content housed within library walls. As a National Library we have a role in helping people to connect to this content even if we aren't physically collecting and preserving it.

What actually is a virtual collection of digital content, as opposed to just a catalogue? For us, it is an aggregated digital collection where we don't necessarily hold a digital copy of the item, but people can still get direct access. The DigitalNZ collection is currently made up of items that are distributed across many partners. We don't hold copies of all that material, but we do have pointers to all the items.

We also make a distinction between a digital record of an item, and any associated digital object. For us, digital content is the combination of a digital record plus a digital object. There must be a digital object of some kind, and it can be in any form such as a document, image, video, sound, or data file; or even a textual object like an HTML webpage.

Catalogues, repositories, and websites

In the era of internet search services like Google, what need is there for a virtual national collection? In fact one of the primary justifications for such a virtual collection is that materials can be hidden or buried on internet search services. There is an aspect of national and cultural identity that is hidden that we can help open up for New Zealanders and the world. The risk being that, if materials are not exposed more fully, we are limiting the opportunity for insight, understanding, and knowledge creation.

There were, and still are, many sources of digital content that have not been indexed by internet search engines because they exist in catalogues, databases, and repositories that web crawlers can't get to. It is also the case that some content can be buried in search results and is difficult to find. For example, many members of the general public would not think to visit museum and library websites around the world for NZ material, but it is there, and not easily found. Unfortunately, it is still true that “if it's not on Google it doesn't exist” for many people.

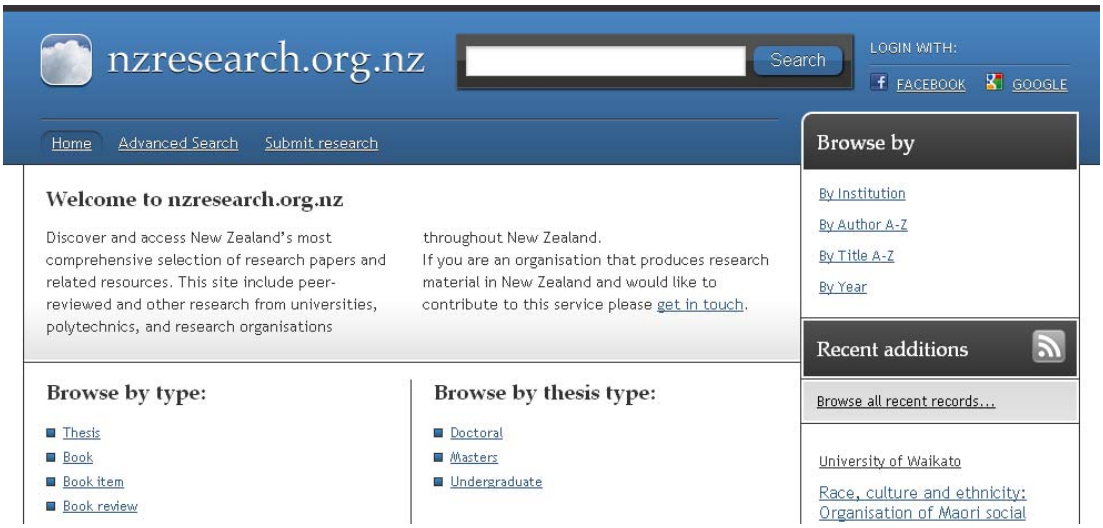
DigitalNZ currently creates its virtual collection from contributions by more than 120 partners. There is no type of digital content that is out of scope, with material coming from across government, universities, the private sector, community groups, broadcasting, museums, and of course libraries and the cultural sector. DigitalNZ collects most of its metadata from NZ organisations, but increasingly also from partners overseas who hold digital copies of material relevant to NZ. In this way we are beginning the process of digitally repatriating NZ-related material.

The rise of the aggregators

DigitalNZ was not the first aggregation service in the world, nor was it even the first in New Zealand. Part of the reason that DigitalNZ was able to grow so quickly is that New Zealand already had a number of sector-specific aggregations that it could build on top of:

- Matapihi (2004) aggregating across New Zealand galleries, libraries, archives, and museums
- KRIS (2007) aggregating across NZ research repositories, particularly universities
- NZMuseums (2008) aggregating across NZ museums

- NZ On Screen (2008) aggregating across NZ film and television
- Data.govt.nz (2009) aggregating across government datasets
-



KRIS / NZ Research service is now hosted by DigitalNZ

Building a national aggregation service

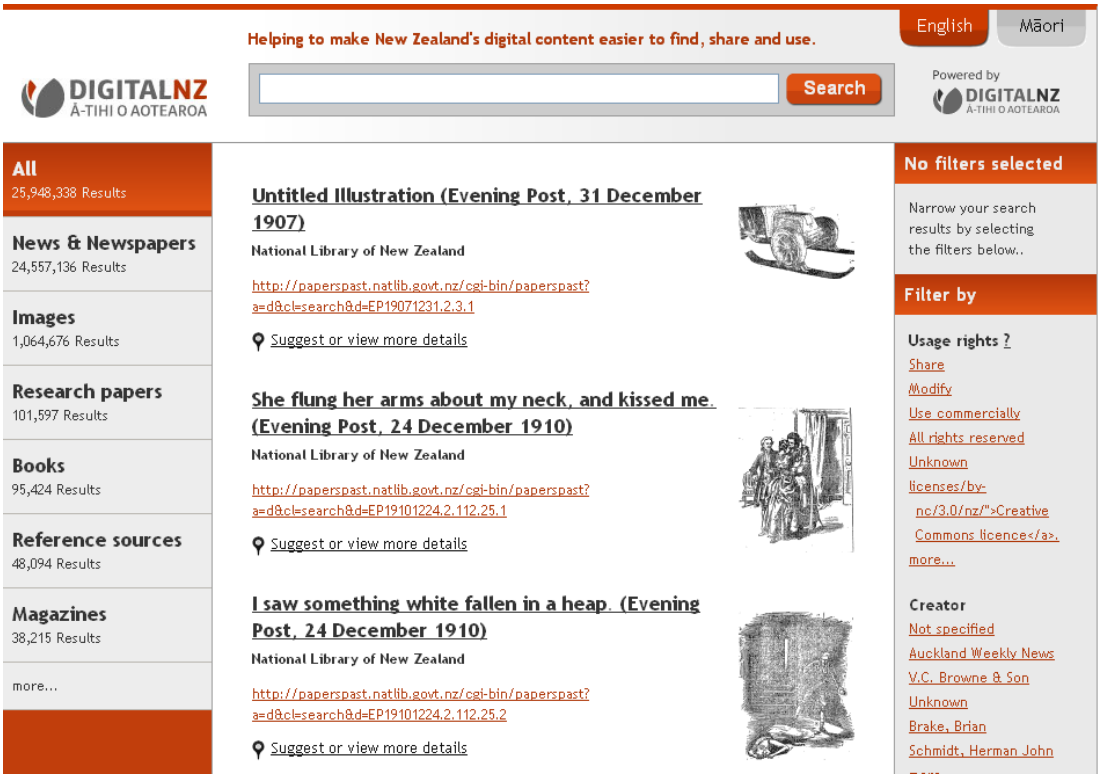
There were several reasons why DigitalNZ was able to quickly establish itself as a national platform with broad support:

1. Eco-system of sector-specific aggregators to build on
2. Low barriers to entry for partners to join, specifically it was technically easy and at no charge
3. A low threshold for metadata quality meant that we did not turn away digital content because of poor digital description
4. Revocable commitments so that partners can pull out at any point in time
5. Use of open source technology allowed quick iteration as new requirements were formed
6. The effort was neutrally branded, not only as a National Library project

In addition, the project started off with a cause to rally partners around. In 2008, the 90th Anniversary of Armistice day (commemorating the armistice signed on 11 November 1918, which ended World War I) was recognised in NZ with the *Coming Home* commemoration. This event was used as the first call for digital content to be part of DigitalNZ. The mini-projects used to launch the service were:

1. Search across collections related to NZ and the First World War
2. Embedded search widget

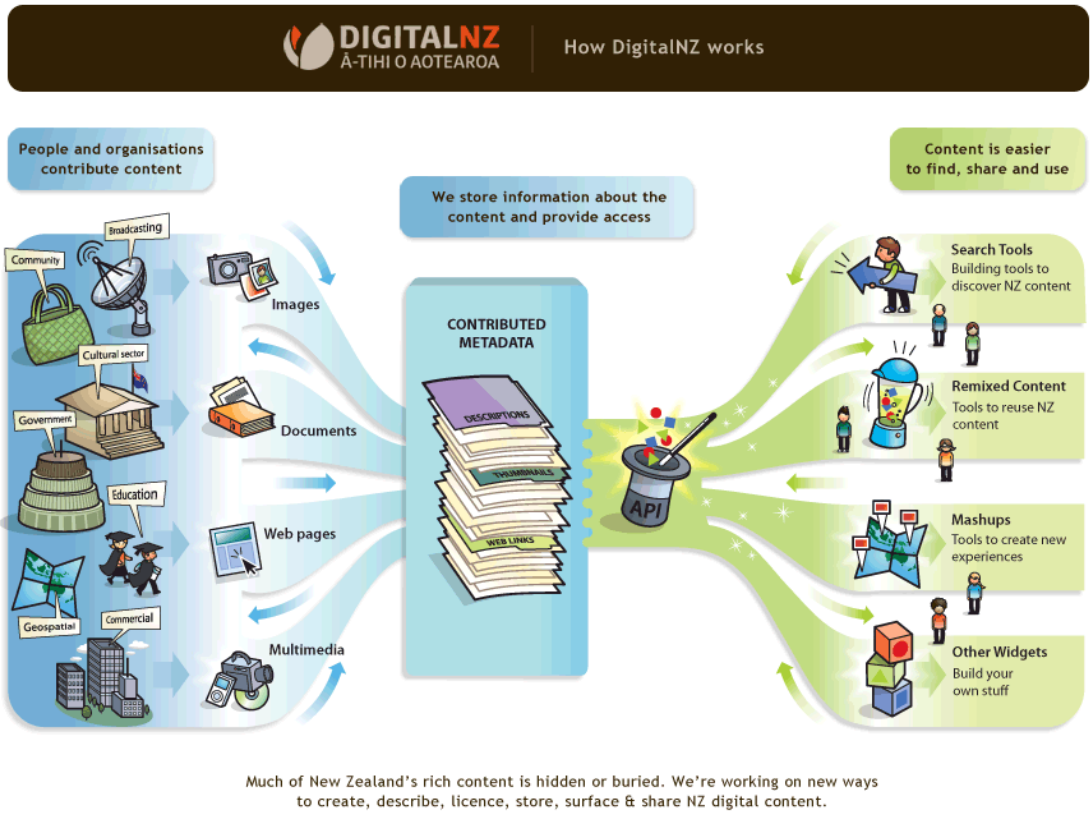
3. Multimedia remix application



The original search service interface at http://search.digitalnz.org

In the process contributing to the commemoration project, partners were in fact taking the first steps toward building a full national aggregation service. At the time, 20 partners contributed approximately 100,000 items to the service. Over the past four years DigitalNZ has grown to be supported by more than 120 partners who now contribute over 25 million pieces of NZ digital content.

Data sharing and APIs



How the DigitalNZ Search Service works

The behind-the-scenes magic that makes a national aggregation service possible is fairly simple:

1. Metadata is extracted from partner systems using software tools developed by DigitalNZ (supporting extraction from XML files, HTML web pages, RSS feeds, and via OAI-PMH and third-party APIs)
2. Each metadata field of interest for a record (e.g. title, description, creator, date, URL link to the resource) is mapped to a common schema based on Dublin Core, and copied to a DigitalNZ database
3. The database of digital content records is shared with other third-party systems by an Application Programming Interface (API) which allows other systems to use the aggregated data

The metadata available through DigitalNZ describes the digital items held by content partners across all walks of NZ life. The APIs can be used to search across titles, descriptions, dates, and creators of the material; as well as requesting image thumbnails and the full metadata record for each item.

The aggregation service collects only the metadata that describes the digital content, but these APIs also return pointers to the item objects and thumbnails put online by content partners.

Find out more about the DigitalNZ API at http://www.digitalnz.org/developer

From metadata to objects

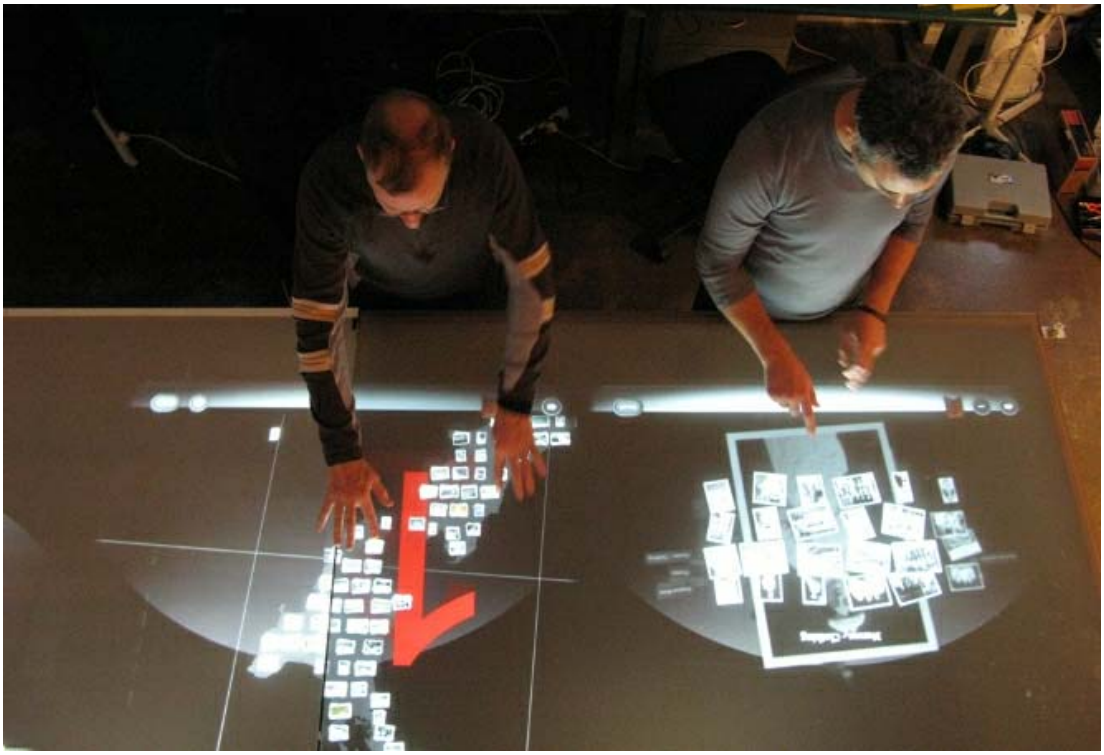
When DigitalNZ first started aggregating digital content, the approach was to point people to the web pages where material could be viewed or downloaded. This is still the approach taken today with many of the search services that are powered by DigitalNZ, such as:

- www.digitalnz.org
- www.nzresearch.org.nz/
- www.ceismic.org.nz/

If you find an item on these sites, you will eventually be pointed to the respective partner's website for the full context and access, not unlike a typical internet search engine.

However, there is growing interest in the DigitalNZ service as a direct provider of digital objects (not via web pages). Imagine the case of a mobile phone or tablet user. If they find an item it is preferable to show that item directly—to let them see the full image or play the video immediately instead of first sending them to another website. As an aggregation service we have a role to play into making objects directly accessible and enabling access on the next generation of devices.

An example of this is a new touchscreen experience called *Lifelines* being developed by the National Library of NZ.



Browsing through people and places on the Lifelines touchscreen

Lifelines is an interesting example of data sharing because the metadata being collected is used by the application to search and display media on a giant touchscreen interface. The data is all driven by the DigitalNZ API, and is also able to locate and display high-resolution images from remote digital delivery services.

Open licensing through NZ GOAL

NZ GOAL is the New Zealand Government Open Access License. It is:

"Government guidance for agencies to follow when releasing copyright works and non-copyright material for re-use by others. It seeks to standardise the licensing of government copyright works for re-use using Creative Commons New Zealand law licences and recommends the use of 'no-known rights' statements for non-copyright material. Creative Commons licences are freely available copyright licences that enable the sharing of copyright works for re-use in a standardised way and in forms that are human, machine and lawyer readable." For more information see http://nzgoal.info/

NZGOAL is an important part of the national approach being taken to make digital content available because it has set a standard approach for government agencies. In turn this raised the profile of the Creative Commons licences and their potential use by the non-government partners that work with DigitalNZ.

The approach taken by DigitalNZ with regard to copyright was an important factor in the set up of the service. The copyright of the metadata, and that of the objects are subject to different rights.

The copyright of the metadata is held by the creating organisations. It is licensed to DigitalNZ in a limited form that allows DigitalNZ to sub-licence its onward use (via APIs) to developers under certain conditions. Partners currently require third-party use to be for non-commercial purposes. This limited license allows partners to withdraw their material from DigitalNZ at any time.

The copyright of the objects is held by the creators of the content, and this does not change when DigitalNZ indexes the material. DigitalNZ will index material that is under any license, including 'all rights reserved'. However, we encourage partners to release materials under less restrictive licenses to encourage reuse.

Remix and reuse

In line with the goal of helping people to find, share, and use New Zealand digital content, an emphasis on the *use* aspect has been important. In particular we have found it important to look beyond just helping people to find NZ digital material. People look for material so that they can use it in some fashion. But organisations often need help in navigating copyright issues.

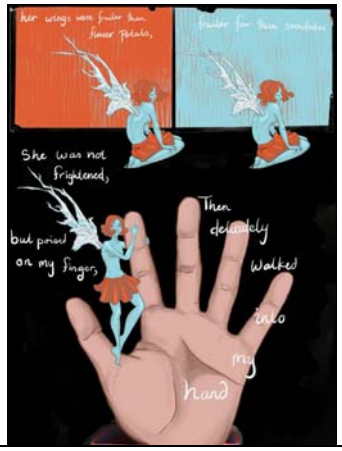
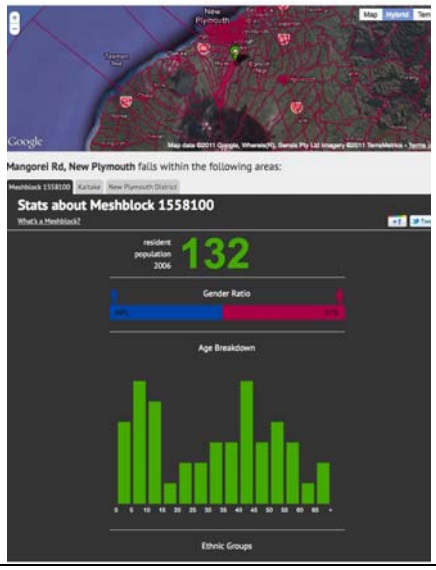
The "Make it Digital" service from DigitalNZ provides a helpdesk service for organisations dealing with digital content, and it includes advice on copyright and for enabling use and reuse. For more information see:

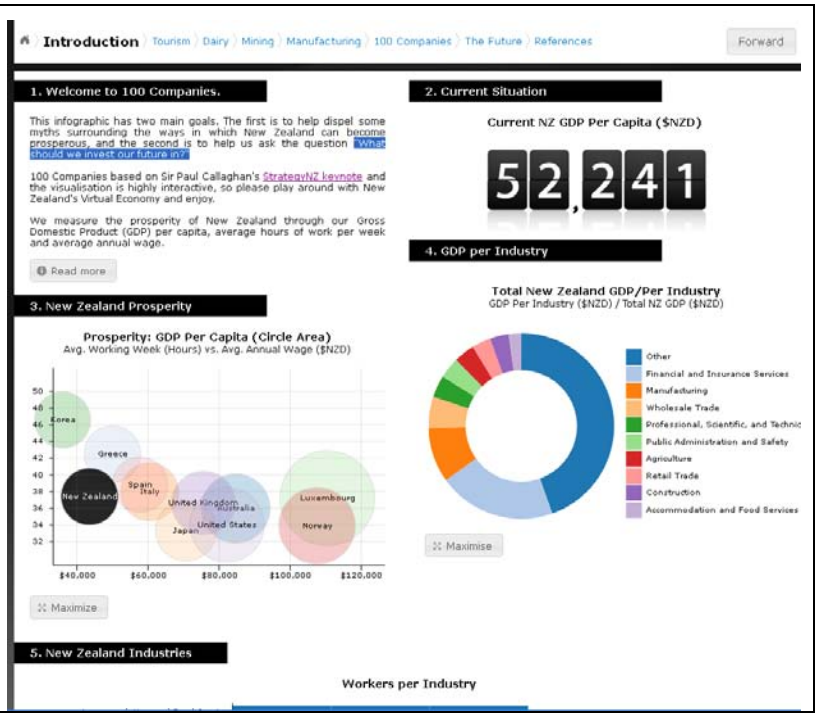
http://makeit.digitalnz.org/guidelines/enabling-use-reuse/

In addition, the Mix and Mash competition was created in 2010 to encourage New Zealanders to make the most of the growing amount of digital content and data available for reuse.

Run twice, in 2010 and 2011, Mix and Mash aimed to encourage the innovative use of NZ content and data so that we could demonstrate the benefits of opening up data and content for reuse. The competition ran over a fixed time period (4 weeks in 2010, and 6 weeks in 2011) and entrants were required to use at least one government dataset or one piece of existing NZ digital content. Entries that were substantially produced before the competition were not eligible.

Over the two years we received a rich diversity of entries:

An Opal Dream Cave, Jem Yoshioka A stunning remix of public domain literature, with heritage photographs combined with the skill of a talented young illustrator	
Mashblock by Cameron Prebble A local demographic data brought together in a more understandable, meaningful and beautiful way	
A Grand Mother by Candy Elsmore The poignant digital story of a great granddaughter discovering her ancestor's name on a dataset of signatories for the New Zealand Suffrage Petitions	

100 Companies by Alex Gibson and Graham Jenson An interactive visualisation that attempts to answer the question ""What should New Zealand invest our future in?"	

Mix and Mash attracted well over 100 entries for each year. In 2011 we were able to encourage the release of twelve new sets of data and content especially in time for the 2011 competition. At least 50 open data sets were used by entrants that year. We also produced a Beginner's guide to Data Mashups and an Educators' Guide to Remix to support entrants. Both resources are Creative Commons licensed and are available for anyone to reuse for their own mashup and remix activities.

Conclusion

The experience in New Zealand has been that a national approach to the sharing of content and data has been beneficial. It has taken some years to build up the relationships that make it possible to co-ordinate initiatives across organisations, and we are now open to the new opportunities that this may create.