KNOWLEDGE SECTOR DIGITIZATION AND THE OPPORTUNITY FOR A BC PROVINCIAL DIGITAL LIBRARY

Prepared by Mike Conroy for the BC Libraries Cooperative
July 6, 2015

THE CHALLENGE

British Columbia libraries, museums, archives, galleries and historical societies have been digitizing historical material for over a decade. At the request of Libraries Branch, Ministry of Education, BC, the following is an attempt to assemble an inventory of the institutions, materials and digitization methodologies that have been or are currently in use in BC. **We provide recommendations for strategic integration and further development of shared services that may reduce costs and provide greater coordination across the knowledge**. Further, the inventory serves to describe what might compose the extent of the currently available knowledge sector contribution to a "Provincial Digital Library".

A Provincial Digital Library brings together electronic resources from across the province making them searchable from a single portal. It is a vital resource for all British Columbians, providing access to unique historical treasures and the intellectual production of BC. Bringing all this content together allows users to make new connections and discover and contextualize our collective past. It is invaluable to researchers and genealogists, students of all ages, government officials and the media. But a portal is really just one aspect and opportunity for a PDL.

In this report we focus specifically on BC's inventory of digitized historical materials which would comprise a significant portion of a PDL. Institutional repositories at BC academic institutions including the upcoming BC Institutional Repository Network (BC IR Network) will comprise another significant component. Over time, a PDL will grow into an index that could encompass an ever-widening array of materials like datasets, vital statistics and electronic records. With the recent agreement between the Government of BC and the Royal BC Museum to address the archival backlog in the Province's archives, the PDL could be of interest as a means of keeping with the principle of providing "citizen-centred" access to archival materials.[1]

Most, if not all, of the content that will be accessible through the PDL has been generated or digitized with public money; a PDL provides the opportunity to maximize these investments and make digitized content discoverable.

DIGITIZATION AND ITS PURPOSE IN THE BC KNOWLEDGE-SECTOR

When we talk about digitization we generally refer to the processes involved in the conversion of analogue formats to digital ones. Some common types of digitization include: scanning photographic prints or negatives to TIFF or JPEG, audio tape to WAV or mp3, printed documents or letters to TIFF or PDF. Other processes in the digitization workflow include file-naming, optical character recognition (OCR), adding metadata according to accepted standards, uploading to a digital asset management system and digital preservation. It's worth noting that transcription of hand-written letters may also be considered digitization and digital photography of museum artifacts is lumped in as well. Digitization is unnecessary for information that is "born digital" such as recently

---

[1] "Agreement reached to address archival backlog." *Government of British Columbia,* April 15, 2015.
    https://news.gov.bc.ca/stories/agreement-reached-to-address-archival-backlog.

published articles, theses and dissertations that comprise a substantial portion of higher-education institutional repositories.

Heritage institutions increasingly view digitization as core to their missions and mandates. Where once it was considered a novel or fringe activity, digitization has cultivated greater interest in local holdings and elevated Institutions' profiles in their communities. They are compelled to digitize for several reasons:

- Institutions in BC possess rare and often unique historical material that is difficult to access.
- Researchers and historians must often travel to access material.
- The public at large may not be aware of the existence of materials in the stewardship of local institutions.
- Digitized materials are easier to access.
- Digitized materials are easier to find.
- Digitized materials get used more often by more people.
- Digitization can reduce wear and tear on the original object.
- Digitization projects are of high interest to local communities who want to share and pass on their history and heritage to younger generations.

Beyond the factors motivating institutions to digitize there are secondary advantages that should be noted. Digital collections are often used in a variety of ways including some not originally anticipated:

- Because historical images are more discoverable on the web, publishers regularly seek out and license images from institutions with historical materials online.
- Some digitizing institutions sell prints or high-resolution images to partially offset the costs associated with digitization.
- TC2, the Critical Thinking Consortium, developed K-12 learning modules on Chinese-Canadian history using digitized materials from the Chinese Canadian Stories Project. The learning modules are downloadable online free of charge.
- Heritage Vancouver used UBC's collection of digitized newspapers to locate years of previously missing Vancouver building permits for their own building permit database.
- The Revelstoke Museum & Archives used UBC's collection of digitized newspapers as research material for an exhibit on the history of snow and avalanche research. During the project they gathered information on previously unknown avalanches and created new entries in the Canadian Avalanche Centre Database.
- As part of its fundraising, the Friends of the Vancouver Public Library sells items screen printed with digitized images from the Vancouver Public Library's Special Collections.

There remains considerable untapped economic potential in BC's current stock of digitized material. Municipalities, regions and the tourism sector could be using our digital past in their promotions. Digitized material could prove to be vital for tomorrow's open textbooks. In retail and marketing there are opportunities that have not yet been exploited. For example, the Nine O'Clock Gun Company, a Vancouver business, makes ball caps emblazoned with the insignias of Vancouver's sporting past. In marketing a cap with the logo of the 1920 Vancouver Athletic Club, the retailer refers its customers to the City of Vancouver Archives for historical images featuring the logo. If a Provincial Digital Library could bring the province's digital history together, entrepreneurs and others could be encouraged to find more ways to make use of, remix, and repurpose these materials.

INITIAL FINDINGS

To date we have identified 85 different organizations or institutions that have begun to digitize their historical collections in BC. Organizations were primarily identified through annual lists of BC History Digitization Program recipients and a list of online photograph databases compiled by the Archives Association of British Columbia. Our list is not exhaustive but provides an adequate view of digitization activities in the province. **Strategies for providing public access to digitized materials vary widely amongst the 85 institutions. We count at least 32 different identifiable platforms.** There are at least 25 institutions whose collections are on unidentifiable platforms. That is, they are on customized platforms or web pages where the type of CMS or database type is not given or not immediately determinable by visiting the site.

Some organizations rely on YouTube and Flickr to host and make material publicly available. Some mount their collections on static html web pages. However, most institutions use locally hosted platforms giving them greater control over their digital assets and the ability to apply structured metadata.

TABLE 1 DIFFERENT PLATFORMS, SERVICES AND TECHNOLOGIES USED TO HOST DIGITAL COLLECTIONS IN BC AND THE NUMBER OF INSTANCES

Platform	#	Platform	#
Newspapers.com	1	ARGUS.net	1
Artefacts Canada	1	Atom	7
HTML	15	CONTENTdm	3
ContentPRo (Innovative Interfaces)	1	Coppermine Photo Gallery	1
Drupal	5	DSpace	2
eXist xml database	1	Flickr	1
Greenstone	1	ImageFolio	1
InMagic	6	Internet Archive	1
Islandora	2	Joomla!	1
MINISIS Inc.	2	Multicultural Canada (Drupal)	2
Omeka	1	Past Perfect	6
Plone	1	Squarespace	1
Ubertor	1	Virtual Museum of Canada	2
Vita Toolkit	1	WebGENCAT	1
WordPress	4	YouTube	1
Unknown	25	Solr	1

At a glance, access to digital collections in BC appears fragmented both in the number of access points and in the methods used to provide access. However, it's worth noting a few examples of alignment and cooperation. For example BC's three largest post-secondary institutions, UBC, SFU and UVic all use instances of CONTENTdm. SFU's CONTENTdm instance also hosts material from Emily Carr University and the Vancouver Public Library. UBC's CONTENTdm instance is home to 2 newspapers contributed by the Sunshine Coast Museum & Archives.

There are also more concerted attempts at integration. The Archives Association of BC maintains an Atom instance called Memory BC which acts as a union catalogue of archival descriptions in BC. While Memory BC does not currently hold digital objects, the site offers users the ability to search across archival holdings from dozens of BC's archives.

BC DIGITIZATION COALITION

The BC Digitization Coalition is the BC knowledge sector's first attempt at cross-sector digitization coordination. The Coalition describes itself as "a group working toward the creation of a provincial digitization strategy" and formed following the BC Digitization Symposium in 2008. Following the event, the Symposium Organizing Committee continued to meet as the BC Digitization Coalition. The Coalition's draft terms of reference describe its purpose and goals as follows:

> ### *Purpose*
>
> - *Continue to implement key recommendations identified by BC Digitization Symposium participants*
> - *Advocate for collaboration and partnerships, especially around opportunities for funding to support ongoing digitization initiatives*
> - *Facilitate dialogue and discussion around digitization strategies in communities across the province*
> - *Work with small and rural communities to support digitization efforts*
>
> ### *Guiding Principles:*
>
> - *In conducting its activities the Coalition will consult widely with a broad range of stakeholders and will ensure that communication is open and transparent.*
> - *The Coalition will work to support digitization efforts without duplicating the work of others.[2]*

Made up of representatives from academic, public and k-12 libraries, museum, archive and other sectors, the Coalition's two main efforts to date are the West Beyond the West search portal and the West Beyond the West Hosting Service. The search portal is an attempt to provide search across BC digital collections by aggregating disparate collection metadata from BC institutions. When the user conducts a search, results point back to host institutions' collections. There are currently contributions from 12 institutions that are searchable in the WBTW search portal. The portal is hosted by the BC Libraries Co-operative.

The hosting service operates a multi-repository instance of Atom, an open-source digital asset management system designed specifically for the archival community. Atom is the same software that powers Memory BC. While Memory BC only contains archival descriptions, WBTW Hosting is intended to host digital objects. Atom is also used by at least 6 other institutions in BC mostly for archival collections. The hosting service currently only has one client but has the potential to host collections currently locked in old or unsuitable platforms and from institutions just starting out in digitization.

The BC Digitization Coalition has been less active in 2015. With no operating budget, the continued development of the portal and hosting service are uncertain. While the services in their current form may be unsustainable, they present proofs of concept for what cross-sector collaborations and shared services could look like.

---

[2] "Terms of Reference (Draft)." *BC Digitization Coalition*, Accessed June 30, 2015.
http://coalition.westbeyondthewest.ca/about-2-2/reports-and-documents/terms-of-reference-draft/.

BC HISTORY DIGITIZATION PROGRAM

In its 7th year, the BC History Digitization Program provides matching funds to BC institutions undertaking digitization projects. The Program is an initiative of the Irving K. Barber Learning Centre and fulfills the institution's commitment to "support lifelong learning by people throughout B.C. and the world ...[and] enable them to work effectively together in continuing to build a better British Columbia through the sharing of knowledge and experience."[3] As stated in the BCHDP application, the goal of the Program is to "promote and enhance access to unique British Columbia historical materials by providing financial support for digitization activities."[4] The Program provides grants up to $15,000 per project and gives out approximately $180,000/year to institutions province-wide.

Since its inception the BCHDP has given out well over $1million. In addition to funding, the program has arguably had a positive effect on the overall quality of digitized material available: Applicants are compelled to submit a scoped project plan that includes goals, budget, personnel and equipment requirements, and technical details like how their proposed scanning resolution and file formats fit with best practices. A volunteer adjudication committee made up of librarians, archivists and academic historians selects projects to be funded based on their viability and value to the history of British Columbia. Though not all projects have been successful, this process is meant to protect against failure and ensure a minimum standard for digitization projects. A list of BCHDP recipients is provided in the appendix.

GOING IT ALONE: WHY ARE THERE SO MANY DIFFERENT PLATFORMS IN USE IN BC

There are a surprising number of platforms used to host digital collections across the province. At last count there are 32 different digital asset management systems, databases, websites and hosting platforms in use. We can only speculate the decision making processes at each institution with regards to selecting a specific platform. It may come down to costs or the availability of in-house expertise or recommendation from their local webhost. However there are several key factors or functional requirements that necessarily limit the range of options for a given institution. They include, but are not limited to, media types and metadata:

- Not all platforms are equal to the task of displaying video or streaming audio for example.
- Not all platforms have file structures that allow for compound objects or archival hierarchies that are required or desirable for contextualizing digital objects in a manner that resembles the order of a physical archive.
- Metadata standards differ between libraries, archives and museums. In the library world, digital objects are most commonly assigned metadata according to rules laid out in the Dublin Core metadata schema. In Canadian Archives, RAD (Rules for Archival Description) are commonly followed for the creation of metadata.
- There are many more metadata schemas in use across all sectors.

While there would doubtlessly be advantages to narrowing the range of platforms in use across the sector, it is unlikely that there is a single platform that would suffice for all institutions. Creating user communities around a few key platforms could be one approach to coordination. This would allow for the exchange of expertise between

---

[3] "Statement of Purpose and Charter of Principles." *Irving K. Barber Learning Centre*, Accessed July 2, 2015.
http://www.ikebarberlearningcentre.ubc.ca/aboutus/charter/.

[4] "BC History Digitization Program: Guidelines and Application Material (2015)." *Irving K. Barber Learning Centre*: 2.
Accessed June 6, 2015. http://ikebarberlearningcentre3.sites.olt.ubc.ca/files/2011/09/digiapp20151.doc.

and amongst institutions of similar size and mandate, and could also be used to set norms for metadata application. To some extent, user communities already exist for some platforms but there isn't a geographic component to bring together regional users who could also meet in person.

The BC Digitization Coalition's WBTW Hosting service is still an option for a hosted solution. Over time, more features may be added to the Atom hosting software making it viable for a greater number of digital collections. Other options include hosting a multi-repository instance of Islandora, the software recently chosen as the platform for the shared institutional repository in the post-secondary sector. Islandora is currently used by the Prince Rupert Library and Prince George Public Library to host digitized newspapers. Vancouver Public Library and Simon Fraser University Library also have plans to migrate to Islandora in the next year.

MODELS FOR A PROVINCIAL DIGITAL LIBRARY

Existing portals in Canada and around the world can be categorized as

1. Metadata aggregators (like West Beyond the West) where the site allows users to search metadata that has been brought together from multiple repositories. Search results point users to hosting institutions pages. Other examples include:

- [Canadiana Discovery Portal](): Canadiana brings together digital collections from institutions across Canada making them searchable from a single portal. The site relies on metadata contributed from institutions across Canada and has focused on larger institutions. Canadiana has its own unique metadata schema used to normalize data from a broad array of disparate collections. Canadiana's repository is the data store for the West Beyond the West. Canadiana relies on metadata updates submitted by partner institutions. Consequently, in many instances the portal does not contain metadata for more recent collections.
- [Digital Public Library of America](): Similar to Canadiana, the DPLA aggregates metadata from larger partner institutions as well as the Internet Archive and Hathi Trust. Notably, partner institutions act as "service hubs" and do the work for metadata aggregation and transformation at the regional level allowing for greater capacity and the ability to index digitized collections from smaller institutions.

2. Single repositories where the organization hosts all of the digital objects even when the origin of those digital objects are from other organizations. These portals tend to be "curated" collections related to a province or region rather than an attempt to gather together every available resource. Example:

- [Manitobia](): Manitobia is a selection of resources on Manitoba history brought together by the Manitoba Library consortium and Council of Post-Secondary Education (now Education and Advanced Learning). All content is hosted by Manitobia allowing for full control of metadata and objects. The site contains books, maps, newspapers and photographs along with a special section for k-12 students and educators.

3. A combination of both.

- [Trove](): Trove is the digital library run by the National Library of Australia. It links together content from external partners, earlier services developed by the NLA and its own repository of digitized newspapers.

West Beyond the West is currently the only British Columbia example of a platform that aggregates metadata from multiple repositories. However, The University of British Columbia is developing a front-end interface called Open Collections to link together three of its digital collection platforms. When complete, the portal will link UBC's

CONTENTdm, Atom and DSpace instances. UBC's approach should be examined for its viability as a potential model for a PDL.

DIGITAL PRESERVATION

Digital preservation remains an emerging practice in British Columbia. While larger institutions like UBC and SFU have developed and are implementing digital preservation plans, such activities are beyond the capabilities of smaller organizations.

The Council of Prairie and Pacific University Libraries (COPPUL) currently runs a digital preservation service called Archivematica-as-a-Service (AaaS). The services uses Artefactual's locally developed open-source digital preservation system called Archivematica and UBC's EduCloud server service. While AaaS was developed and is run for COPPUL members, the service model may be of interest if developing a digital preservation component to the PDL if it is deemed in scope. Although digital preservation is complex and expensive, there could be substantial savings by investing in a model on a provincial scale.

RECOMMENDATIONS

1. Implement a Provincial Digital Library that aggregates metadata from existing digital collections and electronic resources, and that is flexible enough to accommodate growth as new resources become available.

The most practical approach to establishing a Provincial Digital Library is through aggregation of metadata from British Columbia's existing repositories including the upcoming post-secondary shared IR. While the approach will be similar to the West Beyond the West, the PDL will necessarily be wider in scope, encompassing more than the digitized materials described here. In order for a project of this scale to be successful, it will require consultation and buy-in from key institutions and partners.

2. Host a symposium bringing together knowledge-sector stakeholders and government.

The BC Digitization Coalition, with a membership made up of knowledge-sector organizations, can play a role in bringing together the community with government officials and others to craft an agreed upon vision for the PDL.

3. Monitor developments at the national level

Inasmuch as the PDL will be based on best practices, it will anticipate and influence any emergent national standards.

4. Establish a PDL metadata working group.

A metadata working group with all-sector representation will need to develop and promote a comprehensive data model for the PDL. The working group will look at existing data models including those of Canadiana and the DPLA, and ensure that the PDL model is flexible enough to accommodate a growing variety of content and open enough to allow for potential export to other aggregators like DPLA.

5. Maintain WBTW Hosting. Consider a second service using Islandora as a platform.

Though it's unclear whether WBTW Hosting currently provides a service that is demanded by smaller provincial institutions, an available, stable and standards-compliant option is required if smaller institutions are to participate in the PDL. WBTW is relatively inexpensive to operate compared to maintaining numerous separate digital asset

management systems locally. In most cases, data will also be more secure in WBTW Hosting, which is remotely backed up, than in local digital asset management systems. **An appropriate "owner" and management plan for the service should be determined by the BC Digitization Coalition as soon as possible.**

The number of larger institutions adopting Islandora as the platform for their digital collections suggests that it may also be an appropriate platform for a shared digital collection hosting service. An Islandora instance could be administered similar to the way WBTW Hosting is currently administered, providing additional flexibility for smaller institutions. However before embarking on the creation of a second service, target institutions should be canvassed to gauge interest and to ensure such a service would meet their needs. **Oversight of both WBTW Hosting and a possible Islandora service could occur together within a single organization.**

6. Provide grants for data migration and metadata remediation.

Migrating existing digital collections to new digital asset management systems can be costly and requires expertise that may not be available at small and even mid-size institutions. Furthermore, in order for collections to be aggregated into a PDL, a significant amount of metadata and mapping work will need to take place. Grants directed specifically at these tasks will increase cooperation and lead to better metadata.

7. Tie funding for digitization to full participation in the PDL.

Ensure that any Provincial funding for digitization is conditional on compliance with standards required by the PDL.

8. Align BC History Digitization Program funding to support objectives of the PDL.

The BC History Digitization Program has been instrumental in the digitization of historical material throughout the province. As the BCHDP is a major source of digitization funding in BC, we view it as a critical contributor to the success of the PDL. The Program could be further aligned with the vision of the PDL by requiring future recipients to expose their collections to aggregation, ensure their metadata is appropriately mapped, when possible, to the PDL data model, and encourage the utilization of acceptable repository platforms. A portion of BCHDP funding could also be diverted to support specific recommendations such as data migration.

9. Provide funding to the BC Libraries Cooperative to continue the next phase of planning.

As a cross sector member organization, the BC Libraries Cooperative is uniquely positioned to facilitate a collaborative approach towards a PDL. The Cooperative can gather stakeholder feedback from public libraries and, as a member of the BC Digitization Coalition, provide supports for a cross-sector symposium. The Cooperative is also suited to taking on responsibility for WBTW Hosting. Temporary funding for WBTW Hosting could allow the Co-operative to develop a sustainability model for this service.

QUESTIONS FOR DISCUSSION

We end with a few questions designed to stimulate discussion and solicit feedback from stakeholders on the 9 recommendations above.

1. Is your institution interested in leading the way?

2. Is your institution interested in a shared service approach?

3. Do the recommendations contained herein capture existing best practice? Is anything missing?

4. Is the vision here big enough? Too big?

5. Would you be interested and able to participate on committees as we move forward?

6. Who are the key individuals and institutions we should be in discussions with?

ABOUT THE AUTHOR

Mike Conroy has been involved in digitization efforts in British Columbia since 2010. He previously managed digitization projects at the University of British Columbia, coordinated the BC History Digitization Program and chaired the BC Digitization Coalition's technical subcommittee. Mike received his Master's of Library and Information Studies from the University of British Columbia in 2009.

APPENDIX: 85 BC INSTITUTIONS THAT HAVE OR ARE ACTIVELY DIGITIZING HOLDINGS.

Alberni Valley Museum	Langley Centennial Museum	Squamish Public Library
Barkerville Heritage Trust	Lund Community Society	Sunshine Coast Museum and Archives
BC Dairy History Society	Mission Museum	Surrey Archives and Museums
BCIT	Morris and Helen Belkin Art Gallery	Terrace Public Library
Beaty Biodiversity Museum	Museum of Anthropology	TNRD Libraries
Bella Coola Valley Museum	Museum of Vancouver	Touchstones Nelson
Bowen Island Historical Society	New Westminster Public Library	Trail Historical Society
British Columbia Medical Association	Nicola Valley Museum & Archives	Trinity Western University Archives
Campbell River and District Museum and Archives Society	Nikkei National Museum	Tumbler Ridge Museum Foundation and Peace Region Paleontology Centre
CFB Chilliwack Historical Society	North Vancouver Museum and Archives	Two Rivers Gallery
Chilliwack Museum & Archives	Northern BC Archives	UBC Library

CiTR (The Student Radio Society of UBC Vancouver)	O'Keefe Ranch and Interior Heritage Society	Union of BC Indian Chiefs
City of Burnaby Archives	Potters Guild of BC	United Church of Canada
City of Richmond Archives	Powell River Historical Museum and Archives Society	University of Victoria Libraries
City of Surrey Archives	Prince George Public Library	University of Victoria Archives
City of Vancouver Archives	Prince Rupert City & Regional Archives	University of Victoria Art Galleries
City of Victoria Archives	Prince Rupert Library	Vancouver Holocaust Education Centre
Delta Museum and Archives	Quesnel and District Museum and Archives	Vancouver Island University
Emily Carr University of Art + Design Library	R.J Henry Museum and Archives	Vancouver Public Library
Esquimalt Municipal Archives	Reach Gallery and Museum	Vanderhoof Public Library
First Metropolitan United Church Archives	Revelstoke Railway Museum	VIVO Media Arts Centre
Greater Vernon Museum & Archives	Richmond Museum	West Kootenay Women's Association
grunt gallery (Visible Art Society)	Royal BC Museum	West Vancouver Archives
Gulf of Gerogia Cannery Soc.	Royal Roads University	West Vancouver Memorial Library
Hallmark Society	Saanich Archives	Western Front
Hazelton Area Historical Association	Salt Spring Island Archives	Whistler Museum and Archives
Jewish Museum and Archives	Satellite Video Exchange Society	White Rock Museum and Archives
Ladysmith Archives	Sechelt Community Archives	
Lake Country Museum and Archives	Simon Fraser University	