

Social Media Collecting at the National Library of New Zealand

Bill Macnaught

National Librarian

National Library of New Zealand, Wellington, New Zealand

bill.macnaught@dia.govt.nz



Copyright © 2018 by Bill Macnaught. This work is made available under the terms of the Creative Commons Attribution 4.0 International License: <http://creativecommons.org/licenses/by/4.0>

Abstract:

Collecting content from the internet is an increasingly significant part of collection building at the National Library of New Zealand. Social media collecting is a new aspect of our digital collecting. We currently collect social media both under our legal deposit legislation and through donation as part of personal papers or archives. Social media offers unique content and voices, not always available in other formats. While this gives us new opportunities to diversify our collections, it isn't without challenges. Content is shifting away from traditional websites to social media. This is understandable – it's easier to post content, quick to circulate and cheaper. However, it is also comes with new collecting challenges.

Keywords: Collection development, Digital collecting, Social Media collecting, Legal Deposit legislation.

Building social media collections: the New Zealand legislative environment

The National Library of New Zealand (Te Puna Mātauranga o Aotearoa) Act 2003 specifies that the purpose of the National Library is “to enrich the cultural and economic life of New Zealand and its interchanges with other nations by (...) collecting, preserving, and protecting documents, particularly those relating to New Zealand, and making them accessible for all the people of New Zealand, in a manner consistent with their status as documentary heritage and taonga” [S.7(a)] - in both physical and digital formats.[*Taonga = ‘treasure’ in the Māori language, which is an official language of Aotearoa New Zealand.*]

The division of the National Library which holds our heritage collections is the Alexander Turnbull Library (founded by a bequest to the nation by Turnbull in 1918). The Act specifies that these collections are owned by the Crown in perpetuity [S.11(2)], and that the Alexander Turnbull Library must “preserve, protect, develop, and make accessible (...) the collections of that library in perpetuity (...)” The Alexander Turnbull Library is also required to “develop and maintain a comprehensive collection of documents relating to New Zealand and the people of New Zealand.” (S.12)

Legal deposit - provided for in the Act - enables this, ensuring that New Zealand publications will be available to users now and into the future.

The National Library Requirement (Electronic Documents) Notice 2006 (“the Requirement”) authorises the National Librarian to make a copy of a public document published on the Internet, effectively allowing us, under certain conditions to collect digital content directly from the Internet.

Digital collections are also acquired for the Library through donation and purchase. For example the first Twitter archive of a politician came to the Library as part of a donation of personal papers in 2009, but the second came in through our web archiving programme in 2013.

In a rapidly and continually evolving digital environment we need to understand whether the legislation and associated legal deposit requirements, as well as our own institutional policies are still fit for purpose. Some of the questions we grapple with are:

- Should all social media be treated the same?
- Or are there different approaches depending on the way a social media platform is being used by society?
- How do we balance our mandate to collect with our need to protect the privacy of New Zealanders? How do we ensure that we are a trusted repository for people’s personal data (and metadata)?
- Do social media users see themselves as ‘publishers’ and expect their content to be in scope for collecting by a state agency?
- What kind of access to these collections can we provide if we haven’t sought permission prior to collecting?
- It is worth noting that in New Zealand legislation there is no ‘right to be forgotten’ corresponding to decisions that have been seen in some other jurisdictions; and it is not clear, at the moment, that we need an ‘RTBF’: the issue can perhaps be better addressed in other ways.

A number of issues arise when collecting Internet material under current New Zealand legal deposit law. These are: availability, territoriality, copyright, and the terms and conditions of use for online social media Internet sites. These issues also surface the application of New Zealand law to global Internet sites on which New Zealand content is published.

There are no easy answers, and for the Library to realise its purposes under our current legislation, we need to balance the risk of missing out on collecting content because of increased time and effort required to collect it, against the processes that help manage risks with Internet platforms and providers.

These issues can be further complicated by the extent to which social media content ‘published’ on the Internet may still be considered private or personal information by the content creators. This requires the library to apply long established library selection and collection development principles to a new kind of content in order to build useful digital collections, but also to maintain the public’s trust in the National Library as a public good.

Availability

In New Zealand legal deposit ‘availability’ means the document is available to the public on the Internet at the time of collection (even if it is behind a paywall or other restriction). If content is made available on the internet it has essentially been published.

If the published content is freely available online, the National Library can provide open access to its archived copies, otherwise it can only provide access onsite in the Library.

If the original content is taken offline or online access is restricted, the content creator can require that we provide access to the archived versions only onsite, in our Reading Room.

What does that mean if someone deletes one tweet online that was part of a much larger Twitter data set already collected by the Library?

Territoriality

This issue is concerned with whether the document is produced in New Zealand, or whether its production is commissioned from New Zealand.

In this context “produced” relates to the operational aspects of the production or publication that are comparable to printing a hardcopy document. For Internet documents this is usually the content creation and uploading to a platform, whether to the “producer’s” own website or a shared platform like a social media site.

Complications arise if the producer or commissioning party is only resident in New Zealand for some of the time in which the applicable document was produced. This is a particular concern in relation to documents like blogs that build over time, or social media feeds. In the current legal context in New Zealand, if posts to an individual’s blog or social media feed are produced in different countries over time, some parts may be eligible for collection, while other parts are not. This is difficult to prove either way without the cooperation of the author or the producer and we might end up in the position of having to remove sections of a blog, if challenged.

A related issue is the way in which “publisher” is currently defined under our Act. The publisher is the entity that we can require assistance from to collect copies of content that is determined to be in scope for legal deposit, but under our definition it could often be considered to be the foreign entity that runs a particular social media platform, rather than the producer uploading content to the platform.

All of this requires us to have alternative mechanisms available for collecting this content if we decide it is valuable for our collections, e.g. negotiated deposits (donations). It also means looking forward to what new legislation might look like that empowers collecting New Zealand-produced content in an increasingly networked, global context.

Copyright

Copyright must exist in a document before it can be in scope for legal deposit in New Zealand. In collecting Internet material we are almost certainly collecting some material that is an infringing copy of other work. However, it would be onerous and unreasonable to

expect the National Library to background check all collected Internet material for possible copyright infringement.

We need to be aware of the risk that the Library's collecting of digital material may sometimes include the collecting of infringing material. To manage this risk properly, we need to have clear policies requiring the immediate suppression of all content when a claim is made, and we need to be prepared to remove infringing content from the collection if a claim is upheld.

Terms and conditions

It is very common for the terms of use of a website or platform to prohibit the use or downloading of content other than for personal and non-commercial use. If the website or social media platform is operated by a New Zealand entity, and is therefore subject to New Zealand law, the terms of use for that website should not be able to override the Library's statutory ability (and responsibility) to collect content under legal deposit. In other words, a New Zealand website operator should not be able to use terms and conditions to contractually restrict the Library from fulfilling its statutory function. Further, if that entity is governed by New Zealand law, it can be compelled to assist the Library as a "publisher" under section 33 of the Act.

However, this will not be the case for a website or platform operated by a foreign entity from outside New Zealand, as is often the case. Such an entity is not ordinarily bound by New Zealand statutory obligations, notwithstanding the fact it may fall within the definition of the "publisher" in the Act.

As more and more New Zealand content is sited on global platforms, we will need to work carefully to understand the limits of our current legislation and build a fit-for-purpose authorising environment.

Some social media collecting case studies

When the Kaikōura earthquake hit in November 2016, we ran a pilot Twitter crawl to document how the disaster was covered, discussed, and reacted to on Twitter.

Creating a set of parameters of what to collect from Twitter is different from identifying a list of books or even websites to collect. This is a constructed appraisal decision by web archivists who determine the parameters of the harvest, including which accounts to crawl, which hashtags and keywords or subjects to search, and which dates to use to limit the crawl.

Once we analysed the data which we had collected we were able to determine, based on how the hash tags and keywords had been used, that they couldn't be limited to New Zealand, and thus couldn't be collected under our current Legal Deposit framework.

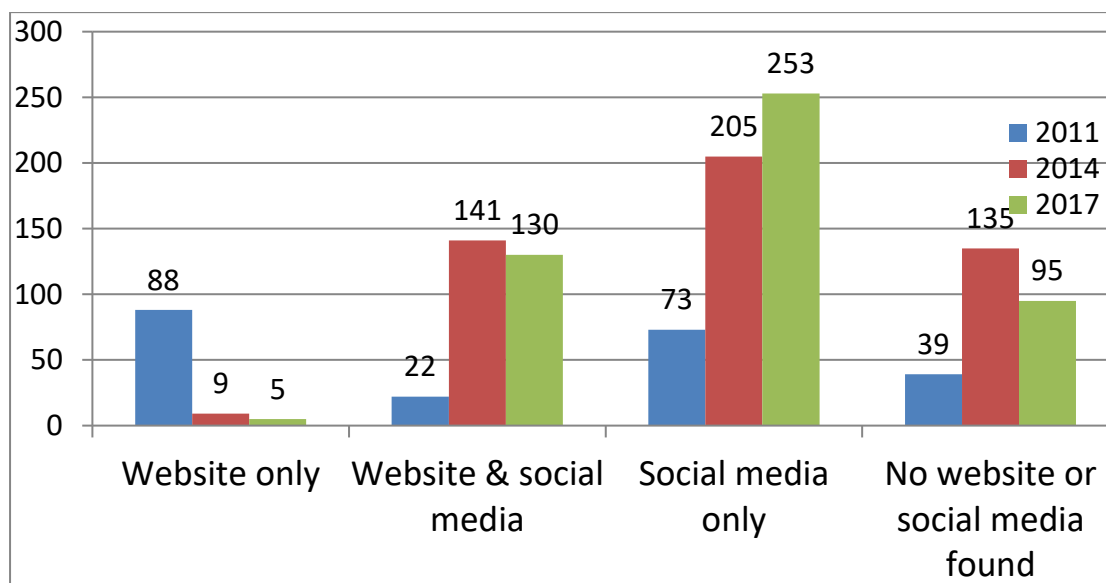
An excellent example of the global reach of social media is an image of stranded cows on a rural hillside. The content started with an aerial photo from a ruined farm on the fault line near Kaikōura. The image then went all around the world in subsequent days and was re-used across multiple media types including television. (In our Twitter harvest data set, the image of these cows was referenced over 28,000 times, definitely extending beyond the reach of New Zealand's legal deposit territoriality.)

We know that documenting responses to natural disasters is important - from the real and important work of rescue and recovery, to the whimsy of stranded cows on a hill-side. Together they give us a richer understanding of their impact on the natural world, and the people and animals that rely on it.

The National Library of New Zealand has been doing general election web archiving since 1999 – (in New Zealand, Parliamentary elections are called ‘general elections’). In 2005, we began collecting blogs as part of our web archiving, noting that they offered “unique and varied perspectives” on an election which were not otherwise available. In the 2014 general election, we noted another significant shift from traditional websites and blogs to using social media almost exclusively. Many candidates either had a largely static website, or abandoned using a website altogether. Instead they opted to use social media as their main campaign platform.

While a lot of candidates in 2011 only had a website, very few stuck with that choice in 2014, with only 9 candidates having only a website, compared with 205 using only social media to communicate with their constituents, (Table 1). Hundreds added some form of social media, or went with just Facebook and Twitter, and no website. The number of candidates using only Facebook stayed level, while some used Twitter only. Interestingly, a large number of candidates had no website or social media presence in 2014.

Table 1: Comparison of Candidates’ Online Presence, 2011, 2014, & 2017 General Elections



Preparing for the 2017 general election, we knew that the majority of candidates had moved exclusively to social media and we would have to include that content where possible if our harvests were to accurately document the elections.

While we have harvested websites for previous New Zealand General Elections, 2017 was the first time we also collected social media content (Twitter). We tried a multi-pronged collection process for the 2017 elections harvest. Our Twitter crawl included identified hashtags, candidates for office, political parties, keywords to search, and the dates.

We also contacted individual candidates and asked them to download their entire Twitter archives directly from Twitter and donate those to the Library as well.

What we've learned

Social media platforms have proliferated over the last couple of decades. The Library has so far collected small, representative examples from Twitter, BlogSpot, WordPress, and Tumblr, along with some videos provided by content owners that are posted on YouTube or Vimeo.

You might notice that a certain rather large social media platform is missing from that list. While Facebook is used by large numbers of people throughout the world, including in New Zealand (currently 2.3 billion people, compared with Twitter's 336 million people), Facebook is also much more restrictive about harvesting or collecting people's content from the site, as it's in the business of selling that data to advertisers. While we now know Facebook has large amounts of data on all its users, its data policy does not allow for web crawlers to crawl the site, nor does it provide the kind of easy API access to content that Twitter provides.

It's also worth thinking about the different ways people use Facebook. Where Twitter is by default an open platform, with the expectation that you are sharing with the world, Facebook users often perceive their Facebook account as a private area, restricted to their "friends". As a trusted repository, for the Library it is worth thinking about how that perception should or should not inform our collecting decisions.

While we don't yet have definitive answers in that area, and expect that any answers we do come up with will continue to be context-dependent, we are exploring a donation-based Facebook collecting pilot this year, one in which we seek donors' permission before collecting their content and restrict access for a period of time by default.

Conclusion

This paper has cast an eye over some of the issues confronting the National Library of New Zealand's legal deposit collecting and web archiving practice in the social media context, as we grapple with increasingly complex materials that challenge traditional workflows and concepts.

We continue to experiment and explore new content on social media platforms, seeking to understand what we can collect and how should we collect it. As we embark on this work we need to ensure that as we collect, we are also paying attention to emerging issues related to privacy, the 'right to be forgotten', the balance between people's experience of natural disasters, for example, and the record of government (the social media feeds of MPs in a general election, for example).

Our challenge is to ensure we strike the right balance between representative and comprehensive collecting. That we understand the technical and legal issues involved in determining what is above the line (publicly available) and what is below the line (direct mail, photographs, and other private material which would enrich our collections over time). We need to be mindful of the rise of artificial intelligence and data mining/analytics techniques, both in how they can assist our collecting and processing activities, but also how we can protect our donor and user privacy.

We also need to be asking ourselves more explicitly about the impact on individuals of our collecting of social media. Some of the questions we are working to understand more fully include:

- do people see their lives reflected on Facebook accounts and other social media as public or personal information?
- are there different approaches to different social media and how do we agree the shape of a multi-layered framework and the related ethics of collecting?
- when do we need to ask permission to collect and when do we not, and what are the access considerations attendant on those decisions?

We are also exploring digital ‘archival’ material created entirely online, but not ‘published’ in the sense intended by the legislation on legal deposit as we understand it. This means content such as correspondence, financial records, minutes, photographs and reports now being documented behind logins on online platforms including Google Docs, Facebook or direct messaging on Twitter and Instagram. How will these changing means of communication and information management impact our efforts to continue building documentary heritage collections that reflect New Zealand’s diverse society? How do they correspond or counter our collecting under Legal Deposit?

In working through these issues we must balance the ideal with the practical, acknowledging that sometimes the answers to many of our questions may be “it depends”.

However, we also know that not collecting this content means we would be failing to serve the people of New Zealand and failing to realise the purposes of our legislation in relation to New Zealand’s published and unpublished documentary heritage.

There is a long way to go yet, but the benefits for our citizens in the future will be enormous in terms of collections that reflect the diverse online activities of all sectors in our society.

Acknowledgments

I’m glad to acknowledge the contributions to this paper made by my staff members Jessica Moran, Amy Joseph and Steve Knight.