

Web Archiving in Singapore: The Realities of National Web Archiving



Ivy Huey Shin Lee and Shereen Tay

Abstract This chapter describes the challenges of web archiving in Singapore, where awareness about web archiving is low. It focuses on the legislation amendments, infrastructure enhancements and public engagement activities required to facilitate web archiving work at a national scale.

1 Background

The National Library, Singapore (NLS) is a knowledge institution under the National Library Board (NLB). The NLB manages 26 public libraries, the National Library, the National Archives and the Asian Film Archives. NLS is the custodian of Singapore's documentary heritage and has a mandate to preserve the published heritage of the nation through the legal deposit of works published in Singapore. This mandate came into force in 1958 through the enactment of the Raffles National Library Ordinance, which eventually came under the NLB Act in 1995. The legal deposit scheme has served NLS well in the last 60 years, enabling the library to systematically collect and preserve more than 1.3 million items over time. However, this mandate is limited to physical items and does not extend to electronic formats such as e-books and online materials. NLS can thus only encourage publishers of such electronic publications to deposit these materials voluntarily as it cannot legally enforce this under the legal deposit scheme that covers only physical materials. While NLS can use web archiving tools to capture online content, it has to seek the written consent of website owners before doing so, which is administratively cumbersome and yields a low response rate. As the volume of digital content explodes, it has become a priority for NLS to update its legislation to save materials from disappearing into the digital black hole.

I. H. S. Lee (✉) · S. Tay
Statutory Functions and Research, National Library, Singapore, Singapore
e-mail: Ivy_LEE@nlb.gov.sg

Electronic and Internet materials are increasingly important resources that capture national identity and culture; hence, it is essential for NLS to have a legal mandate to collect digital content. This is especially true for materials published on the Web. As of December 2019, there are more than 181,000 websites in Singapore (.sg domain),¹ and this number is growing rapidly. If published materials on the Web are not systematically collected in a timely manner, there will be a permanent loss of Singapore's published cultural heritage.

2 Amendments to the Legislation

NLS takes its cue from forerunners at other national libraries who have in place legislation and processes that facilitate the collection of digital legal deposit materials and web archiving. In the drafting of legislative provisions, NLS became aware of the common issues that other national libraries encountered, especially those based in developed economies. Privacy concerns and copyright matters consistently turn up at the top of the list and need to be addressed. Bearing in mind these two key areas of concern, NLS drafted proposed amendments to balance the rights of content creators and publishers while also meeting the library's mandate to preserve digital cultural heritage and provide access to the public.

Guidelines were also prepared for handling operational matters which the law may not cover in detail, such as the number of access points and takedown requests for the materials. According to these guidelines, NLS will provide two terminals (or two concurrent accesses) on the library's premises for viewing digital legal deposit materials and archived websites. Two access points were chosen as this mirrors the way in which the physical legal deposit collection is available to the public in Singapore, that is, two copies. This restricted access will help to protect the content owners' commercial interests while also ensuring that members of the public have access to these materials.

Based on NLS's research and findings, it proposed four main areas for update to the legislation:

- (a) To extend the scope of NLS's legal deposit scheme to incorporate digital materials, including electronic counterparts of print publications as well as online content.
- (b) To empower NLS to collect online materials through web harvesting without requiring it to seek written consent from the owners/producers. Without amending the NLB Act to authorise NLS to collect online content, the library would have no protection against potential legal liabilities such as copyright infringement. NLS will only collect online content within the .sg domain, as its focus is on Singapore. In addition, the library will not harvest protected online

¹See Singapore Network Information Centre (SGNIC) at www.sgnic.sg

publications and materials (e.g. sites protected by password), unless they are voluntarily deposited by the website publisher or owner.

- (c) To give NLS the right to digitise, copy, preserve and use the deposited materials for non-commercial purposes. This is important for NLS to carry out tasks like digital preservation, where content will need to be preserved digitally (including migration to new formats, as old formats become obsolete due to technological advancements) to enable perpetual access.
- (d) To enable digital materials collected/archived by NLS to be accessible to researchers.

NLS also consulted legal agencies in Singapore and was advised that a complementary amendment to Singapore's Copyright Act would be necessary. A decision was made to include updates to the Copyright Act simultaneously.

In July 2018, the Bill to amend the NLB Act was passed by the Singapore Parliament and the changes came into effect on 31 January 2019. The updated NLB Act now empowers NLS to collect, preserve and provide access to digital legal deposit publications, as well as to Singapore websites.

3 Engagement with Stakeholders

During the drafting of the legislation, NLS engaged with stakeholders and the general public to get feedback and support for the proposed amendments. In Singapore, it is common practice to seek inputs from stakeholders and the general public via a public consultation exercise whenever government bodies implement or amend legislation.

Due to the complexity of the legislative amendments, NLS designed a series of infographics to explain the benefits of the legislative changes to stakeholders in easily understood pictorial forms (see an example on Fig. 1).

NLS began with a shortlist of key stakeholders who would be most affected by the proposed legislative amendments, to invite them for focus group discussions. The objective was to judge sentiment on the ground, to gain feedback on the legislative changes and to obtain buy-in, as well as to identify advocates for future media interviews when the legislation change was made public. Several sessions were conducted from November 2017 to February 2018 with publishers, writers, academics, heritage researchers and website owners. A dialogue session was subsequently extended to other members of the same group of stakeholders so as to widen NLS's reach and ensure that as many voices as possible could be heard and addressed. Both content providers (publishers, writers and website owners) and users (researchers and academics) were generally supportive of the proposed amendments and appreciated the custodian role that NLS plays in preserving documentary heritage. They were in favour of the web archiving initiative and suggested that NLS should archive websites more frequently and provide greater access to these materials. Stakeholders also encouraged NLS to consider archiving



Fig. 1 Infographics explaining the legislative changes to stakeholders and members of the public

non .sg websites that contain Singapore content. A public consultation exercise was conducted simultaneously to seek feedback from members of the public. The consultation paper detailing the proposed amendments was published through the Singapore government's feedback unit, REACH (www.reach.gov.sg), for a month at the end of 2017. At the end of the consultation period, feedback received included general support for the amendments and some technical queries pertaining to web archiving because the general public is not familiar with the web archiving process.

The key takeaway from the engagement with stakeholders and the public consultation exercise was that established good working relationships with publishers and researchers were advantageous, as both parties were willing to have open discussions and share their concerns. The challenge, however, was engaging stakeholders with whom NLS had less association, that is, website owners. It took multiple attempts to reach out to them before feedback was received, but after a connection was made, website owners lent their support and gave NLS the necessary encouragement to push forward with the update to the legislation.

4 System Enhancements to Support the Legislative Changes

Apart from legislative changes and engagement with stakeholders, NLS also looked at improving its systems and infrastructure so that it would be better able to process and support the web archive collection. Web Archive Singapore (WAS) (eresources.nlb.gov.sg/webarchives) hosts NLS's collection of archived Singapore websites.

As the number of archived websites increased exponentially, from a handful to about 180,000 registered .sg websites a year, effort was focused on rebuilding WAS. The increase in volume of data called for a review of the existing workflow, policies, budgetary requirements and manpower needs. WAS was launched in 2006 when NLS first embarked on collecting Singapore websites. The portal had two functions—keyword search and subject browsing. Websites are harvested with Heritrix and displayed via the OpenWayback.²

A project to revamp the WAS portal started in late 2017. Figure 2 presents a screenshot of the WAS portal. An environmental scan was done of other established web archiving portals, such as the British Library, the National Library of Australia and the Library of Congress, to baseline what functionality the redesigned WAS website could adopt and improve upon.

Five new functions were added to the new website—curation, full-text search, the public nomination of websites, a data visualiser and rights management. Curation takes centre stage on the homepage of the new website, with the showcasing of highlights from the collection. This helps to generate interest and encourage further exploration of the website. Figure 3 illustrates that there are two types of curation activities: Special Collections, which cover a significant event or topic relating to

²<https://github.com/iipc/openwayback>



Fig. 2 The revamped Web Archive Singapore portal, <https://eresources.nlb.gov.sg/webarchives/>, 2018

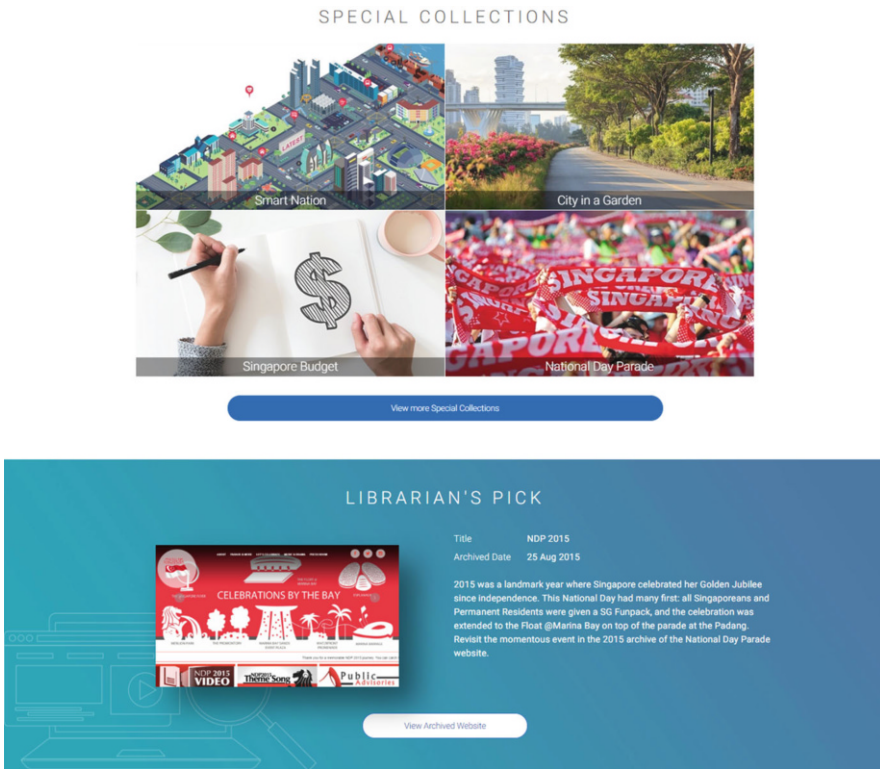


Fig. 3 Two types of curations are available—Special collections and Librarian’s Pick, 2018

Singapore (e.g. National Day Parade General Elections) and Librarian’s Pick, a monthly highlight of an archived website related to a topic or event of interest for that month (e.g. Singapore Arts Festival). The selection of the topics and curation of archived websites are done by librarians.

The search function, based on Solr,³ has been enhanced and facilitates full-text search right down to the individual webpages and files within an archived website. Facets are also included to enable users to refine their search results by format (e.g. websites, documents or images), year, subject category and URL.

To encourage public participation, an online nomination form was included so that the public could recommend Singapore websites from non .sg domains for NLS’s consideration. This will help librarians to build the web archives collection and also allow members of the public to contribute to NLS’s published heritage through their nominations.

Figure 4 presents another new feature of the portal is a data visualiser, which charts the frequency of a keyword or phrase written in archived documents over a period of time. This visual representation may be a useful resource for researchers who are interested in studying trends in the use and frequency of certain words. NLS is exploring other features that will be of use to researchers, in particular the provision of derivative data that highlights the unique voices of Singaporeans through the country’s websites.

Finally, the new WAS has also built in rights management to control access to the collection. Although NLS now has the mandate to collect .sg websites, these by default are only viewable within the library’s premises due to prevailing copyright

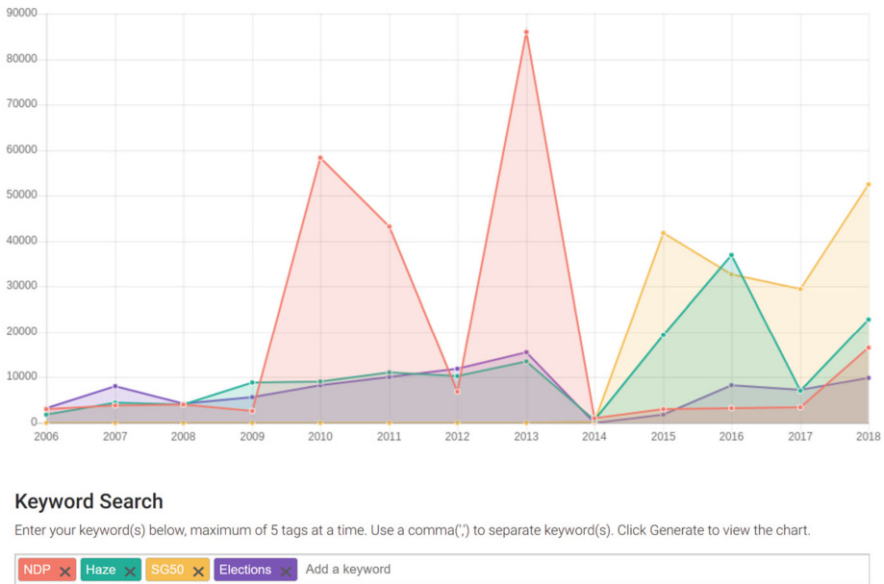


Fig. 4 Users can trace the frequency of use of a keyword or phrase over a period of time

³<https://lucene.apache.org/solr/>

laws. A simple lock icon will be displayed for archived websites that can only be viewed onsite.

5 Building a Comprehensive Collection of Singapore Websites

With the implementation of the new law, NLS has adopted a multi-prong approach to archiving the nation's online published works. NLS conducts both domain and selective archiving to ensure that the breadth and depth of online information is captured. This includes annual domain archiving, quarterly archiving of pre-identified subjects and monthly thematic archiving.

A domain crawl of .sg websites will be conducted at least once a year. This is done via a Memorandum of Understanding with the Singapore Network Information Centre, the national registry of .sg websites in Singapore, which allows NLS to obtain a yearly list of registered .sg websites. For selective archiving, NLS conducts quarterly archiving of websites from pre-identified subjects that contain substantial heritage value for Singapore, such as government websites, the arts, heritage, culture and politics. A quarterly frequency has been chosen as it strikes a good balance between capturing new content as well as ensuring that the web crawler does not affect the websites' performance.

NLS also conducts thematic archiving of websites that are event-driven. Event-based websites with a short lifespan will be archived more often to allow timely capture and minimise loss in case they are taken down or updated frequently. Examples of such topics are annual National Day celebrations, the Southeast Asian Games, new government policies and trending events that have substantial news and social media coverage.

While these three approaches help to build a good collection of .sg websites, NLS is also aware that there are many non .sg websites that contain significant Singapore content. As non .sg websites are excluded under the new legislation, NLS is still required to seek written consent from website owners before these can be archived. NLS will continually identify Singapore websites from the non .sg domain strategically. It has written to thousands of Singapore businesses and registered societies from non .sg domains for permission to archive their websites. This may be a tedious process and one which might not yield a high success rate, as NLS had experienced before the legislation was amended, but it will help the library to build a more comprehensive collection of Singapore websites. To help ease the permission-seeking process, NLS has also enhanced its public nomination form to allow website owners to grant permission directly to archive their websites. With enhanced marketing and promotion efforts, NLS hopes that this channel will encourage more Singaporeans and local companies to join it in preserving their institutional memories.

6 Public Education on Web Archiving and Legal Deposit

The passing of the NLB Act amendment Bill in July 2018 was a milestone event for NLS. Momentum arising from coverage in the media and public attention was harnessed for the purposes of public education and to strengthen the National Library’s position as the custodian for Singapore documentary heritage. This was a “once-in-a-lifetime” opportunity afforded only by the timing of the legislative amendments. NLS seized the opportunity and organised an outdoor event at the National Library Building, as well as roving exhibitions with supporting programmes at NLB’s three regional public libraries, to launch the implementation of digital legal deposit and domain web archiving, which are the primary components of the amendments to the legislation. The event was positioned to achieve the objective of creating greater public awareness of and appreciation for Singapore’s published heritage and web archiving, as well as to encourage the public to contribute towards this collection.

A 2-day event in November 2018 was held at the National Library Building that included a showcase of legal deposit and web archive content, public talks and performances, Instagram-worthy backdrops and photo booths to attract young people and families (see Fig. 5). The event was attended by about 1800 people, who came to enjoy the activities and soak in the carnival-like atmosphere. A public talk by Mr. Brown, Singapore’s most famous blogger, that highlighted the role of web archiving to preserve online materials was standing-room only in a venue that can hold 300 people.



Fig. 5 Elderly guests from nearby communities having fun at the “Instagram-worthy” photo booth

7 Conclusions

NLS's journey to extend the remit of the legal deposit scheme to the digital realm required perseverance to push through the proposed legislation. Through this process of amending the legislation, NLS learnt many valuable lessons, such as the importance of investing time and effort to cultivate relationships and goodwill with content creators, publishers and website owners so that they can be persuaded to support web archiving initiatives. It was also essential to keep up with the developments of the various laws governing copyright and privacy worldwide as these could potentially impact how the legislation was drafted. Internally, NLS had to make a convincing case to persuade the relevant ministries to support the long-term activities of archiving and preserving digital works in annual budgets.

The journey does not end now that the legislation has been passed. NLS will need to continually fine-tune its work process and systems, train its personnel and develop their skillset. Externally, it will be an ongoing uphill task for NLS to create awareness of web archiving, in order to maintain its relationship with content creators and publishers. It will be a long road ahead, but NLS takes heart that at least it has started on the right track and is positive that its work will create a lasting legacy for future generations of Singaporeans and researchers.