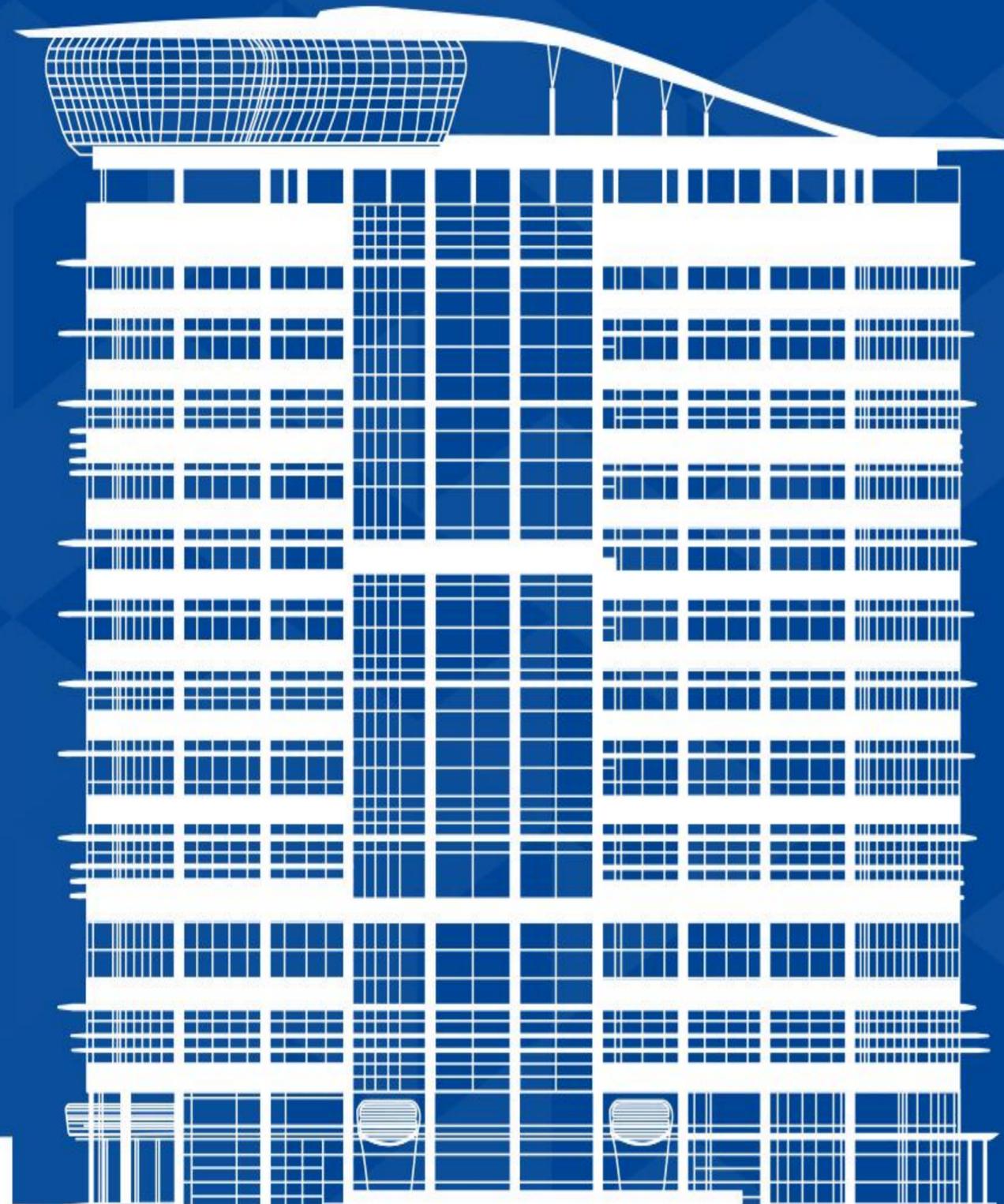


WEB ARCHIVING SINGAPORE

Ivy Lee (Ivy_LEE@nlb.gov.sg)

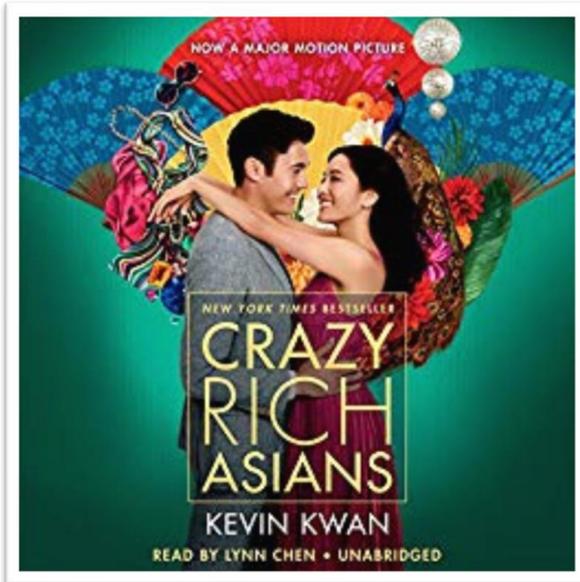
National Library, Singapore

IIPC Zagreb, 6 & 7 June 2019



NLB | National Library Board
Singapore

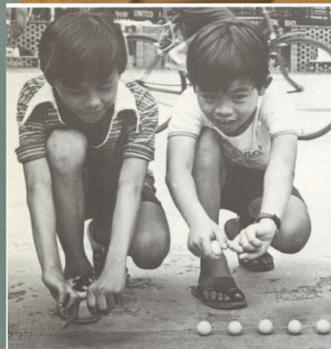
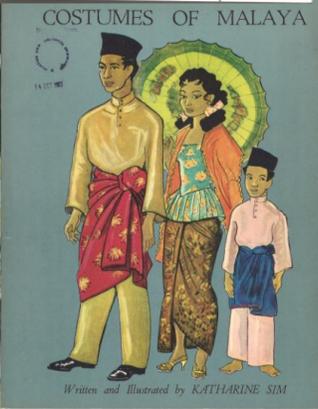
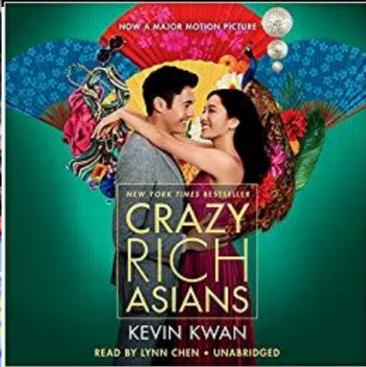
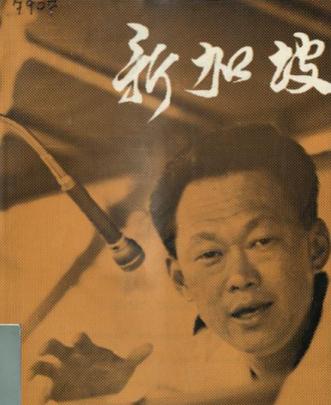
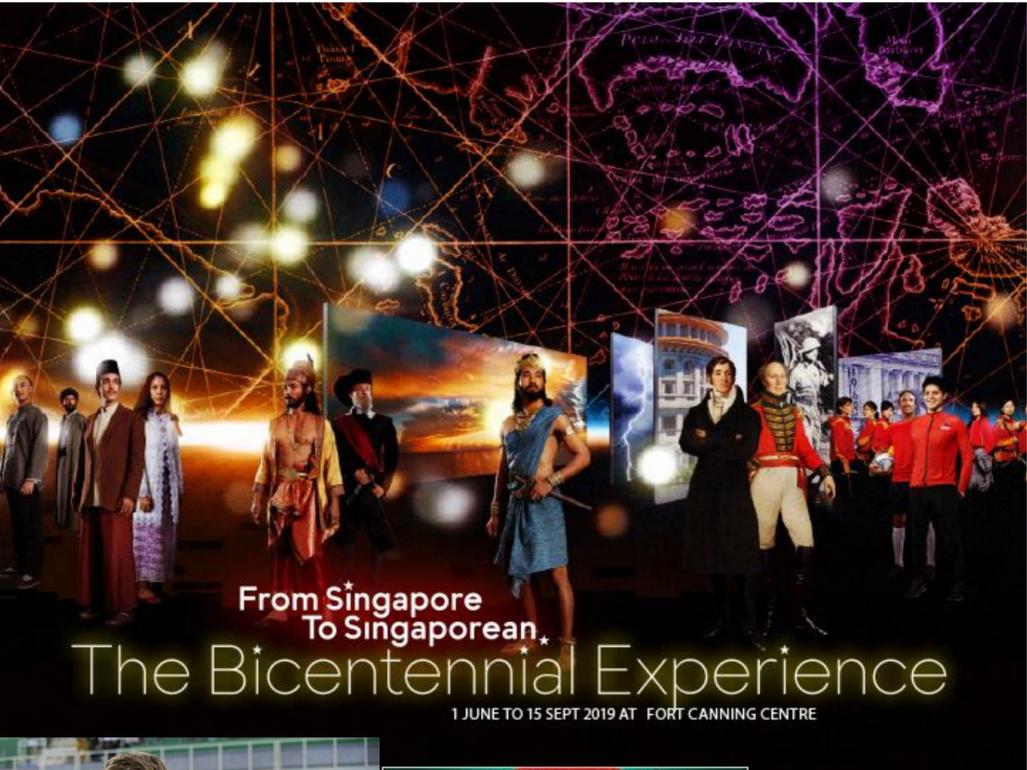
WHERE IS SINGAPORE



The **Little Red Dot** known as **Singapore**

Singapore, with a land area of only 718 km², is one of the smallest countries in the world. As a result, Singapore is often depicted on world maps as just a **little red dot**.

SMALL, BUT WITH A UNIQUE HERITAGE WORTH PRESERVING



NATIONAL LIBRARY SINGAPORE



- A knowledge institution under the National Library Board that collects, preserves and makes accessible Singapore's documentary history and heritage.

WEB ARCHIVING IN SINGAPORE

- Started selective web archiving of Singapore websites in 2006
- Challenges with consent-based approach
 - Written consent of website owners is required
 - Difficulty in identifying website owners
 - Low response rate from website owners
- This approach is inadequate for an institution like a National Library mandated to build a comprehensive collection of Singapore published materials.



AMENDMENTS TO THE NATIONAL LIBRARY BOARD ACT (EFFECTIVE 31 JANUARY 2019)



Legal Deposit

Preserving electronic publications

Extending the **scope** of legal deposit to include **electronic materials**.



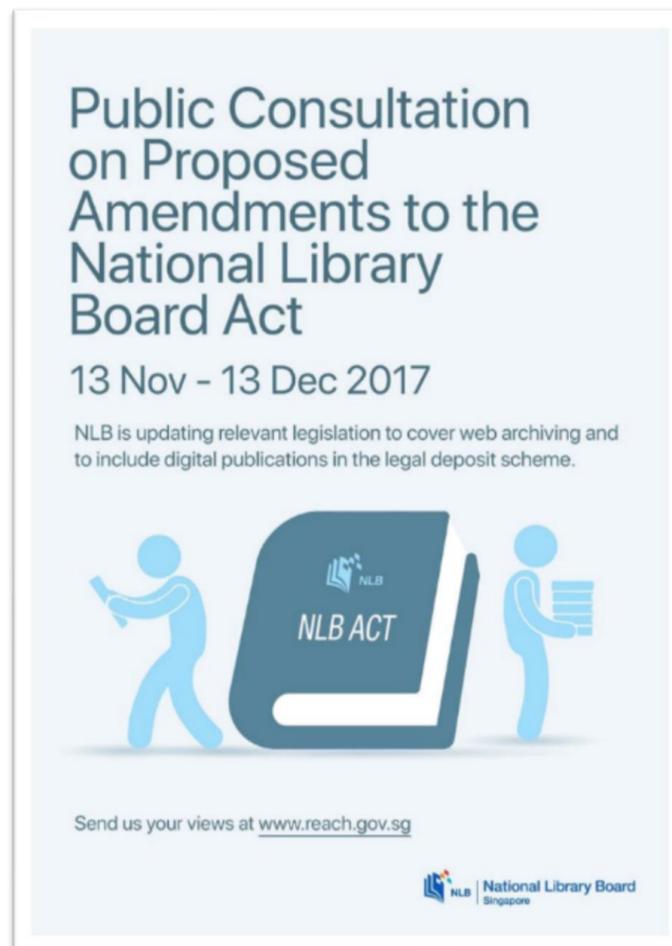
Web Archiving

Preserving online publications

Empowering NLB to **collect, make accessible and preserve websites from the Singapore .sg web domain** without the need for written consent.

STAKEHOLDERS' ENGAGEMENT

Focus Group Discussions (FGDs) with key stakeholders to consult representatives from academia, heritage researchers, writers, website owners, social media influencers, bloggers and librarians.



Public Consultation on REACH (Singapore government's feedback unit) to provide an online platform for the public to comment on the proposed amendments.

PUBLIC ENGAGEMENT

Op-ed article on preserving digital and online published heritage by Prof Ang Peng Hwa, Nanyang Technological University and Advisor to Singapore Internet Research Centre was published in the Straits Times on 15 Dec 2017.



Online archives matter – it's only right to bring them into NLB fold

Ang Peng Hwa
For The Straits Times

We can expect that Singapore's Smart Nation thrust will have a major impact on our lives, economic and social. Just how extensive that impact might be is well-illustrated by a Bill to amend the National Library Board Act. The draft Bill is aimed at capturing and preserving for posterity our online world.

Historical artifacts have tended to be ignored, if not discarded, in our headlong rush for progress. The proposed Bill to catch up. The Act was last amended in 1995, a year after the internet was made available to the public. Many other countries – Britain and France in Europe, Australia, New Zealand, South Korea and Japan in the Asia Pacific – have recognized the importance of the online world not just as a repository of information, but also of culture, and they have passed bills that empower libraries to archive online material.

The proposed Bill will require electronic versions of digital materials such as e-books to be deposited with the National Library Board (NLB), empower the board to compulsorily archive websites registered under the .sg domain name, and afford restricted access to these materials.

Under the current regime, the board is required to seek the approval of the site before archiving. The consent of site owners is that such material should not be made available in such a way as to breach copyright. A consequential amendment to the Copyright Act is the reform needed.

The protocol globally for online archives is that such materials will be made available through strict access control within the National Library's premises. Researchers viewing the archived websites on computer terminals will not be able to download, copy and print the archived content. Researchers can contact the content owners, who are usually the website creators or organizations, for permission to use the content.

The Bill will enable online material to be comprehensively collected and preserved. Without it, active digital material will be lost.

Archives serve the important function of keeping documents for future use. The precise use cannot be predicted.

For example, how important is it that one keeps the records for the running of a lightshow? In the first Budget drawn up by the Singapore Government for his project comparing our national budgets over the years.

Just how important an archived document may be cannot be easily predicted. When Ukrainian separatists brought down Malaysia Airlines Flight MH17, the post with links to video footage of the wreckage was online for just two hours. It happened that the site had been marked as one to watch and so the post had been archived.

While some of us may regret having a silly photo or remark of us by us online, the reality is that webpages have a short lifespan, averaging just 100 days. For important events such as the National Day Parade, the archived websites for the various years would show how the event has evolved over the years. Our identity is tied up with our memories. And our archives inform us of our identity.

The current consultation for the Bill is necessary to balance competing interests from publishers and content creators. The board had engaged these stakeholders earlier through focus group sessions even before the public consultation was conducted since mid-November by government feedback agency Reach.

The comments received touch on the security of public access to deposited digital publications, clarification on the scope of archiving, eg websites, and on copyright intellectual property rights to deposited content.

The Bill is a welcome addition to our laws. It brings digital archives within the purview of national depositories. The consultation period provides an opportunity for stakeholders to give their views.

In this regard, as a researcher, I would like to suggest that Singapore pay forth more forward-looking policies and break away from the current norm of having very limited access. For example, online archives typically allow extremely limited access – only at desktop terminals and no downloading or copying. The comments on over the dissemination of digital copies.

Perhaps indices and abstracts of the archives can be made available online to encourage greater use of the resource. This would be a small but significant step towards giving citizens and researchers greater access to our archives, and yet another step towards being a Smarter Nation.

Ang Peng Hwa is a professor at the Nanyang Technological University, where he teaches and researches internet law and policy.

Drives in a Merlion light display against the Marina Bay skyline at the National Day Parade held at The Float @ Marina Bay in August. Online archives for the National Day Parade would help Singaporeans to remember the event, and our identity, according to the writer. It is tied up with our memories. THE S.I.T. PHOTO

Singapore-Malaysia territorial dispute over Pedra Branca, these other two questions and mandate research proved critical to supporting Singapore's case as they showed how we had been operating.

In research, an economist friend raved over being able to obtain the

arranging just 100 days. For important events such as the National Day Parade, the archived websites for the various years would show how the event has evolved over the years. Our identity is tied up with our memories. And our archives inform us of our identity.

The current consultation for the

Media Interview for Minister, Ministry of Communications and Information at the Legal Deposit exhibition.

CONCERNS OF WEBSITE OWNERS AND CREATORS



- Lack of understanding about web archiving
- Privacy, personal data; availability of takedown measures
- Comprehensiveness of web archiving; who gets to be archived and content hosted on social media platforms
- Frequency of web archiving; disruptions to websites
- Vulnerability (government websites)
- Need to balance between website owners' concerns and researchers' needs

Limit access to digital legal deposit materials and archived websites online and provide full access at the library premises only

WEB ARCHIVE SINGAPORE (WAS) PORTAL

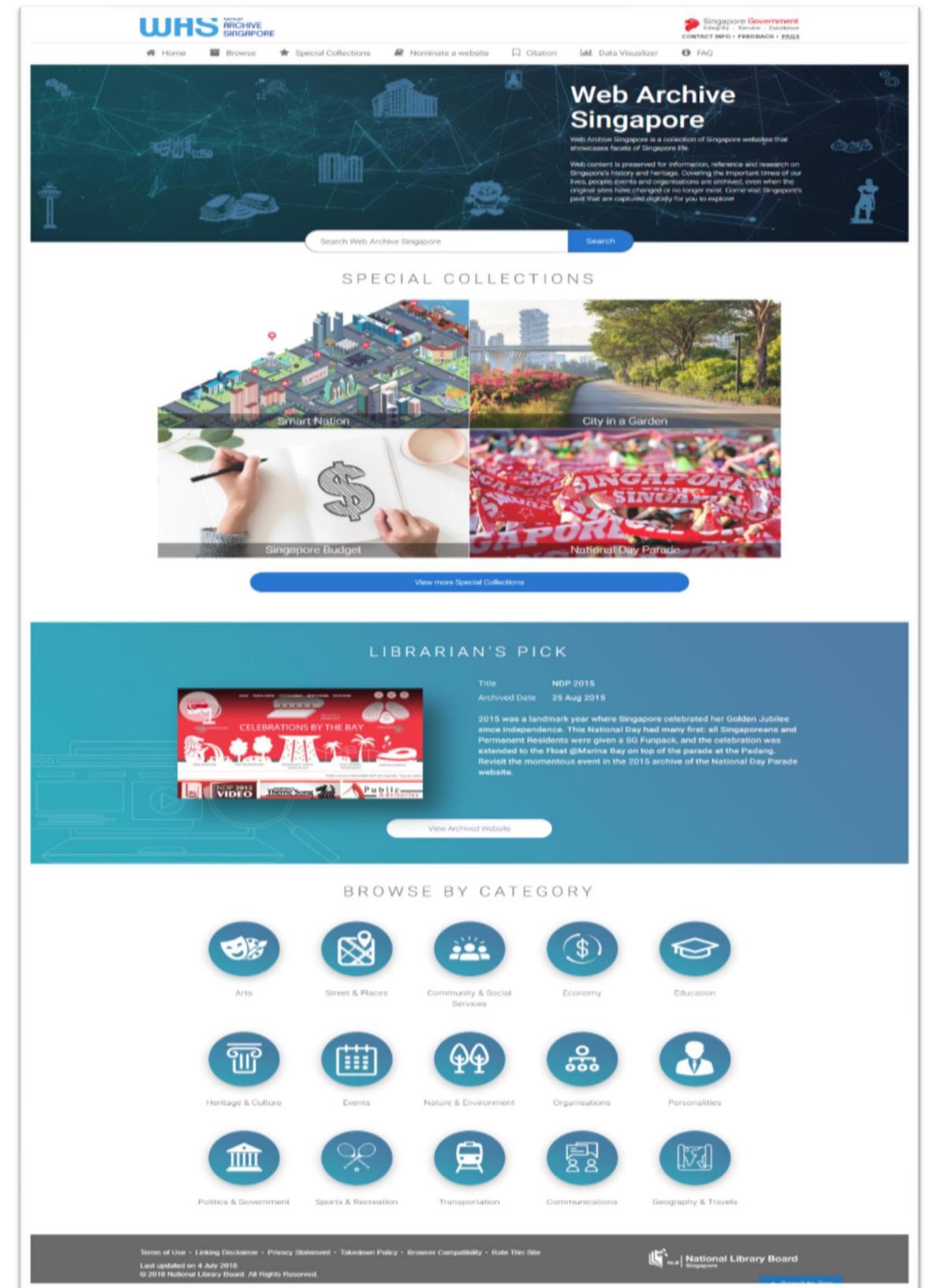
- First edition of the Web Archive Singapore started with a small dataset of selected websites in 2006
- Until 2019, there were only 2,000 websites on WAS
- Only had two functions – keyword search and browse by subject



WEB ARCHIVE SINGAPORE REVAMP IN 2018

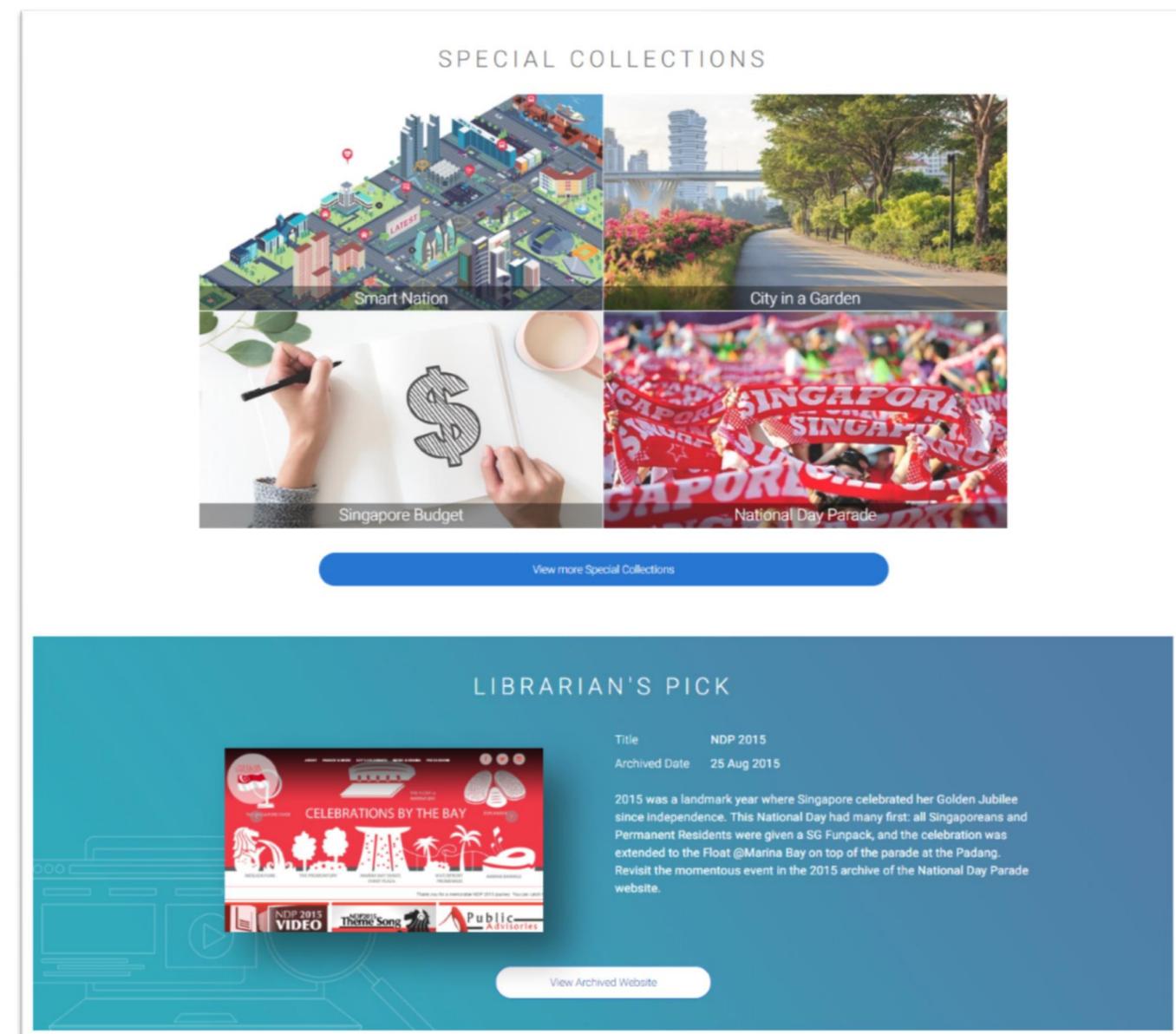
New features in the revamped Web Archive Singapore

1. Curation
2. Full text search
3. Self and Public nomination of websites to be archived
4. Data visualiser
5. Rights management



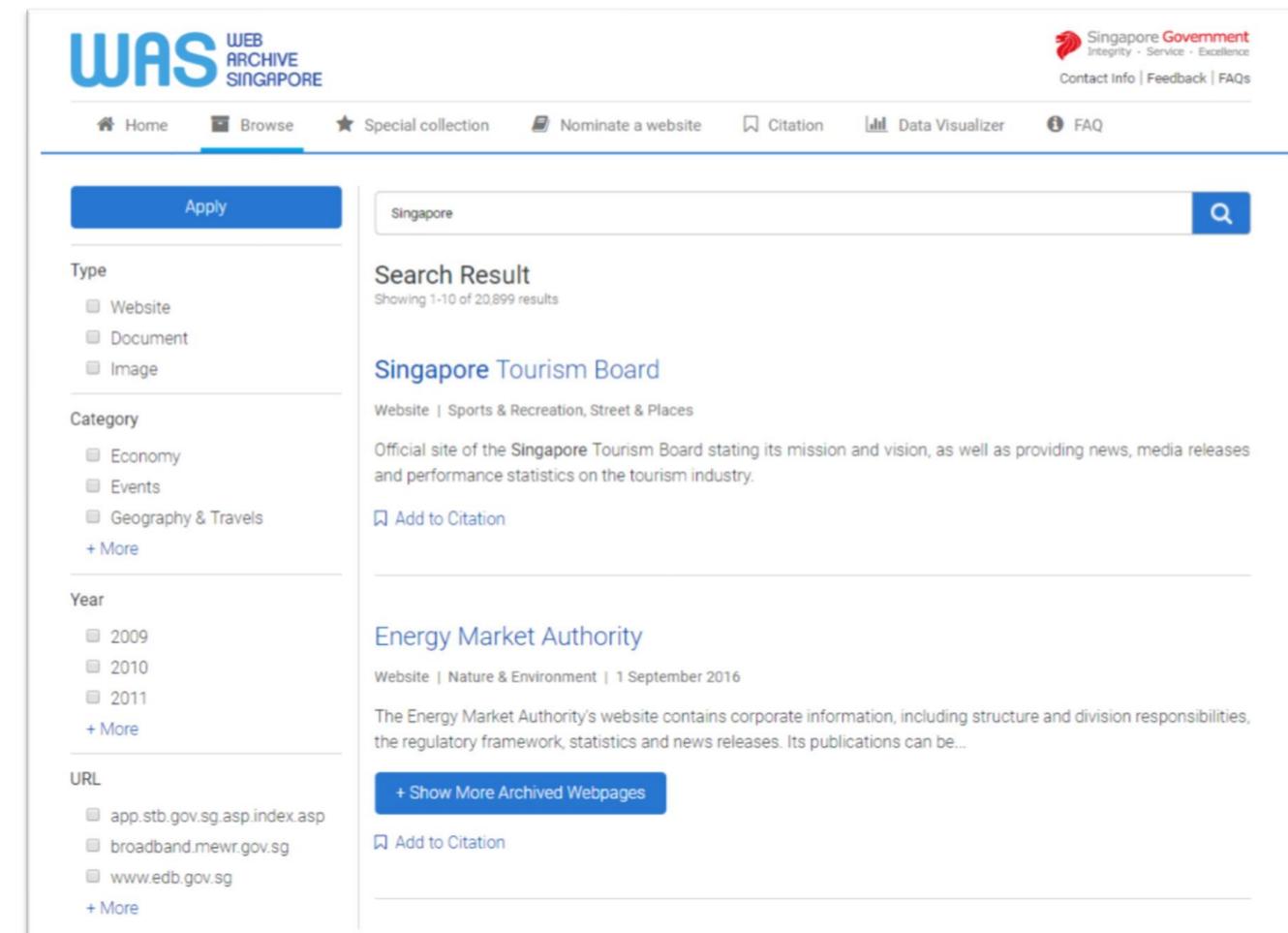
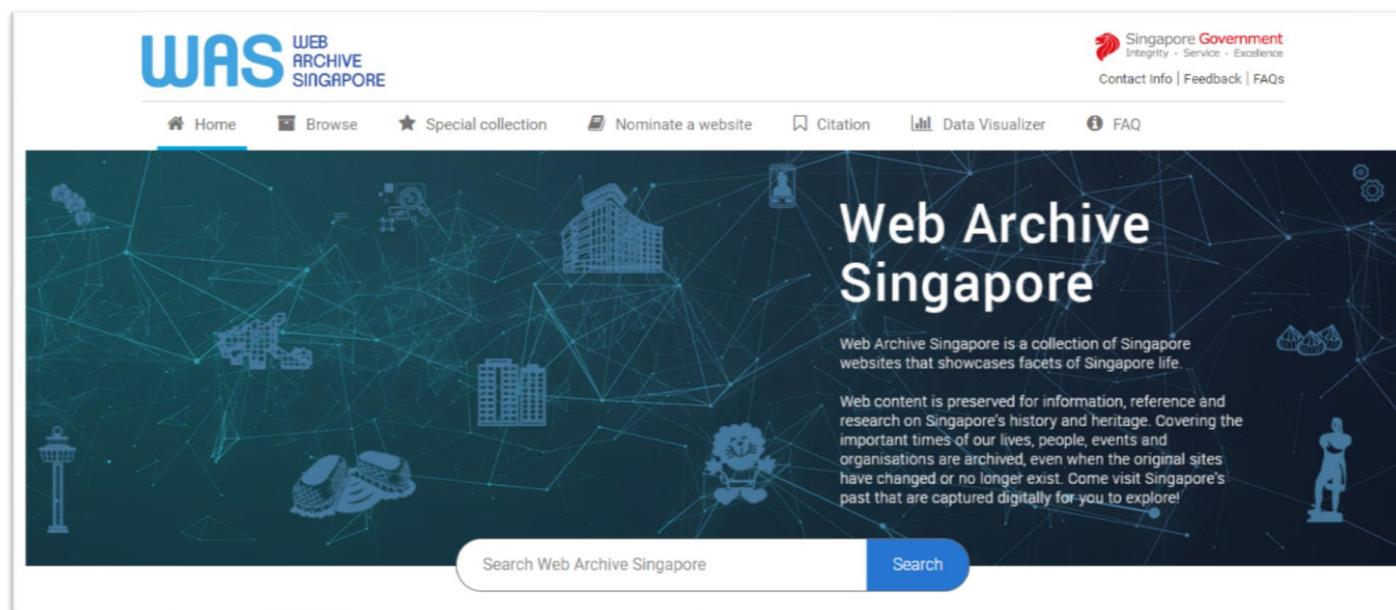
ENHANCED FEATURES IN WEB ARCHIVE SINGAPORE REVAMP – CURATION

- Showcase highlights from the collection
- Two types of curation:
 - Special Collections: cover a significant event or topic relating to Singapore
 - Librarian's Pick: a monthly highlight of an archived website related to a topic or event of interest for that month



ENHANCED FEATURES IN WEB ARCHIVE SINGAPORE REVAMP – FULL-TEXT SEARCH

- Implemented Solr for full-text search
- Facets to fine tune their search results by type (i.e. websites, documents, or images), year, subject categories, and URL



Prominent search bar with full text search capability



ENHANCED FEATURES IN WEB ARCHIVE SINGAPORE REVAMP – PUBLIC NOMINATION

- The public can nominate websites that do not belong to the .sg domain for archiving, as long as there is relevant Singapore content
- This gives the public a chance to contribute to our archiving work

NOMINATE A WEBSITE

Have you come across a Singapore website that is not under a .sg domain? Recommend the website for our consideration for preservation in the Web Archive Singapore.

We are interested in websites with:

- Substantial content about Singapore
- Written or owned by a Singaporean or Singapore organisation

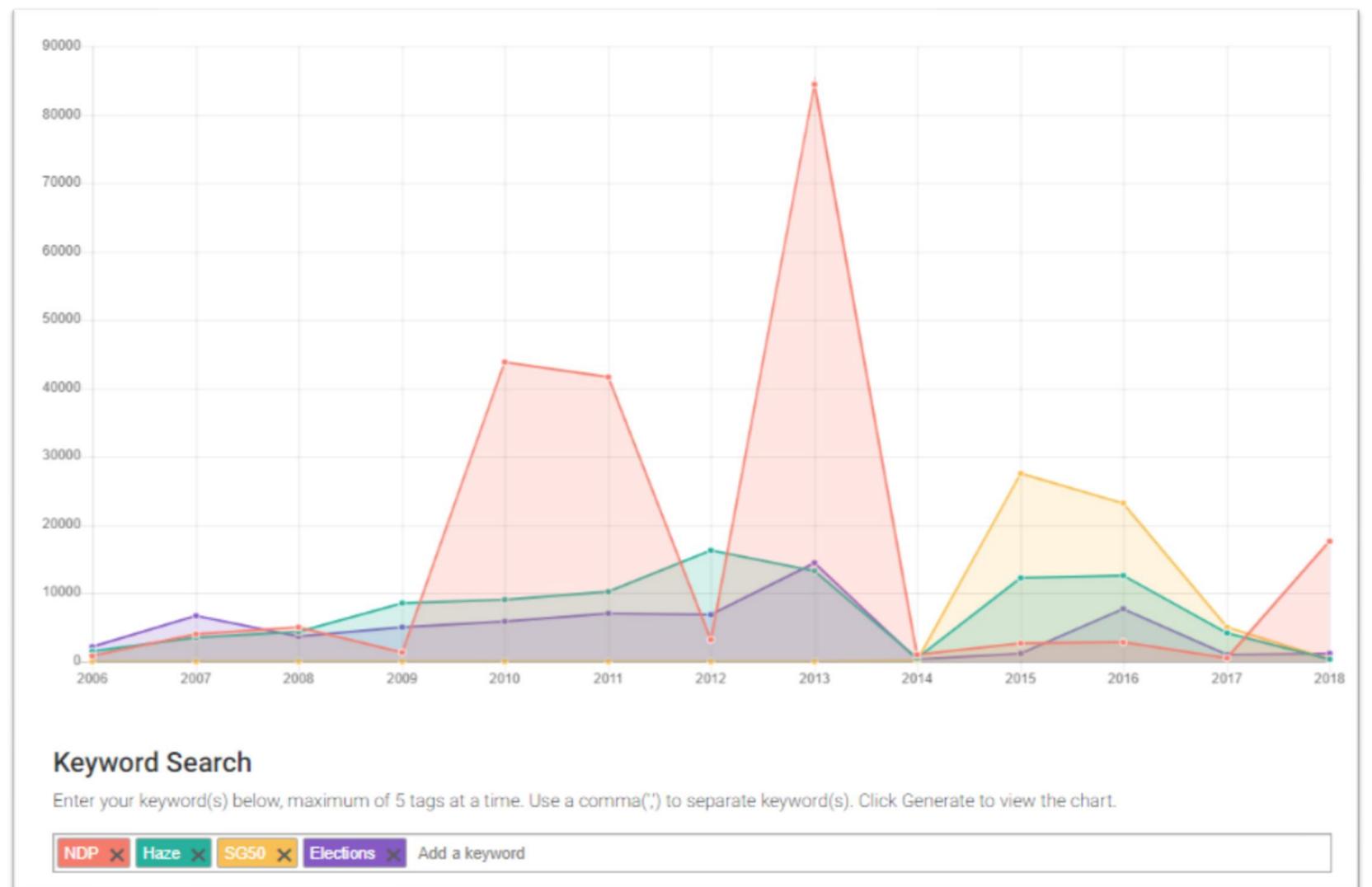
The nominations will be assessed based on the National Library, Singapore's selection guidelines as well as web archiving priorities.

Once the nominations have been accepted, we will contact the owner or publisher of the websites for their consent before archiving.

Name	Email
<input type="text" value="Name of nominator"/>	<input type="text" value="Email address of nominator"/>
Website URL	Are you the owner of the website
<input type="text" value="URL of website to archive"/>	<input type="text" value="Please Select"/>
Reason(s) for recommendation	
<input type="text"/>	
Maximum 1000 characters (0 character)	

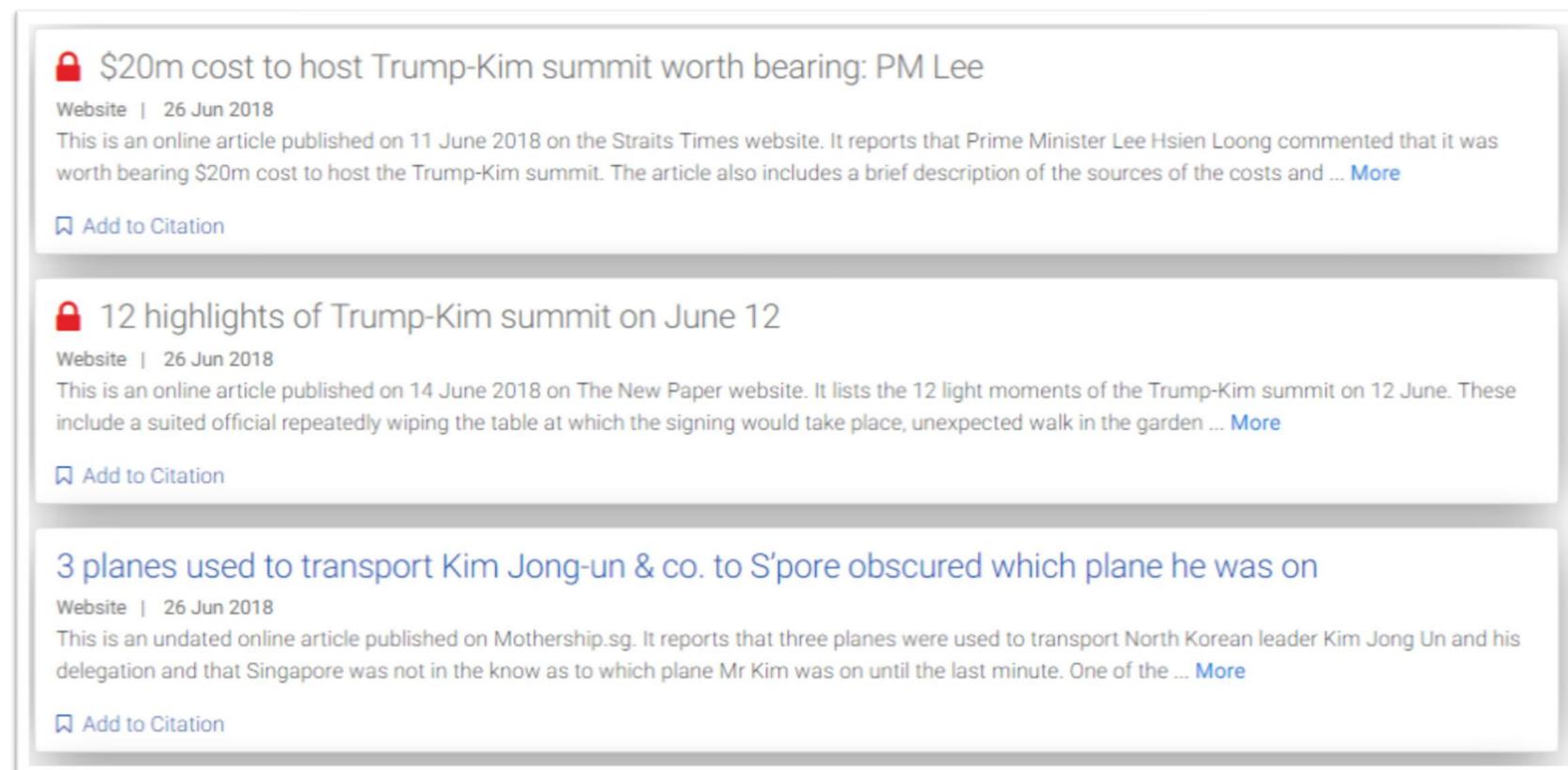
ENHANCED FEATURES IN WEB ARCHIVE SINGAPORE REVAMP – DATA VISUALIZER

- Visualizer that charts the frequency of a keyword or phrase over a period of time
- This visual representation may be a useful resource for researchers who are interested in studying the trend of certain words



ENHANCED FEATURES IN WEB ARCHIVE SINGAPORE REVAMP – RIGHTS MANAGEMENT

- .sg websites by default can only be viewable within the library's premise
- A simple lock icon will be displayed to alert users that viewing is only available at the library



The screenshot displays three article entries from the Web Archive Singapore. Each entry is enclosed in a light gray box and features a red lock icon on the left, signifying that the content is restricted to library premises. The entries are as follows:

- Entry 1:** Title: "\$20m cost to host Trump-Kim summit worth bearing: PM Lee". Website: "26 Jun 2018". Description: "This is an online article published on 11 June 2018 on the Straits Times website. It reports that Prime Minister Lee Hsien Loong commented that it was worth bearing \$20m cost to host the Trump-Kim summit. The article also includes a brief description of the sources of the costs and ... [More](#)". Action: "Add to Citation".
- Entry 2:** Title: "12 highlights of Trump-Kim summit on June 12". Website: "26 Jun 2018". Description: "This is an online article published on 14 June 2018 on The New Paper website. It lists the 12 light moments of the Trump-Kim summit on 12 June. These include a suited official repeatedly wiping the table at which the signing would take place, unexpected walk in the garden ... [More](#)". Action: "Add to Citation".
- Entry 3:** Title: "3 planes used to transport Kim Jong-un & co. to S'pore obscured which plane he was on". Website: "26 Jun 2018". Description: "This is an undated online article published on Mothership.sg. It reports that three planes were used to transport North Korean leader Kim Jong Un and his delegation and that Singapore was not in the know as to which plane Mr Kim was on until the last minute. One of the ... [More](#)". Action: "Add to Citation".

ANNUAL DOMAIN CRAWL

- Obtain a yearly list of registered .sg websites from the agency responsible for registering Singapore websites, currently standing at more than 180,000 websites.
- Annual archiving includes:
 - .sg websites
 - non .sg websites that have granted NL the permission to archive their website
- First domain archiving since the commencement of the legislation was completed in March 2019.

COLLECTING SINGAPORE WEBSITES OUTSIDE OF THE .SG DOMAIN

- Strategic approach that will cover key aspects eventually - start with Singapore businesses and registered societies. To write to them for permission to archive their websites and make available online.
- Enhancing the public nomination form - make it easy for website owners to grant permission directly to National Library Singapore to archive their websites



LOOKING TO THE NEAR FUTURE - COLLECTING SOCIAL MEDIA

- Internet activities are increasingly moving to social media platforms
- Almost all government agencies and political holders in Singapore use at least one social media platform



TESTS DONE TO COLLECT SOCIAL MEDIA

- Conducted initial testing with Twitter API
- Further exploration needed on how to metatag, preserve and make these content available, and handling of legal aspects such as third-party comments, website links, images and videos to third-party content
- Facebook is the more well-used social media platform in Singapore and it does not provide any free open API

Social Media Users in Singapore

25%
Twitter 

70%
Facebook 

Statistics from <https://wearesocial.com/sg/digital-2019-singapore>

WHAT'S NEXT

- Mitigate the legal limitations of providing onsite access only by getting permission to make more websites available in the public domain
- Explore alternative methods to collect social media content such as social media archiving companies; explore alternative collecting business models such as via donation
- Develop greater understanding of how researchers use web archives to allow for better info-packaging to meet their information needs
- Develop better integration of web archives into other resources at the National Library to present a complete suite of resources on Singapore to researchers
- Integrate Singapore content published overseas into the National Library's collection
- Develop more in-house digital capabilities in web archiving

Thank you



Ivy_LEE@nlb.gov.sg