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
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.

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
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
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
COLLECTING SINGAPORE WEBSITES

Adopts a combination of selective and whole domain approaches:

1. Whole domain archiving (annual)

2. Archiving of pre-identified subjects (quarterly)

3. Thematic archiving (monthly/weekly or more often if required)

Adapted from https://www.sg/sg50/collection.aspx
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

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