

Collaboration and aggregation for innovative and rich content services - the NLB experience

Peter PAK

Technology & Innovation, National Library Board, Singapore
peter_pak@nlb.gov.sg

Siang Hock KIA

Technology & Innovation, National Library Board, Singapore
kia_siang_hock@nlb.gov.sg



Copyright © 2015 by Peter Pak and Siang Hock Kia. This work is made available under the terms of the Creative Commons Attribution 3.0 Unported License: <http://creativecommons.org/licenses/by/3.0/>

Abstract:

Libraries and archives collect and provide access to resources of significant national, cultural and heritage values. While these efforts continue unabated, we cannot undertake this enormous task alone. We must collaborate, and over the years, libraries and archives have established extensive and close co-operative relationships.

The National Library Board of Singapore (NLB) manages the National Library, the National Archives and the Public Libraries in Singapore. When the National Archives of Singapore (NAS) joined the NLB family in November 2012, it brought with it a huge and highly valuable collection of primary and unique materials on Singapore's history. OneSearch, the unified search service across the National Library, NAS and the public libraries launched in August 2014, clearly demonstrated the opportunities and benefits of the close collaborations amongst libraries and archives.

Libraries and archives, with their patrimonial roles, are in the unique position to collaborate with content owners on copyrighted content. The NewspaperSG and the MusicSG services are two examples of such powerful collaborations with the newspaper publishers and composer and lyricist association in Singapore that bring otherwise unavailable copyrighted content to the public. The eReads service aggregates all e-book titles from the various commercial providers NLB subscribed to and provides a one-stop portal to discover and find e-books of various genres using an intuitive and attractive user interface. At the national level, the Singapore Memory project brings the entire nation together to build a national memory bank.

A recent research by NLB of national libraries in 14 countries identified 3 key digital library trends: digitisation of collection and offering digital library services; digital preservation; and collaboration.

Strategic collaborations (at local, regional and international levels) and aggregations, coupled with the innovative use of information technologies, enable NLB to better preserve its national collections while making valuable content accessible.

Keywords: National Library Board of Singapore (NLB), Collaboration, Aggregation, Rich Content Services

1. INTRODUCTION

The National Library Board of Singapore (NLB) manages the National Library, the National Archives and the Public Libraries in Singapore. When the National Archives of Singapore (NAS) joined the NLB family in November 2012, it brought with it a huge and highly valuable collection of primary and unique materials on Singapore's history. This important milestone provided the opportunity for the libraries and archives to work together to collaborate and jointly provide access to resources of significant national, cultural and heritage value to the residents of Singapore at a more holistic level.

As a result, new services were conceptualised and developed to leverage on this important milestone. OneSearch, a unified search service across the National Library, NAS and the public libraries, was launched in August 2014. Collaborations also extend beyond NLB, with partnerships forged with the private sector and copyright owners to make more accessible copyrighted content to the public (e.g., NewspaperSG and MusicSG).

2. THE NLB DIGITAL COLLECTIONS

2015 marks fifty years of independence for Singapore. A nationwide effort called SG50¹ was conceived to celebrate this major milestone with activities and programmes that reflect on how far Singapore has come together as a nation and people.

Over the years, NLB has built up a huge collection of primary and unique materials on Singapore's history (Figure 1). Upping the digitisation efforts to preserve as well as to give its content wider access through various online platforms put NLB in good standing to showcase its collections for SG50 and beyond. In summary, the digitised content from NLB constitutes a valuable repository of the memories of a nation.

¹ SG50 –About SG50 <https://www.singapore50.sg/SG50/About>

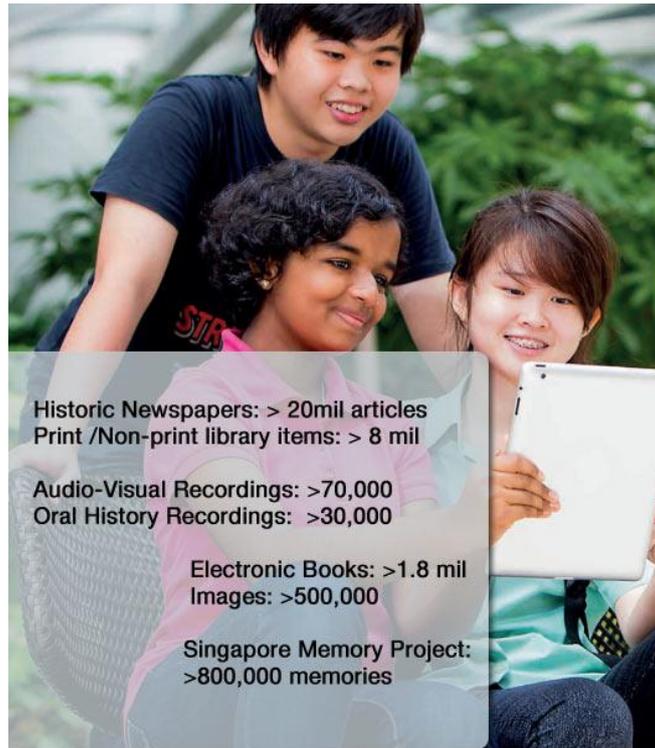


Figure 1: NLB digital collections

3. CONTENT COLLABORATION AND AGGREGATION IN ACTION

However, NLB cannot go about digitising and archiving content solely on its own. In the digital age where enormous amount of content is “created” every minute, decisions to preserve content for future generations are very much dependant on the budget available. Skillsets for different projects and collections also vary and additional manpower may not necessary be available to address the skill gaps, especially gaps in domain knowledge.

Co-creation, content collaboration and public donation become increasingly important in the digital explosion age. Factors in the technology enabler to deliver content in a robust and cost effective manner are important considerations. An agile suite of re-usable technology stack is needed to deal with the additional complexity in collaborations, particularly the diversity of content, formats, standards and access rights management needs. Robustness and the ability to scale are also critical to handle the expanded repositories made possible through collaborations.

Table 1 highlights some of the NLB services that encapsulate the ethos of the NLB collaboration and aggregation strategy.

NLB Service	Brief description	Collaboration and/or Aggregation
OneSearch	Unified search service across libraries, archives and museums. (http://search.nlb.gov.sg)	National Library, Public Libraries, National Archives and National Heritage Board museums.
NewspaperSG	Digitised newspapers with over 30 newspapers in English, Malay and Chinese from 1839 to 2010. (http://eresources.nlb.gov.sg/newspapers/)	National Library, Singapore Press Holdings, MediaCorp.
MusicSG	A non-profit digital archive set-up to digitise, archive and provide access to all forms of published musical works. (http://eresources.nlb.gov.sg/music)	National Library, Composers and Authors Society of Singapore (COMPASS), Artists.
eReads	An online portal that aggregates the subscribed eBooks, eMagazines and eNewspapers in an attractive manner. (http://eresources.nlb.gov.sg/ereads)	Various Vendors (including Overdrive, Library Press Display, Gale, EbscoHost, etc.).
Singapore Memory Portal	An online portal to support the Singapore Memory Project, a national initiative to collect, preserve and provide access to Singapore's collective memory. (http://www.singaporememory.sg/)	Individuals and Institutions.
Citizen Archivist	An online portal where members of the public help to make archives records more discoverable by describing images and transcribing hand-written documents. (http://www.nas.gov.sg/citizenarchivist)	Individuals.
ASEAN Digital Library	The ASEAN Digital Library (ADL) is a platform that provides users with a seamless way to search for and to gain access to the wealth of digital materials held by the National Libraries in ASEAN.	ASEAN National Libraries.
Biodiversity Heritage Library	The Biodiversity Heritage Library (BHL) is a consortium of libraries that cooperate to digitise literature of biodiversity and make that literature available for open access. (http://www.biodiversitylibrary.org)	Libraries.

Table 1 Collaboration and Aggregation

3.1 Unified Search Service

OneSearch is NLB's search service that provides access to the rich content available across the National Library, Public Libraries, National Archives of Singapore and museums under

the National Heritage Board² of Singapore with extensive search coverage beyond books, magazines and audio-visual materials from the library's catalogue. Launched in August 2014, the search now spans across a wide range of databases and online content including digitised newspapers, audio-visual recordings, local music, literary works, manuscripts, speeches, biographies, museum artefacts, among others (Figure 2).

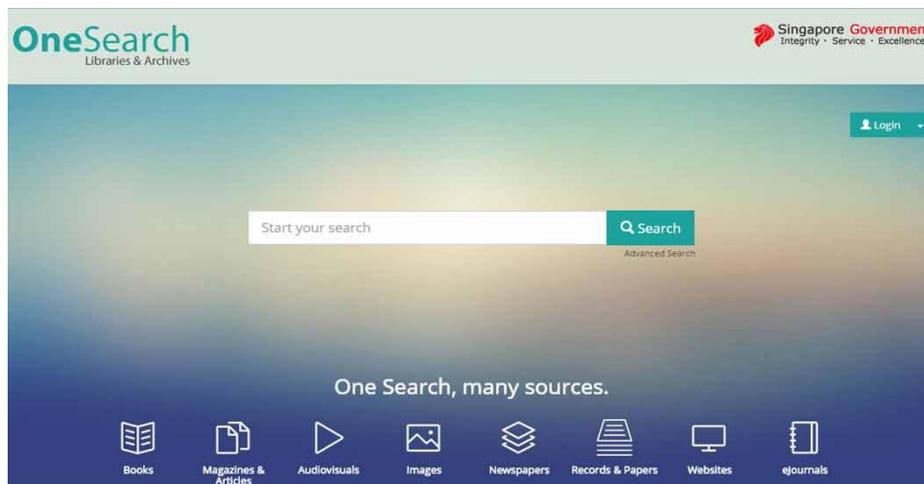


Figure 2: OneSearch unified search service

3.2 Public-Private Sector Content Collaboration

NewspaperSG and MusicSG are two examples of portals whose rich content was made possible with the close collaboration with the private sector institutions as well as individuals.

3.2.1 NewspaperSG

In the case of NewspaperSG, NLB worked closely with the newspaper publishers (including the Singapore Press Holdings³) in Singapore to expand the accessibility and availability of historic newspapers through large-scale digitisation at the article level. The portal enables search and access to over 30 newspapers published between 1831 and 2009 (Figure 3). Most content on NewspaperSG are accessible online, with those issues published from 1990 onwards available onsite within NLB premises.

With over 20 million newspaper articles, NewspaperSG is the most popular content service in NLB. It attracts over 1 million page views a month.

² National Heritage Board - <http://www.nhb.gov.sg/>

³ Singapore Press Holdings - <http://www.sph.com.sg/home/>



Figure 3: NewspaperSG portal

3.2.2 MusicSG

MusicSG is a yet another great example of a rich content service where NLB worked closely with the Composers and Authors Society of Singapore⁴ (COMPASS) and individual artists to bring alive Singapore's music scene. It is a portal dedicated to music composed or published by Singaporeans, music produced or published in Singapore, and music related to Singapore (Figure 4). Similar to NewspaperSG, not all tracks are available via the Internet. Due to copyright reasons, some of the tracks are only available for access within library premises. Beyond the tracks, NLB digitised and made available album inserts, scores, lyrics and research articles.

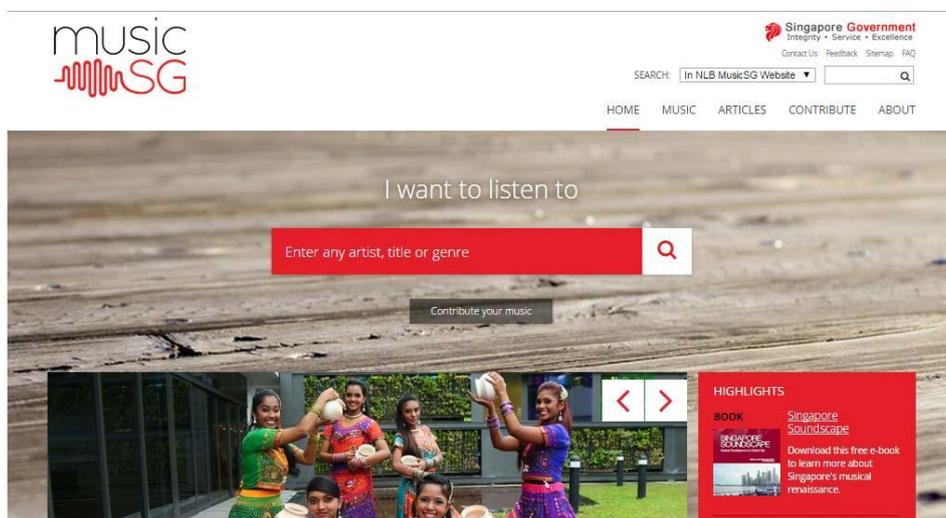


Figure 4: MusicSG portal

For MusicSG, giving the assurance to budding or indie artists (independent artists) of a website that will display their works yet protect them from distribution also provided an impetus for this category of artists to donate their works.

⁴ Composers and Authors Society of Singapore is an organisation created to protect and promote the copyright interests of composers, authors* (and their heir) and publishers of musical works and their related lyrics (<http://www.compass.org.sg/>)

In both instances, without the explicit permission granted by private sector partners, the content of both portals will not be as rich as it is right now. Although full access may not be available for now due copyright restrictions, the content is archived for perpetuity, ensuring that the valued content will be available for generations to come.

3.3 Nationwide Collaborations

Two examples of nationwide initiatives to get the general public to contribute are the Singapore Memory Project and Citizen Archivist Project.

3.3.1 Singapore Memory Project

The Singapore Memory Project (SMP) actively engages the general public, institutions and community groups to collect, preserve and provide access to Singapore's collective memory. Since its launch in 2011, it has garnered over 845,000 memories (Figure 5). The SMP team continues to engage the citizens of Singapore to contribute through various online and on-the-ground programmes. It has worked with grassroots leaders to gather community memories of townships (including Queenstown and Toa Payoh) and commissioned short films to document different facets of Singapore that are or have undergone drastic changes due to the rapid development in Singapore.



Figure 5: Singapore Memory Project Portal

3.3.2 Citizen Archivist Project

The Citizen Archivist Project Portal was recently launched by the NAS in March 2015 (Figure 6). It provides a platform for volunteers to help describe images and transcribe handwritten documents so as to enrich the metadata of archival items with limited metadata. The call for contribution recognises that libraries and archives may not necessary have the capacity to describe an increasing number of items in their collections. The project taps on the domain knowledge of volunteers on specific topics to further enrich the metadata.

In the three months since the launch, the portal has generated substantial interest by the public in general with over 4,700 items described and transcribed.



Figure 6: Citizen Archivist Portal

3.4 Regional and International Collaborations

Collaborations at the regional and international levels are not new. Success factors however are determined not just by NLB this time around, but by active participation and collaboration by libraries around the region and globally. Two examples cited here are the ASEAN Digital Library and the Biodiversity Heritage Library.

3.4.1 ASEAN Digital Library (ADL)

The Association of Southeast Asian Nations⁵ (ASEAN) consists of ten member states (Indonesia, Malaysia, Philippines, Singapore, Thailand, Brunei Darussalam, Vietnam, Laos, Myanmar and Cambodia). The ASEAN Digital Library (ADL) aims to foster greater awareness of the diverse cultures and heritage in this region as the national libraries from the ASEAN nations hold a great wealth of library resources. ADL is a regional effort to draw together these national databases so that the repositories can be seamlessly searched and accessed.

Accessing the diverse content in different languages requires regional initiatives such as ADL to cross the barriers of access due to language, budgetary and technological differences. The current participating countries in ADL are Cambodia, Laos, Thailand, Vietnam and Singapore.

3.4.2 The Biodiversity Heritage Library (BHL)

The Biodiversity Heritage Library⁶ (BHL) was formed in 2006 as a consortium of ten (now 23) natural history, botanical garden and university libraries with the purpose of digitizing key biodiversity literature in an open environment. The content is freely and openly available to the world and currently the BHL Secretariat is based at the Smithsonian Libraries (Washington, DC). In 2014, the National Library Board of Singapore became the 16th member as well as the first institution to join BHL as both a member of the global BHL consortium and a global node (BHL Singapore).

⁵ ASEAN - <http://www.asean.org/asean/about-asean/overview>

⁶ Biodiversity Heritage Library (BHL)- <http://www.biodiversitylibrary.org/>

The BHL Singapore collection contains 47 volumes from 46 titles, made up of 14,864 pages as of this writing (Figure 7). More content from NLB and its Singapore partners are being prepared for inclusion to BHL Singapore.

Page-level metadata need to be created before the digitised materials can be uploaded to BHL. The Smithsonian Libraries created an open source software known as the Macaw Metadata Collection tool⁷ to assist contributors to perform the page-level metadata tagging and to manage the uploading workflow. For the purpose of BHL Singapore, NLB installed the Macaw software in its data centre. The technology team in NLB worked closely with the developer of Macaw and suggested several code changes to Macaw needed to meet the requirements of NLB. These changes were incorporated into the new releases of Macaw.

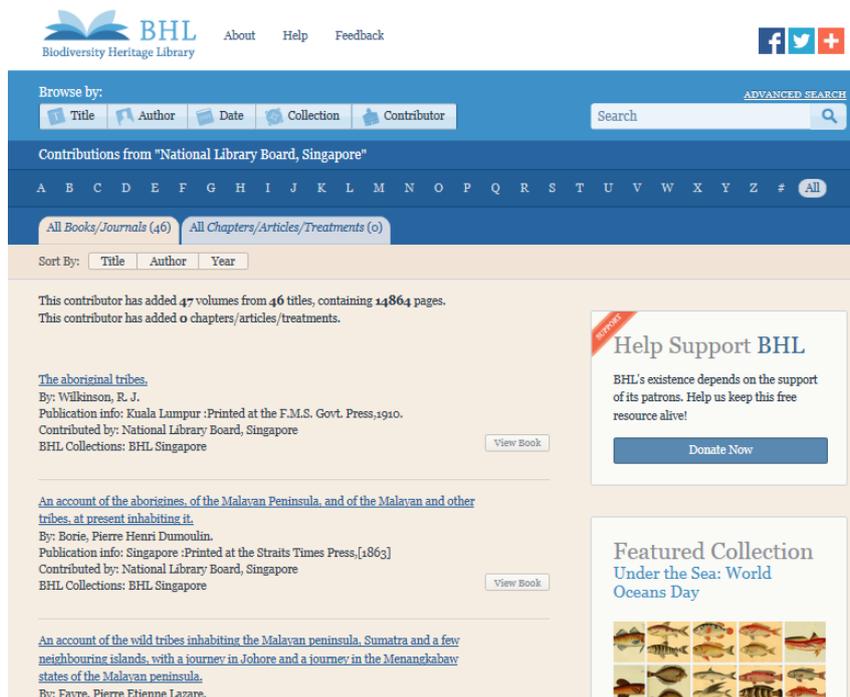


Figure 7: Biodiversity Heritage Library (BHL) Singapore collection

4. CONTENT AGGREGATION TO TRANSFORMATION

NLB acquires e-book resources from vendors such as Overdrive, EbscoHost, Beijing Apabi and World EBook Library to meet the needs of its users. In the past, library patrons need to visit various websites to discover these resources. NLB took a leap forward in simplifying the user experience by aggregating the various websites into a single search and discovery service called eReads⁸ (Figure 9). This was achieved using something that most of these vendors already provide when NLB subscribed to their services - the metadata of their content. The metadata of all the e-books were consolidated into a central repository.

The differentiating factor next was to ensure that the human element was still intact. A backend service was built that allowed librarians to curate and showcase titles based on themes, thereby giving a fresh discovery element for the users. These themes can be changed

⁷ Macaw Metadata Collection tool - <http://www.sil.si.edu/Galaxy.cfm?id=2.174>

⁸ eReads (<http://eresources.nlb.gov.sg/ereads>)

regularly allowing the librarians to collaborate and share curated themes or categories not only on eReads, but other NLB e-book platform services. This opens up a world of the possible services beyond the traditional website, into new innovative concepts such as remote e-book kiosks and walls (e.g., having e-book kiosk in schools, bus stops and other government libraries).

This role of curating e-content is very similar to what librarians do in physical libraries creating booklists and book, newspaper and magazine displays, the only exception being that the displays done on eReads extend beyond the physical library space and into the digital realm.

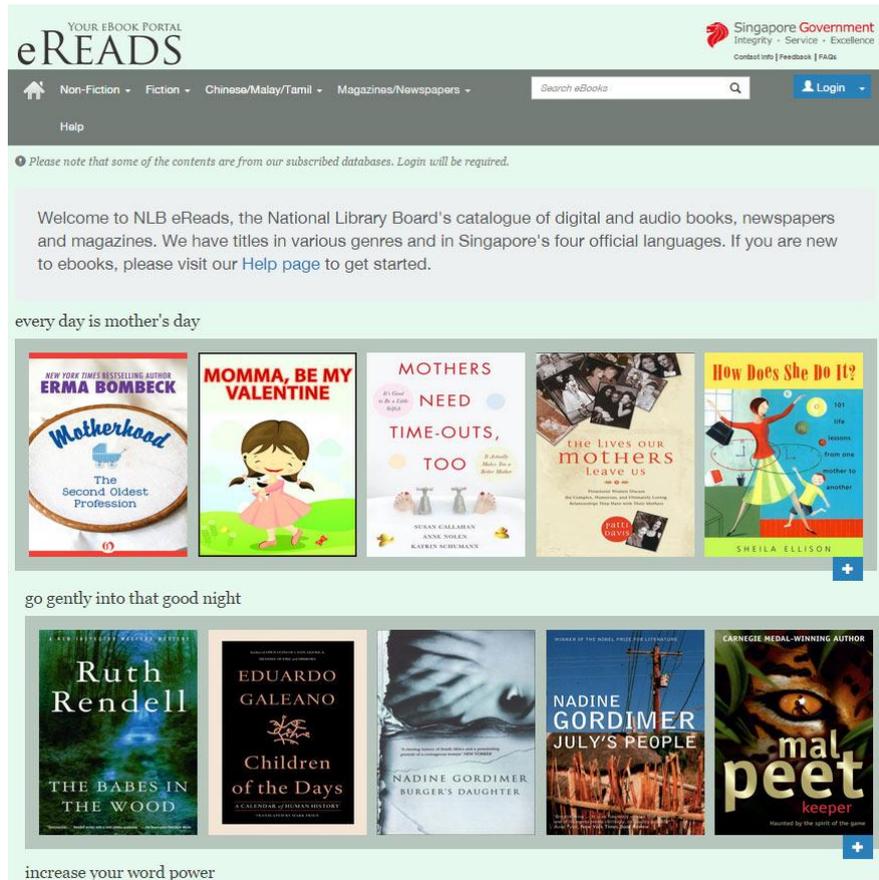


Figure 8: eReads Portal

5. CONTENT AGGREGATION ENABLER

Over the years, NLB has developed a suite of re-usable application components (collectively termed the NLB Service Enablement Architecture) to power service innovations (Figure 9). A key layer is the set of components for ‘rich media content management’ which handles content and metadata management and transformation.

The Content Management Service (CMS) is NLB’s central service that manages access to the vast and myriad collections of NLB content (including HTML text, JPG, PDF, ePUB, MP3, MP4 and many more). NLB has also adopted the NLB Application Profile (NLBAP), a Dublin Core Application Profile (DCAP) as its internal metadata standard to describe digital

objects, providing the necessary scalability and semantic interoperability with other applications.

The CMS is a content-as-a-service platform (using the Alfresco⁹ open source software) that is not tightly coupled with any online services. It is open and extensible to new content websites that aggregate content from different collections. Importantly, it provides full flexibility in the design of the user experience of the website, while consuming the APIs provided by the CMS.

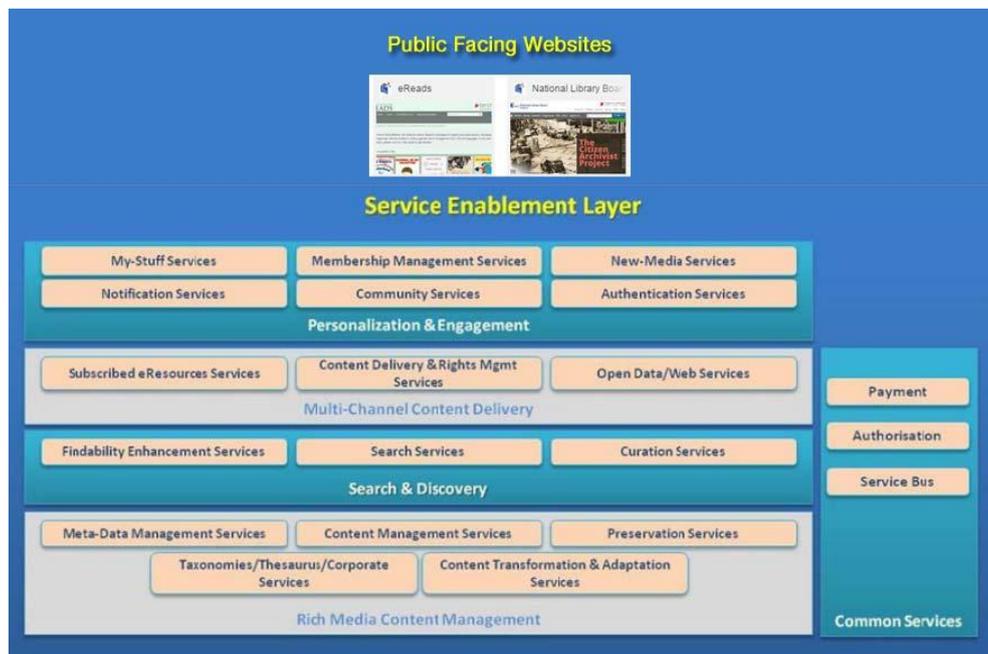


Figure 9: Service Enablement Layer

The CMS uses the Apache Solr¹⁰ search engine that provides fast and robust search capabilities, including metadata and full-text indexing, and facets. The OneSearch service leverages on Solr to enable searches across library, archives and museums content.

The configuration of the production CMS platform is shown in Table 2.

⁹ NLB is using the Alfresco Community Edition as its content-as-a-service platform since 2010. Alfresco is an open source Enterprise Content Management System (ECM) that manages all content in an organisation, and provides services and controls to manage the content.

¹⁰ <http://lucene.apache.org/solr/>

Servers	Brief description
Alfresco (Authoring) Server	Used for bulk ingest. Redhat Enterprise Linux Operating System Tomcat Application Server Alfresco Application Apache Solr
Alfresco (Read only) Server Cluster	Used for online service access. Redhat Enterprise Linux Operating System Tomcat Application Server Alfresco Application Apache Solr
DB Server Cluster	Redhat Enterprise Linux Operating System MySQL DB Server

Table 2: Production CMS

The platform has been used in many content-rich NLB services and has become core to NLB. The agility of this loosely coupled, component-based architecture provides NLB the flexibility that enables service innovation to be built upon established, tested and repeatable processes. To fully leverage on this, NLB has also developed a number of its own APIs (in the form of web services) on top of the out-of-the box ones. Examples of these include search, retrieve metadata, and retrieve content.

Retrieval of metadata and content was only the first step. To increase findability and provide recommended titles, NLB embarked on developing a service that provided high quality associations between its vast amount of resources. This was done using text analytics processing algorithms supported by Apache Mahout¹¹ (an open source machine learning and data mining software) that have been developed to operate in an Apache Hadoop¹² cluster framework.

6. CONCLUSION

Digital content is being created at an exponential level. NLB recognises the tremendous values of collaborations at all levels. Successful collaborations are ‘win-win’ where the collaborating institutions come together to provide a delightful service to the library and archive users.

NLB has pursued all opportunities for collaboration, from the OneSearch service across the libraries and archives, national initiatives such as Singapore Memory and Citizen Archivist projects, to international ones like the ASEAN Digital Library and Biodiversity Heritage Library. Leveraging on technology, NLB sought to reinvent process already long established

¹¹ Mahout (<http://mahout.apache.org/>)

¹² Apache Hadoop (<http://hadoop.apache.org/>)

by librarians in the physical sphere and innovating it for the digital realm (such as book selection by librarians and meta-content description by volunteers).

An agile suite of re-usable technology stack is needed to deal with the additional complexity in collaborations, particularly the diversity of content, formats, standards and access rights management needs. Robustness and the ability to scale are also critical to handle the expanded repositories made possible through collaborations.

NLB seeks to constantly innovate itself by looking at its current process as well as leveraging ideas from various communities of practice and open source solutions to keep up with global digital trends.

References

KIA, Siang Hock and WANG, Zhi Liang (2013) Content-as-a-service platform with the al fresco open-source enterprise content management system. Paper presented at: IFLA WLIC 2013 - Singapore - Future Libraries: Infinite Possibilities in Session 108 - Information Technology.

KIA, Siang Hock and LIAU, Yi Chin and ONG, Ian (2014) Inter-connected Network of Knowledge – The NLB Journey. Paper presented at: IFLA WLIC 2014 - Lyon - Libraries, Citizens, Societies: Confluence for Knowledge in Session 208 - Knowledge Management. In: IFLA WLIC 2014, 16-22 August 2014, Lyon, France.

ONG, Ian and GOH, Cindy and CHUA, Lilian and PAK, Peter (2014) Empowering the Library Patron: The Public Libraries of Singapore's experience with transactional services delivered through a mobile application. Paper presented at: IFLA WLIC 2014 - Lyon - Libraries, Citizens, Societies: Confluence for Knowledge in Session 210 - Information Technology. In: IFLA WLIC 2014, 16-22 August 2014, Lyon, France.

Pak, Peter (2014) eReads – NLB's On-going Journey in Enhancing the e-Book User Experience. Paper published at: Trends & Issues in Library Technology, IFLA IT Section Newsletter, December 2014.

Zaccheuss, Melody. (2015, March 18). 74 projects under the Singapore Memory Project to go on show in lead up to SG50. The Straits Times. Retrieved from <http://www.straitstimes.com/news/singapore/more-singapore-stories/story/74-projects-under-the-singapore-memory-project-go-show-1>

Making Lasting Connections. (National Library Board Singapore). Retrieved May 25, 2015, from <http://www.nlb.gov.sg/Portals/0/Reports/fy13/>

About Us. (Biodiversity Heritage Library). Retrieved May 25, 2015, from <https://drive.google.com/file/d/0B00hDkSQMhfDTTBwUG9MS2JPcGM/view?pli=1>