

From portal to platform – building new frameworks for user engagement

Sherratt, Tim, National Library of Australia, Canberra, Australia

Abstract

The recent launch of the Digital Public Library of America (DPLA) has highlighted the continuing importance of metadata aggregation services as a way of overcoming institutional and disciplinary divides and providing customised discovery services. But the DPLA's catchy slogan of 'portal, platform, public' points beyond web pages and search boxes towards a more complex stream of user engagement, innovation and advocacy.

Trove, the National Library of Australia's discovery service, provides access to a wide range of cultural heritage collections drawn from libraries, museums, archives, universities and elsewhere. Best known is the collection of digitised newspapers -- a vast resource that currently includes more than 100 million articles. But Trove has always been about more than discovery. Integral to the system's success and growth has been its capacity to provide a platform for user engagement. Trove's annotation and organisation features allow users to customise their experience while, at the same time, enriching discovery metadata.

In addition to existing facilities for user engagement, Trove, like the DPLA, Europeana and DigitalNZ, provides machine-readable access to its aggregated collections through an API. The provision of an API enables the creation not only of new content, but the creation of new applications and interfaces – new ways of using, visualizing, analyzing and enriching the existing metadata.

Based on the experience of developing and maintaining Trove, this paper will critically examine the shift from portal to platform represented by the increasing importance of user annotations and the opportunities provided by the provision of metadata in machine-readable forms.

The Digital Public Library of America was launched in April 2013. Explaining what it actually was, Dan Cohen, the Executive Director, pointed to three key elements: the DPLA was a portal, a platform, and an advocate for open public access to scientific and cultural content.¹

We understand portals – they're just web gateways or starting points. Similarly, the need for advocacy around open access is well-recognised within library and research communities. But what makes the DPLA a 'platform'?

The DPLA is not just a database or a website, it provides a set of tools that anyone can use to build their own application or interface on top of the DPLA's aggregated data. This toolset is called an Application Programming Interface (API). APIs let computer programs talk to other computer programs, enabling application components to fit together like Lego blocks.

DPLA's API, Dan explained, would make 'millions of items available in ways so that others can build creative and transformative applications upon them'. The creators of one such application, StackLife, noted that DPLA's technical architecture meant that 'anyone can write a new way of browsing the DPLA without asking permission'. The API made 'the DPLA's collection an open-ended resource for innovation'.²

Portals are for visiting, platforms are for building on. While a portal can present its aggregated content in a way that invites exploration, the experience is always constrained – pre-determined by a set of design decisions about what is necessary, relevant and useful.

Platforms put those design decisions back into the hands of users. Instead of a single interface, there are innumerable ways of interacting with the data. Instead of a single website, the data is free to be displayed anywhere on the web.

APIs are everywhere. If you use a Twitter or Facebook client you're using their APIs. Most of the social media services you're familiar with are, to some extent, platforms, providing APIs that allow third-party developers to create new apps that interact with the core service.

For some years cultural heritage institutions around the world have been exploring ways of exposing collection data in forms that machines understand – as APIs, but also as Linked Open Data, or simply as CSV files. The idea that this data has value not simply because of what it describes, but because of why it might become, is not new. So why is the DPLA's focus on the importance of the platform so interesting?

What's interesting, I believe, is what happens when you pair the possibilities of the platform with the reach and scale of an aggregation service – when a single key can unlock the cultural heritage of a nation or a continent.

It's early days for the DPLA, they currently provide access to around 5 million resources. But then there's Europeana, pulling in content from countries across Europe and pumping it through their own API – 29 million objects, 36 countries, 1 API.

¹ Dan Cohen, 'Welcome to the Digital Public Library of America', 14 April 2013 <<http://dp.la/info/2013/04/18/message-from-the-executive-director/>> [accessed 17 October 2013].

² StackLife FAQ <<https://stacklife-dpla.law.harvard.edu/#faq/>> [accessed 17 October 2013].

Closer to home, of course, DigitalNZ has long been showing the way – they've had an API since the beginning back in 2008, and it now opens up more than 26 million resources.

And of course, Trove. Our API went live last year and can currently use it to retrieve the details of around 300 million books, articles, objects, images, manuscripts and more.

So let's add those up – 360 million resources, through just 4 APIs. And of course unlike websites, APIs can themselves be connected together.

Earlier this year twelve digital humanists were brought together at George Mason University in Virginia and given a task – in the space of a week they had to design, develop and promote a new digital tool. This was a learning exercise for participants, but what they produced was a web app called Serendip-o-matic.³

If you feed Serendip-o-matic a text or a Zotero library, it will extract keywords from the content and then use these keywords to search for images in a number of collection databases. It's a simple idea, but one that encourages us to think more broadly about connections, about the power of serendipity to point us in new directions – to think differently.

At its launch Serendip-o-matic was hooked up to the Flickr Commons, DPLA and Europeana. 34 million possibilities (plus whatever's in Flickr), 3 APIs, one web app. But it didn't take much to build on those figures. The source code was released to the public at the same time as the app itself. So after a couple hours hacking, I'd connected up Trove and contributed my modifications back to the project.⁴ More than 41 million possibilities, 4 APIs, one web app.

But it's not just about scale and mind-boggling statistics. APIs help aggregators position themselves as something more than discovery services – as information infrastructure – key components in a broader landscape of access, democracy, creativity, research, and economic development.

The desire to foster creativity was one of the motivations cited by the DigitalNZ team back in 2008: 'We want to be surprised by what people come up with', they explained, 'the whole point of putting the open API out there is to drive others to make new, exciting things with the content that we've made available.'

Since that time, many events and initiatives have sought to connect cultural and government data to the people who have the skills and ideas to do something interesting with it. In New Zealand there has been Mix and Mash, in Australia LibraryHack. Europe seems awash with culture hack events, and the DPLA brought developers together to pitch ideas before their API was even ready.

³ Serendip-o-matic <<http://serendipomatic.org/>>, [accessed 17 October 2013].

⁴ Brian Croxall, 'One Week | One Tool: Introducing Serendip-o-matic', ProfHacker, 5 August 2013 <<http://chronicle.com/blogs/profhacker/one-week-one-tool-introducing-serendip-o-matic/51449>> [accessed 17 October 2013].

Earlier this year, the Trove team participated in GovHack 2013. Across a single weekend, 130 teams around Australia were let loose on a tempting buffet of data sources – from weather recordings to public toilet locations.⁵ Fighting for attention amongst them was the Trove API, supported by \$4000 in prizes from NSLA. By the end of the weekend around twenty teams had made some use of Trove. The winners of the NSLA prizes, Pixtory and Advintage, were both mobile applications that brought new perspectives to bear on our aggregated content.⁶

Events such as these are obviously useful in drawing attention to the possibilities of open data. They encourage clever developers to start playing with your stuff, and help data creators and managers see how their painstaking labours can give birth to all manner of unexpected wonders. But there are challenges too. How do you ensure that this explosion of creativity is rewarded, and not exploited – that there are opportunities for clever ideas to be developed further in a way that benefits both the coder and the data provider. How do you broaden the pool of ideas, creating spaces where need and disadvantage are identified and addressed? How do you enable those who have questions, but no technical skills, to be heard?

More generally, how do you demonstrate the connection between creativity and innovation? If aggregation services are going to demonstrate their broader value as a locus for economic, social and cultural development, they have to show that their APIs can sustain innovation – that there is a path from playspace to workplace and beyond.

Europeana describes itself as a 'Core Service Platform' supporting the cultural heritage sector and creative industries.⁷ Focused on 'providing the tools and infrastructure for others to build end-user services', Europeana aims to extract efficiencies from its network and provide 'a catalyst for innovation'. Specific projects such as Europeana Creative, and Europeana Cloud are exploring how these connections can be created and strengthened, while studies of other sectors such as cultural tourism have helped identify opportunities for collaboration.

An aggregator-powered API can also provide new efficiencies and options back to its contributing partners. Europeana presents a series of case studies showing how cultural organisations can use its API to enhance their own local search offerings. DigitalNZ offers the possibility of developing specialised portals, as well as an easy-to-configure, customisable search widget. Trove provides access to user annotations attached to collection items. APIs give contributors the chance to retrieve enriched metadata for use within their own descriptive systems. They can get back more than they give.

For the rest of us, the promise of the platform is that collection data will, as Europeana puts it, find it's way into our workflow:

Instead of trying to bring the user to Europeana, we will take the material to the user. We will do this by developing strategic partnerships, by paving the way for creative re-

5 'GovHack 2013 Report – How the Beast Went National' <<http://www.govhack.org/govhack-2013-report-how-the-beast-went-national/>> [accessed 17 October 2013].

6 Pixtory <<http://hackerspace.govhack.org/?q=groups/pixtory>> [accessed 17 October 2013]; Advintage <<http://hackerspace.govhack.org/?q=groups/advintage>> [accessed 17 October 2013].

⁷ Europeana Business Plan, 2013.

use by developers and by providing the infrastructure that offers opportunities for creating new meaningful ways to access and interpret culture.⁸

The 'in your workflow' model offers possibilities for more efficient discovery, annotation and use. But as Europeana suggests, the decoupling of access from portals also brings opportunities for innovation through contextualisation and interpretation – new interfaces can be built, new visualisations framed, new meanings made and stories told.

As a digital historian myself, I was playing around with the Trove API before I became part of the official team. What excited me was the possibility of moving beyond discovery as the primary mode of online interaction with cultural collections. Instead of just finding interesting newspaper articles, for example, I could start to explore trends and patterns across the aggregated resource.

QueryPic is a web app I built that graphs search results from the Trove newspaper database over time. It's a simple tool, but it allows you to ask new new types of questions. And when you've hooked up one API, why not another. By using DigitalNZ to access PapersPast, I made it possible to compare results from Australia and New Zealand.⁹

But why should we just be targetting our workflows? Why shouldn't we be working to get cultural collections 'in our lives'. We also have to remember to that while APIs can liberate collections from the prison of portals, they themselves have boundaries and limitations. You can still only do what the API lets you do. We have to guard against creating a new set of silos.

After attending a DPLA planning meeting, Ed Summers reflected on the characteristics of a 'generative platform', suggesting that projects like DPLA need to look at the broader environment within which metadata is published, harvested, exposed and linked:

I guess it's a no-brainer that for it to succeed the DPLA needs to be aligned with the ultimate generative platform of our era: the World Wide Web. Name things with URLs, create typed links between them, and other people's stuff.¹⁰

Ed argues that aggregators, like the DPLA, should work with the the grain of the web, not against it. APIs are convenient and well understood by developers, but they shouldn't define the principles of our platforms.

David Weinberger, a well-known thinker about the culture of the net, directed the technical team that developed the DPLA's API. But he also reflected more broadly on the nature of platforms and how they relate to the work of libraries.¹¹ Switching our thinking from portals to platforms, he suggests, means thinking about libraries 'as infrastructure that is as ubiquitous

⁸ Europeana Business Plan, 2013.

⁹ QueryPic <<http://dhistory.org/querypic>> [accessed 17 October 2013].

¹⁰ Ed Summers, 'The Dpla as a Generative Platform', 25 May 2011 <<http://inkdroid.org/journal/2011/05/25/the-dpla-as-a-generative-platform/>> [accessed 30 August 2013].

¹¹ David Weinberger, 'Library as Platform', Library Journal, 2012 <<http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/>> [accessed 30 August 2013].

and persistent as the streets and sidewalks of a town, or the classrooms and yards of a university'. But embracing the possibilities of the platform means giving up the ability to control, or even predict, what will be constructed using this infrastructure. Indeed, he notes that 'a platform gains value the less can be predicted about what will be built with it'.

By exploring these two characteristics – ubiquity and unpredictability – we can move beyond the platform as set of methods, or a technical framework, to conceive of it more broadly as an open, creative space full of possibilities. Thinking like a platform, not a portal, or a website, means enabling a set of opportunities, experiences and emotions, that are yet to be known, or even imagined.

EverywhereNess

Aggregation services collect stuff. We vacuum up metadata, do a bit of cleaning, then file it away for future reference. Through our websites and APIs we show people all the cool things that we've found – Look here! Did you know about this? But platform thinking asks us to expand our notions of dissemination to think about how we can be everywhere at once – part of the fabric of online existence, and not just a service to be queried. How can we, supernova like, explode our aggregations, sending the building blocks of life out into surrounding space? Perhaps we can find some hints in the ways new communities are already emerging around our collections.

Every Tuesday a loosely-organised group of bloggers post about their latest Trove discoveries using the tag #TroveTuesday.¹² It wasn't our idea. Indeed, of our involvement is generally limited to a few retweets. It was the brainwave of one passionate Trove user, taken up by others who simply want to share the excitement of discovery – as of course we all do.

In a similar way, we've noted that one of our main referrers, alongside the likes of Google and Wikipedia, is a service called Ravelry. It's a site for knitters. Ravelry users trawl through newspapers and the Australian Women's Weekly looking for knitting and other craft patterns, which they collect and share on the site.¹³ Again, this was not something we planned or initiated. It just happened.

In the age of social media we are all sharers. But what these examples remind us is that there's something really fundamentally human about the desire to share that need not be mediated through the APIs of social media services. There is more to sharing than a 'Like' button – it can evolve its own forms in and around the specific technologies we provide.

But of course there are things we can do to help. We can develop tools, employ standards, and pursue practices that not only simplify the process of snipping and sharing, but add a bit of useful structured data into the mix.

Paul Hagon, from the National Library of Australia, has created a Wordpress plugin that makes it easy for users to embed the contents of a Trove newspaper article in their own

¹² Amy E H Lehmann, 'Trove Tuesday', <<http://branchesleavespollen.blogspot.co.nz/p/trove-tuesday.html>> [accessed 17 October 2013].

¹³ 'Trove Users Still Crafty!', 18 July 2013 <<http://trove.nla.gov.au/forum/showthread.php?1529-Trove-users-still-crafty!>> [accessed 17 October 2013].

posts – complete of course will a full citation and a link back to the site. One Trove user has created their own browser plugin for quickly capturing newspaper citations, while I've developed a translator for Zotero that let's you save structured metadata from the newspapers site into your own research database.

There's no mystery now about embedding useful structured metadata in your web pages to support discovery and re-use. Far from being the arcane pursuit of semantic web nerds, embedded metadata is increasingly powering tools like Google and Facebook. But we don't have to be constrained by the needs of search engine optimisation. We can give to Google what is Google's, while adding richer metadata to support uses yet unfathomed.

Hypothes.is is a new service that promises to provide annotation at web scale – the possibility of linking resources anywhere within a structured network of discussion, debate, interpretation and meaning. And it's building all this using, amongst other things, some simple embedded metadata. By working with the grain of the web we leave our options open.

UnclosedNess

The more we become aware of the power of networked information, the more we become concerned with making and preserving its 'openness'. To me open data is a process not a product - each visualisation, or interpretation can challenge our assumptions and help us to see things differently. Each use is an opening into new contexts.

'A library platform', David Weinberger argues, 'should be measured less on the circulation of its works than in the circulation of the ideas and passions these works spark'. He imagines platforms giving rise to 'messy, rich networks of people and ideas'.¹⁴ But how can we avoid the temptation to clean up some of the messiness, to close some options for the sake of efficiency or a more familiar user experience?

Game designers are grappling with similar questions. Intrigued by the success of games like Minecraft and services like Pinterest, one designer has contrasted the experience of 'sandboxes' – online spaces for open, collaborative play – with 'pathways', where the narrative journeys are largely predetermined.¹⁵ Another has rather evocatively described this as the difference between 'exhaustibles' and 'possibility engines'.¹⁶

So how do we create possibility engines? One suggestion is that we should underspecify the tools we offer users – we shouldn't seek to tie them too closely to particular tasks. We need to leave some space for subversion, for people to use our tools in ways we don't expect. The humble tag, for example, can have many uses – from search enrichment to community

¹⁴ David Weinberger, 'Library as Platform', Library Journal, 2012
<<http://lj.libraryjournal.com/2012/09/future-of-libraries/by-david-weinberger/>> [accessed 30 August 2013].

¹⁵ 'Poetpainter: From Paths to Sandboxes' <<http://www.poetpainter.com/thoughts/article/from-paths-to-sandboxes>> [accessed 13 September 2013].

¹⁶ Sebastian Deterding, 'Don't Play Games With Me! Promises and Pitfalls of Gameful Design', 2011
<http://www.slideshare.net/dings/dont-play-games-with-me-promises-and-pitfalls-of-gameful-design?from=ss_embed> [accessed 12 October 2013].

organisation. However, the range of uses can also be easily curtailed by attempts to normalise text strings or control vocabularies – to clean up some of the messiness.

Users can enrich the content of Trove in a number of ways: by correcting OCR'd newspaper articles, or by adding tags and comments. They can also create lists. Lists are simply collections of resources and provide a convenient way for people to save and share their research. DigitalNZ has something similar, called 'sets'.

There are currently more than 27,911 public lists on Trove created by 5,796 users. Those lists contain almost half a million items. Analysing the frequency of words in the titles of these lists reveals, unsurprisingly, that most are related to family history. But there are also some unexpected joys – such as the fact that there are 107 lists about lawn mowers. Lists are a very simple tool, but with that simplicity comes an open invitation to the making of meanings.

And to take us back to our starting point, the contents of public lists are themselves accessible through the Trove API. So people can re-use their collections in other contexts, or embed in other tools.

But who will?

If we're trying to kick start the engines of possibility we have to admit that the biggest barrier is not the over-specification of our APIs, but the lack of access to technical skills, knowledge and support. Few of us can take an API and simply build something.

This means, I think, that we have to consider another aspect of platforms – elevation – platforms can lift us up.

Elevation

'The DPLA', Dan Cohen noted, 'will serve as an on-ramp, allowing local and regional organisations to move into the digital realm, and to help DPLA users build essential digital skills'.¹⁷ Europeana's API has similarly been described as offering a 'migration path' for organisations, helping them shift their thinking towards 'a layered, web service based information architecture'.¹⁸

DigitalNZ's 'Make it Digital' service is a great example of this sort of skills development in operation, and we have to do more of it, recognising the effects of unequal access to technical expertise. Otherwise whose collections will be aggregating?

¹⁷ John Palfrey, 'What Is the DPLA?', *Library Journal* <<http://lj.libraryjournal.com/2013/04/future-of-libraries/whats-is-the-dpla/>> [accessed 17 October 2013].

¹⁸ Cesare Concordia, Stefan Gradmann and Sjoerd Siebinga, 'Not (just) a Repository, nor (just) a Digital Library, nor (just) a Portal: A Portrait of Europeana as an API', 2009 <[http://www.researchgate.net/publication/228539235_Not_\(just\)_a_Repository_nor_\(just\)_a_Digital_Library_nor_\(just\)_a_Portal_A_Portrait_of_European_as_an_API/file/79e4150acb481b79f8.pdf](http://www.researchgate.net/publication/228539235_Not_(just)_a_Repository_nor_(just)_a_Digital_Library_nor_(just)_a_Portal_A_Portrait_of_European_as_an_API/file/79e4150acb481b79f8.pdf)> [accessed 17 August 2013].

I think platforms also have a special obligation to their users to be open about their limitations and biases – about the artificial, constructed nature of the aggregation itself. I would like to think that this discussion could broaden into a deeper critical understanding of search technologies, an opportunity to open some black boxes, and pull apart some of the uncontested algorithms that are shaping ever more aspects of lives.

Neither can we forget our non-users. According to a recent evaluation survey of Trove the people we are not reaching include ‘the young, the less affluent, the less well educated, Indigenous Australians, and the large proportion of the Australian population for whom English is not the primary language spoken in the home’.¹⁹ How do we lift these groups up to join us on the platform?

All of this seems rather daunting, but I think the thing that excites me most about the DPLA, Europeana, DigitalNZ and Trove, is the possibility that we might be exercise the power that comes through aggregation to deliver some leadership in these sorts of areas. We should remember that platforms are also for speaking from.

19 Marie-Louise Ayres, ‘Singing for Their Supper’: Trove, Australian Newspapers, and the Crowd’, 2013 <<http://library.ifla.org/245/>> [accessed 12 October 2013].