British Columbia Government Publications Digitization Project: Proof of Concept

Proposal to the British Columbia Research Libraries Group

January 17, 2011

(Corrected version – February 16, 2011)

by

Marilyn Carr-Harris, Acquisitions/Serials Librarian: British Columbia Government Publications, LLBC

Gail Curry, Instruction/Data Librarian, UNBC

Carla Graebner, Liaison Librarian: Archaeology, Economics and Government Information, SFU

Susan Paterson, Librarian, Government Publications and French, UBC

Caron Rollins, Librarian, Government Publications and Political Science, UVic

Table of Contents

1 Executive Summary

The British Columbia Research Libraries Group (BCRLG) struck a task group, in conjunction with the Legislative Library of British Columbia (LLBC), in spring 2010, to identify and investigate means of digitizing historical British Columbia government publications of research value. Identified publications would include both executive publications of the Legislature and original government publications. The goal of the project would be to increase access to these publications by making them widely and freely available online.

After considerable research and numerous conference calls, e-mails, and consultations, the Task Group is pleased to make the following recommendations to the BCRLG. Note that the publications selected for the initial proof of concept project have been carefully chosen for the different issues they present.

Recommendation #1: The Task Group proceed with an initial proof of concept to digitize one or all of the following publications, listed in priority order: the first ten years of the *BC Sessional Papers*; the *Government Gazette, British Columbia,* 1863-1871 (including the *Government Gazette for Vancouver Island*, 1864 – 1866 and the *Government Gazette for the Colonies of Vancouver Island and British Co*lumbia, 1859-1860); and the *BC Newspaper Card Index*.

Recommendation #2: The digitization for the proof of concept project should take place at the University of Victoria (UVIC), with the LLBC providing the publications to digitize.

Recommendation #3: The Task Group consult with digital initiative and technical experts at our institutions regarding decisions related to the technical, digitization, and metadata best practices presented in Appendices 7.4 and 7.5 of this proposal.

At the end of the initial project, the Task Force will report back to the BCRLG with an evaluation of the success of the project, outstanding issues, and recommendations for future action.

2   Acknowledgements

3   Introduction

Until recently, researchers expected to, at best, go to their local library or, at worse, incur travel expenses, in order to examine historical government documents. Now, in the age of digitization and the Internet, library users expect to access these documents easily and efficiently from their homes and offices.

In spring 2010, the British Columbia Research Libraries Group (BCRLG) struck a task group, in conjunction with the Legislative Library of British Columbia (LLBC), to pursue the feasibility of identifying and locating suitable historical British Columbia government publications for digitization. These publications would include both executive publications of the Legislature and original government publications. The end goal would be to increase access by making the digitized publications widely and freely available online.

4   Criteria for Selection & Prioritization

The Task Group defined the following selection/prioritization criteria, in accordance with the goals identified in Section 1, to use in selecting and prioritizing materials or collections to be digitized:

- **Value**: Documents are of high research value to library users, participating institutions, and the Province.

- **Duplication factor:** Duplication avoided, except under mitigating circumstances (e.g., it may be less expensive to digitize large runs of documents rather than to track down whether specific documents have already been digitized).
  Note: The Task Group will continue to liaise with the BC Bibliography Group to ensure that there is no overlap with their project.

- **Copyright**: Must have the right to create digital copies.

- **Condition/availability/format of material:** Strive for clean copy of original material in original format (best copy). Scan from microform only as a last resort. Material must be available and accessible for scanning. Format, size, and condition of items must be conducive to scanning.
- **Funding:** Projects with outside funding take precedence.

With the above guidelines and criteria in mind, and with the assistance of the Council on Library and Information Resources (CLIR) Decision-Making Matrix[1] presented in Appendix 6.1, the Task Group identified potential documents to digitize (Appendix 6.2). All these documents are feasible to digitize, are of high research value, and are considered high priority collections by other librarians within our various institutions. Some of these materials, however, have digital rights issues that need to be resolved.

The Task Group also applied the following broad questions[2] to the items in Appendix 6.2.

- Will the digitization project make the content more widely available?
- Does the project focus on items of enduring value (i.e. historical importance)?
- Does digitization fulfill the needs of preservation and conservation (rescue digitization)?
- Will the digitization project increase the demand of the various collections?
- Is the project in keeping with the policies at an institutional level?

All questions were answered in the affirmative.

5   Digital Rights Management

5.1   Definitions

The process of determining who owns the copyright and how to obtain copyright permission to digitize British Columbia government publications is an ongoing challenge for the Task Group. Depending on the type of document, two sets of rights may apply: parliamentary privilege and Crown copyright.

---

[1] Dan Hazen, Jeffrey Horrell, and Jan Merrill-Oldham, *Selecting Research Collections for Digitization* (Washington: Council on Library and Information Resources, 1998), http://www.clir.org/pubs/reports/hazen/pub74.html.

[2] Stuart D. Lee , *Appendix G: Assessment Criteria for Digitization*, (Oxford: Bodleian Library and Radcliffe Camera, University of Oxford, 1999), http://www.bodley.ox.ac.uk/scoping/assessment.html.

5.1.1 Parliamentary Privilege

Materials printed by the LLBC such as *Hansard*, *Votes and Proceedings*, *Sessional Papers*, bills, and statutes & regulations are protected by parliamentary privilege. The federal Access to Information Review Task Force provides the following definition of parliamentary privilege:

> Parliamentary privilege is the collective and individual rights accorded to parliamentarians to ensure they are able to carry out their functions and perform their duties without obstruction. **The privilege is protected by the Constitution and extends to all matters relating to parliamentary proceedings.** This includes a Member's right to freedom of speech, **the House's entitlement to regulate its own internal affairs** and a Committee's right to call witnesses. [3]

Unlike Crown copyrighted works, parliamentary privilege extended to printed parliamentary publications does not expire; intellectual property rights are held in perpetuity by Parliament.

In British Columbia, legal issues pertaining to parliamentary publications are managed by the Legislative Assembly Law Clerk.

5.1.2 Crown Copyright

Crown Copyright applies to *original* publications of the government, boards, and the Legislative Assembly. According to Section 6.3.4(e)(2) of the Ministry of Finance *Core Policy and Procedures Manual*, "Crown Copyright of any Work means it belongs to the Province and not to individual ministries or any other government agencies."[4]
An important distinction between parliamentary privilege and Crown copyright is the duration of copyright. Crown copyright expires after a fifty year period. Section 12 of the *Copyright Act* stipulates:

> Without prejudice to any rights or privileges of the Crown, where any work is, or has been, prepared or published by or under the direction or control of Her Majesty or any government department, **the copyright in the work shall, subject**

---

[3] Access to Information Review Task Force, "The Legislature – Parliament" in *Access to Information: Making it Work for Canadians: Report of the Access to Information Review Task Force* (Ottawa: Access to Information Review Task Force, 2002), http://www.atirtf-geai.gc.ca/report/report7-e.html.

[4] British Columbia, Ministry of Finance, "Procurement" in *Core Policy and Procedures Manual* ([Victoria]: Ministry of Finance, n.d.), http://www.fin.gov.bc.ca/ocg/fmb/manuals/CPM/06_Procurement.htm#1634e (accessed October 8, 2010).

to any agreement with the author, belong to Her Majesty and in that case shall continue for the remainder of the calendar year of the first publication of the work and for a period of fifty years following the end of that calendar year.[5]

Determining who owns the copyright to government publications can present a challenge. By default, the author (if there is one) owns copyright. Copyright ownership can also depend on the author's relationship with government (contractor versus employee) and the nature of contractual agreements between government and contractors.

The Intellectual Property Program[6], administered through the Ministry of Citizens' Services, is the primary licensing body for all BC Government-owned property. It manages the licensing and sale of government-owned intellectual property and is the point of contact for universities, libraries, and other interested parties seeking permission to reproduce and distribute Crown copyrighted material.

6 Proof of Concept

The Task Group is seeking go-ahead from the BCRLG to proceed with a proof of concept project. This will allow us to

(a) demonstrate the feasibility of this digitization project;

(b) identify, and suggest resolution of, any issues; and

(c) provide the BCRLG with further recommendations for proceeding with additional publications.

6.1 Recommended publications

The Task Group considered the feasibility of obtaining permission to digitize the publications presented in Appendix 6.2. Appendix 6.3 lists the publications, along with associated rights, that we identified as the highest priority publications for digitization. These titles do not appear to be included in any other digital project.

Note that the rights restrictions for the *Sessional Papers* and other parliamentary publications are more stringent than for the other publications. If a parliamentary

---

[5]Copyright Act, Consolidation, Current to September 22, 2010, c. 42, s. 12, http://laws.justice.gc.ca/PDF/Statute/C/C-42.pdf, 18 (accessed October 13, 2010).
[6]British Columbia, Ministry of Citizens' Services, "Intellectual Property Program," Ministry of Citizens' Services, http://www.cio.gov.bc.ca/c io/intellectualproperty/index.page (accessed October 13, 2010).

publication is selected for digitization, further consultation with the Legislative Assembly Law Clerk and other copyright experts will be necessary.

Publications with expired Crown copyright (over fifty years), such as the *Public Accounts*, *Estimates*, annual reports and other ministerial publications, have fewer restrictions. More research on the specific protocols involved with reproducing and distributing expired and more current Crown copyrighted works is required. It will be necessary to contact the government Intellectual Property Program before initiating projects involving Crown copyrighted publications.

We applied the following criteria to the publications listed in Appendix 6.3 to identify publications for a proof of concept project:

(a) Manageable size.

(b) No physical concerns.

(c) No copyright issues.

Based on these criteria, we have identified three potential publications to test in an initial proof of concept project. These have been carefully selected for the different issues they present and are listed in priority order.

- **British Columbia Legislative Assembly. Sessional Papers (first ten years)**
  - Includes all reports tabled during B.C. Legislature sessions.
  - Collection spans the years 1876-1982. For the proof of concept project, we recommend digitizing the first ten years (1876-1886).
  - 1876 (1st session, 2nd Parliament) to 1886 (4th Session 4th Parliament). 10 volumes, pages 7 x 10, paper very brittle. Volumes can include fold out maps and charts. Total page count: 5532 pages plus 10 title pages and approximately 30 index/prefatory pages. *Sessional Papers* are selectively included and listed as individual titles in the *BC Bibliography 1849-1899*.
  - **NOTE:** The *Sessional Papers* for the 1st Parliament of British Columbia (1872-1875) are included in the *Journals and Sessional Papers of the Province of*

British Columbia[7] and were reprinted in 1980 in Hendrickson's *Journals of Colonial Legislatures of the Colonies of Vancouver Island and British Columbia 1851-1871*. The LLBC has digitized the Journals from 1851 to 2009, excepting 1875 (http://test.archives.leg.bc.ca/). Included in Henrickson's *Journals* are the *Vancouver Island House of Assembly Correspondence Book* (1856-1859), the *Vancouver Island House of Assembly Minutes* (1856-1858), *Vancouver Island Legislative Council Minutes* (1851-1861), and the *British Columbia Legislative Council Journals* 1864-1871. These are listed *in BC Bibliography 1849-* 1899, entries #1643, #1644, #1645 and #224.

➢ Crown Copyright has expired for the first 10 years of the *BC Sessional Papers*. Would need to approach the Legislative Assembly Law Clerk for permission to digitize and make available to the public due to Parliamentary Privilege.

➢ Would be a good proof of concept for a Parliamentary Privilege publication.

- **Government Gazette, British Columbia, 1863-1871 (Including the Government Gazette for Vancouver Island, 1864 - 1866 and the Government Gazette for the Colonies of Vancouver Island and British Columbia, 1859-1860.)**

  ➢ Crown copyright has expired.

  ➢ Potential collaboration opportunities with Queen's Printer.

  ➢ Would be a good proof of concept for expired Crown copyright publications.

  ➢ **Government Gazette, British Columbia, 1863-1871**. Purpose was, and continues to be, to disseminate government proclamations, notices, and regulations as required by provincial legislation. On November 14, 1871, the title was changed to the *British Columbia Gazette*.[8] Issued weekly, bound in annual volumes. Ten volumes exist. Page sizes for early years 15 ¾ x 10 ½; later years 11 7/8 x 7 ½. Four to eight pages per weekly issue were printed.

---

[7]Marjorie C. Holmes, *Publications of the Government of British Columbia, 1871-1947* (Victoria: Government of British Columbia, 1950), 8.

[8]Madge Wolfenden, "The Early Government Gazettes," *British Columbia Historical Quarterly* 7 (1943): 171-190.

Total page count: approximately 4500 single-sided pages (2250 double-sided pages). Listed in the BC Bibliography 1849-1899, entry #187.

➢ **The Government Gazette for Vancouver Island:** 3 vols (1864 – 1866). Available microform UVIC Law. Page count 363 pages. Listed in the *BC Bibliography 1849-1899*, entry #242.

➢ **Government Gazette for the Colonies of Vancouver Island and British Columbia:** Published from 10 September 1859 to 28 August 1860. Page count unknown. Listed in *BC Bibliography 1849-1899*, entry #100**.**

• **The BC Newspaper Card Index**

➢ Provides subject access to major Vancouver and Victoria newspapers published between 1900 and 1990.

➢ 1.1 million cards in good condition with some faded text.

➢ Copyright not a concern, as the LLBC owns the intellectual content.

Note that there are still parliamentary privilege and / or copyright issues to discuss with the appropriate government officials. As a result of these discussions, digitization of one or more of the above publications may not be an outcome of the initial proof of concept project.

Recommendation #1: The Task Group proceed with an initial proof of concept to digitize one or all of the following publications, listed in priority order: the first ten years of the *BC Sessional Papers*; the *Government Gazette, British Columbia,* 1863-1871 (including the *Government Gazette for Vancouver Island*, 1864 – 1866 and the *Government Gazette for the Colonies of Vancouver Island and British Co*lumbia, 1859-1860); and the *BC Newspaper Card Index*.

6.2 Location of Digitization

The University of Victoria (UVIC) would be a logical location for digitizing the above publications; it is equipped with necessary resources and expertise, and it is located close to the LLBC.

Recommendation #2: The digitization for the proof of concept project should take place at the University of Victoria (UVic), with the LLBC providing the publications to digitize.

6.3 Best Practices

The Task Group will need to consult with digital initiative and technical experts at our institutions regarding decisions related to the technical, digitization, and metadata best practices presented in Appendices 7.4 and 7.5 of this proposal.

Potential consultants identified:

• Bronwen Sprout and Allen Bell (UBC)

• Mark Jordon (SFU)

• Chris Petter or Nancy Stewart (UVIC)

Recommendation #3: The Task Group consult with digital initiative and technical experts at our institutions regarding decisions related to the technical, digitization, and metadata best practices presented in Appendices 7.4 and 7.5 of this proposal.

6.4 Outcome

At the end of the proof of concept project, the Task Force will report back to the BCRLG with an evaluation of the success of the project, outstanding issues, and recommendations for future action. The report will include recommendations related to the distribution of workload and resources amongst the five institutions.

6.5 Timeline

We propose the following tentative timeline. Note that this is a non-linear timeline; activities will overlap. The project will begin upon acceptance of this proposal. The scheduling of digitization activities will depend upon other projects underway at UVIC.

Three to six months	• Seek parliamentary privilege permissions for the first ten years of the *Sessional Papers*.
	• Confirm that Crown copyright is not an issue with both the first 10 years of the *Sessional Papers* and the *Government Gazette(s)*.
	• Consult with Queen's Printer about potential collaboration for the

	Government Gazette(s). • Identify storage space for digitized files. • Determine administrative, technical, and preservation metadata schema in consultation with digital and technical experts at our institutions (continuing – to be refined for each title). • Start digitizing the first ten years of the *Sessional Papers*.
Three to six months	• Finish digitizing the first ten years of the *Sessional Papers*. • Digitize *Government Gazette*, 1863-1871 (including the *Government Gazette for Vancouver Island*, 1864 – 1866 and the *Government Gazette for the Colonies of Vancouver Island and British Columbia*, 1859-1860). • Work with digital initiative and technical experts and other collaborators (e.g., the BC Bibliography group) to make digitized files accessible through an available platform.
Three to six months	• Digitize *BC Newspaper Card Index*. • Continue work on making the digitized files accessible through an available platform.
Three to six months	• Report back to the BCRLG.

7 Appendices

7.1 Selection for Digitizing: A Decision-Making Matrix

SChapman, DHazen, JHorrell, JMerrill-Oldham, 5/12/97
©Harvard University Library
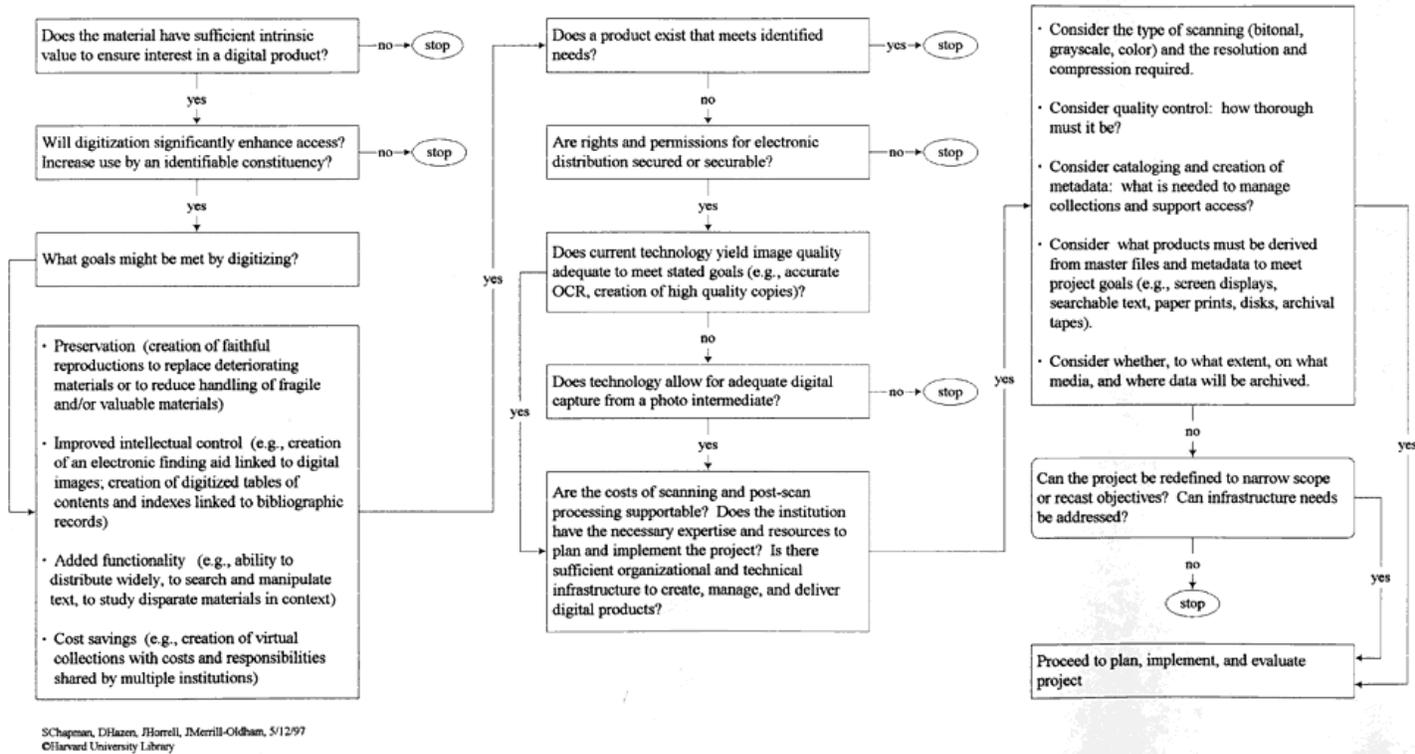
7.2 Potential Documents to Scan: Holdings by Institution

Attached

7.3 Priority Documents for Digitization

Potential Project	Parliamentary Privilege	Crown Copyright	Other Considerations
Sessional Papers, 1876-1982	Yes	Maybe	High priority project. Further consultation with Legislative Assembly Law Clerk required. *Sessional Papers* include Ministry annual reports, *Estimates*, and *Public Accounts* which may be subject to crown copyright (< 50 years).
Decisions of the Speaker of the Legislative Assembly of	Yes	No	Further consultation with Legislative Assembly Law Clerk required.

Potential Project	Parliamentary Privilege	Crown Copyright	Other Considerations
British Columbia, 1877-1943			
British Columbia Gazette, 1863-	No	Yes	Later issues (< 50 years) subject to Crown copyright. Further consultation with Queen's Printer required.
British Columbia Government Gazette, 1863-1871	No	Expired	Expired Crown copyright. Further consultation with Queen's Printer required.
Vancouver Island Government Gazette 1864-1866	No	Expired	Expired Crown copyright. Further consultation with Queen's Printer required.
Public Accounts, 1871- (originals)	No	Yes	Further consultation with the BC Intellectual Property Program required. Crown copyright has expired on pre-1960 publications. *Public Accounts* also contained within *Sessional Papers* (1872-1982). The most recent *Public Accounts* (2000/2001-) are available online through the [LLBC catalogue](#) and the Ministry of Finance website. Therefore, if we digitize the *Sessional Papers*, the highest priority *Public Accounts* to digitize separately would be 1983-1999.
Estimates (originals), 1870/71-	No	Yes	Same considerations as *Public Accounts*. Recent publications are available online are available online through the [LLBC catalogue](#) and the Ministry of Finance website. Therefore, if we digitize the *Sessional Papers*, the highest priority *Estimates* to digitize separately would be 1983-2000.
Dept./Ministry Annual Reports (originals)	No	Yes	Same considerations as *Public Accounts*. Recent publications are available online through the LLBC catalogue, the Ministry of Finance website, and

Potential Project	Parliamentary Privilege	Crown Copyright	Other Considerations
			various Ministry websites. Therefore, if we digitize the *Sessional Papers*, the highest priority annual reports to digitize separately would be older editions of annual reports from after 1982 which are not included available digitally through the LLBC. Potential collaboration with various ministries.
BC Voters Lists	No	Yes	Lists older than 25 years can be digitized.
BC Newspaper Index	No	No	No parliamentary privilege or Crown copyright restrictions. LLBC owns the intellectual content.

7.4 Digitization Guidelines

7.4.1 Processing of Material

- Follow best practices in order to preserve image quality.

- Create a master archival image, which is a faithful reproduction of the original document, and an access version in which all the essential informational value of the original is readable.

- Digitize to the highest resolution appropriate.

- Digitize original (i.e., print rather than microfilm).

- Apply optical character recognition (OCR) when possible.

- Browseable.

- Search capability for at least the index page(s) from each publication, if not the entire text. For example, see the digitized *BC Law Reports* at http://bchistory.library.ubc.ca/?db=bcreports, and the *British Colonist* at http://britishcolonist.ca .

7.4.2   File Formats

- Follow best practices by using standardized open file formats widely used for digital archiving. These include TIFF, PDF/A, JPEG2000, DNG (Digital Negative), and ODF (OpenDocument Format).

7.4.3   Storage and Access

- Consider using open source digital repository platforms.

- Consider whether content should be distributed or hosted on a central server.

- Material is publically accessible via the Internet.

- Select platforms with the following storage and access capabilities.

  - Supports metadata scheme and "crosswalking" between different metadata schemes.

  - Supports OAI-harvesting.

  - Can apply "checksums" or fixity checks for quality control.

  - Supports different levels of access for member institutions and staff.

  - Must easily facilitate the removal of materials upon request.

  - Interface is easy for users to discover, use, and navigate.

  - Capable of providing usage statistics.

  - Supports external OCR applications.

7.4.4   Maintenance and Preservation

- Plan for the long-term maintenance/viability of the collection(s).

- Have a migration strategy that supports the periodic transfer of digital materials from one hardware/software configuration to another or from one generation of computer technology to a subsequent generation.

7.5 Metadata Guidelines

The titles identified for digitization in this proposal are held in our libraries and have MARC cataloging records available. These MARC records contain most of the metadata necessary to identify the digitized versions of titles.

The digitized titles will be hosted at the digital collection repositories of member institutions. These institutions are using a variety of platforms to organize, distribute,

and provide access to digitized objects. The most commonly used applications are *DSPACE* and *CONTENTdm*. Each of these platforms has its own strengths (e.g., images or text), and, depending upon the particular title to be digitized, one may be preferable over the other. Born digital items in such repositories are full-text searchable, but texts that are digitized as images (pdf, tiff, jpeg) often are not, due to the cost of creating OCR files.

Whichever hosting platform is used, metadata additional to that found in the existing MARC records must be created for the digitized versions of identified titles. Most of these titles are serials, so it is likely that metadata for individual volumes, issues, or annuals will have to be created. Metadata allows "harvesting"[9] of the information about the resource by search engines, essentially finding and redisplaying. For example, the Chinese Consolidated Benevolent Association digital collection displayed in the *West Beyond the West* portal links to metadata and images from the SFU ContentDM digital repository. Another example is the Early British Columbia Maps digital collection at UVic. This collection of digital maps and associated metadata is archived in UVic's ContentDM digital repository, and, using OAI-PMH, its metadata is harvested by Google. Finally, a set of environmental maps of Iraq are archived in UVicDSpace. Their metadata is searchable in Google. A goal of this project is the creation of accessible (ie., findable) digital resources.

Two commonly used metadata element sets for digital objects are Dublin Core[10] and MODS. Dublin Core can "crosswalk" to and from MARC. An example of how MARC data fields are "crosswalked" to Dublin Core is shown in Table 1.

---

9. Open Archives Initiative, "Open Archives Initiative Protocol for Metadata Harvesting," Open Archives Initiative, http://www.openarchives.org/pmh/ (accessed October 13, 2010).
10. Dublin Core Metadata Initiative, http://dublincore.org/metadata-basics/.

Table 1: Example of MARC to Dublin Core Crosswalk[11]

MARC Fields	Dublin Core Elements
130, 240, 245, 246	Title
100, 110, 111	Creator
100, 110, 111, 700, 710, 711	Contributor
600, 610, 630, 650, 651, 653	Subject / Keyword
Notes 500, 505, 520, 583	Description
260 $b	Publisher
581, 700 4t, 730, 787, 776	Relationship
008/07-10 260$c	Date

Dublin Core is also the basis for the digital object descriptions in *Alouette Canada*'s and *Multicultural Canada*'s digital collections. Elements of Dublin Core include administrative, descriptive, preservation, technical, and use metadata. We are most familiar with descriptive metadata, as many MARC fields will map to the Dublin Core elements. For this project, administrative, technical, and preservation metadata will need to be added once the schema, Dublin Core or MODS, is selected. See Table 2 for descriptions of this data.

Table 2: Different Types of Metadata and Their Functions[12]

Type	Definition	Examples
Administrative	Metadata used in managing and administering collections and information resources	• Acquisition information • Rights and reproduction tracking • Documentation of legal access requirements • Location information • Selection criteria for digitization
Descriptive	Metadata used to identify and describe collections and related information resources	• Cataloging records • Finding aids • Differentiations between versions • Specialized indexes

---

11. Baca, *Introduction to Metadata*, 41 (adapted).
5. Baca, *Introduction to Metadata*, 9.

Type	Definition	Examples
		• Curatorial information
		• Hyperlinked relationships between resources
		• Annotations by creators and users
Preservation	Metadata related to the preservation management of collections and information resources	• Documentation of physical condition of resources
		• Documentation of actions taken to preserve physical and digital versions of resources, e.g., data refreshing and migration
		• Documentation of any changes occurring during digitization or preservation
Technical	Metadata related to how a system functions or metadata behaves	• Hardware and software documentation
		• Technical digitization information, e.g.,
		• formats, compression ratios, scaling routines
		• Tracking of system response times
		• Authentication and security data, e.g., encryption keys, passwords
Use	Metadata related to the level and type of use of collections and information resources	• Circulation records
		• Physical and digital exhibition records
		• Use and user tracking
		• Content reuse and multiversioning information
		• Search logs
		• Rights metadata

Fifteen elements (see Table 3) are used to define administrative, descriptive, preservation, technical and use metadata. These elements can be qualified by selection of specific values or controlled vocabularies for the expression of each element.

Table 3: Elements Used to Define Metadata

Contributor	Publisher
Coverage	Relation
Creator	Rights
Date	Source
Description	Subject
Format	Title
Identifier	Type
Language	

7.6 Digitization Costs: Examples

Most of the costs associated with digitization projects fall into the activities of planning, set up, and implementation; the physical process of converting print material to digital objects is not as costly. The following examples, based on small-scale digitization projects from our respective institutions, are included to provide some examples of digitization costs. Actual costs will vary depending on equipment; staffing; and the type, format, and amount of material being scanned.

7.6.1   University of Victoria

This example is a cost summary of UVIC's Early BC Maps Project. The principal project costs arose from metadata creation, which was done by higher pay grade staff. Staff time used to organize and set up equipment and to create metadata templates contributed to project costs. Once the equipment and the templates were established, the actual digitization went quickly and involved lower pay grade staff. The end project resulted in 250 maps being digitized. The project involved six staff, including one librarian. The librarian's time of 128 hours to create metadata was not factored into the costs. The average cost worked out to approximately $27 per map.

7.6.2 University of British Columbia

The example from the University of British Columbia (UBC) is based on a fictitious project in which 1,000 pages are scanned, illustrating options of bound and unbound pages. Like the UVIC example, higher costs are incurred from staff time and the preliminary planning of projects. File format and ability to disbind material also affects the cost outcome.

Once the preliminary project work has been completed and the equipment has been chosen and set up by higher pay grade staff (i.e., librarian), lower pay grade staff (i.e., support staff) can assume responsibility for scanning at a cost of approximately $20/hour.

Scanning time depends on whether the Atiz book cradle scanner (bound material) or the Fujitus feeder (disbound material) are used. In addition, the image editing step (adjusting for black and white versus colour, fold-out pages versus pull-out pages, difference in type quality, and whether the text is laid out consistently from page to page) will affect the time needed to produce a final file. The approximate times in Table 4 take these factors into consideration and also include the standard error rate during scanning (e.g., camera loses focus). Please note that, while the first two steps (scanning and editing) need to be closely supervised, the final step of OCR runs automatically, allowing for multi-tasking.

Table 4: Approximate Scanning Times (Hours) by Type of Equipment – 1000 pages in TIFF Format

Equipment	Scanning	Image Editing	OCR/PDF Creation	Total Time
ATIZ Scanner	3.2	2.1	.75	6.05*
Fujitsu Scanner	.5[†]	-	.75	1.25

* For JPG format, total time would be approximately 4.25 hours.

[†] Includes image editing.

Combining the approximate scanning times in Table 4 with the $20/hour staff cost figure provides approximate cost estimates, which are presented in Table 5.

Table 5: Approximate Cost Estimates for Scanning 1,000 Pages by Equipment Type and Format

Equipment	TIFF	JPEG
ATIZ Scanner	$121	$85
Fujitsu Scanner	$25	*

* Not calculated.

Approximate cost estimates for scanning 1,000 pages: $121 using the ATIZ scanner to scan to TIFF file format (6.05 hours @ $20/hour); $85 using the ATIZ scanner to scan to JPG file format (4.25 hours @ $20/hour); and $30 an hour using the Fujitsu feeder to scan to TIFF file format (1.5 hours @ $20/hour).

7.6.3 Simon Fraser University

The costs at Simon Fraser University (SFU) are also dependent on similar criteria to UVIC and UBC (e.g., staffing/equipment set up). The costs work out to $0.50 per page, which generates a high quality TIFF file for a master image and a PDF for general consumption. The PDF is searchable via OCR. The price includes scanning, editing, and OCR but excludes the creation of metadata. Metadata creation is a separate pricing category and is dependent upon whether the cataloguing is original or copy cataloguing.

7.6.4 Legislative Library of British Columbia

Staffing costs at the LLBC for scanning are approximately $0.35 per page. Standard office equipment is used to produce OCR PDFs from bound copies. Existing metadata records are updated; original metadata creation would add to the cost.

7.7 Digitization Equipment Inventory, Member Institutions (All Colour and Black & White Capable)

Scanner Type	Typical use	Make / Model	Institutions	Capacity
Flatbed	Single sheets, bound volumes placed upside down	Epson GT-15000	UVIC, LLBC, UBC	11.7 x 17
		Epson 836XL	SFU	12 x 17 inches
		Microtek Scanner 9800 XL	UNBC	12 x 17 inches
Sheet feed (flatbed with document feeder)	Single sheets, unbound volumes	Epson Expression 10000XL Graphics Arts Scanner	UVIC	12.2 x 17.2 inches

Scanner Type	Typical use	Make / Model	Institutions	Capacity
		Rico IS760D	SFU	11 x 17 inches
		Fuijitsu fi-6770	SFU, UBC	11 x 17 inches
		Xerox DocuMate	UBC	
		Xerox WorkCentre 7328	LLBC	12 x 18 inches
		Rico IS760D	SFU	11 x 17 inches
Book Edge	Bound volumes, placed upside down specially designed edge to prevent spine shadows	Plustek OptiBook 3600 Plus Book Scanner	UVIC	8.5 x 11.7, pdf only
Book Cradle (holds books in a v shaped cradle, face-up, overhead scans using digital cameras).	Bound volumes, 3-D, fragile, oversize, maps	ATIZ Book Drive DIY Book scanner	UVIC, UBC	15.7 x 23.2 inches
Overhead (flat surface, face-up, overhead scans using digital cameras)	Bound volumes, 3-D, fragile, oversize, and single sheet items e.g., maps, posters, architectural drawings	Zeutshel OS10000 TT	SFU	24 x 17 inches , maximum book thickness 4 inches
		TTI Repro-Graphic Workstation 4600/Digifles 45EI	UVIC	24 x 32 inches, no maximum thickness
Wide Format (single sheet feed, cameras NOT overhead)	Large format single sheet items, e.g., maps, architectural drawings, posters.	Vidar scanner	UBC	Handles documents between 24 and 54 inches wide, depending on model
Microfilm/fiche/opaque	Film (16/35), fiche, opaque,	ScanPro 1000	UVIC	Film (16/35mm), fiche, opaque.
Slides	Slide scanner	Nikon Super Coolscan 5000ED with a Nikon SF-210 auto slide feeder	UNBC	35mm slides, 35mm filmstrips,