# Creating a Search Portal for BC Government Publications

VILSC

February 5, 2020

Louise Brittain Boisvert Legislative Library of British Columbia



Philip Wong
Caitlin Samuelsson
Wenchong Wu
Katy Earl
Charles Hogg
Legislative Assembly IT Branch



# Digital Collections

- Collection Policy
- Born digital & digitized documents
- Files stored on Legislative Assembly webserver
- Accessible in catalogue via link in 856 MARC field



# Accessing Monographs



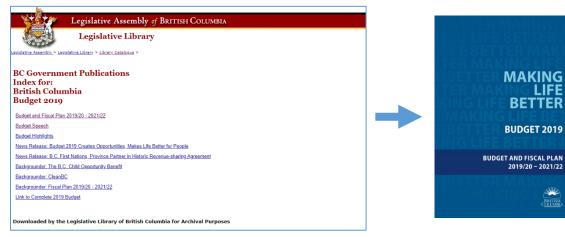








## Accessing Serials



Budget 2019







Budget 1908

#### Problems

- Only top-level titles are catalogued
- Full-text search not available for links on index pages

Example:





#### Goals

 Create a record set that allows for full text indexing and searching of all individual documents.

Implement a discovery interface for searching and access



### Solution



**Scraping Index Pages** 





**Search Engine** 

Components: Apache Solr, PHP, MySQL



#### MARCALIZER

#### Written in PHP

- Queries ILS database for records from our digital collection
- Exports MARC records using Horizon's MARCout Utility and creates a new .mrc file
- Analyzes 856 field in each record to see if multiple items exist 3.
- Traverses through links, creating a new copy of MARC record for each item
- Scrapes index pages, editing various MARC fields as needed: **Examples:**

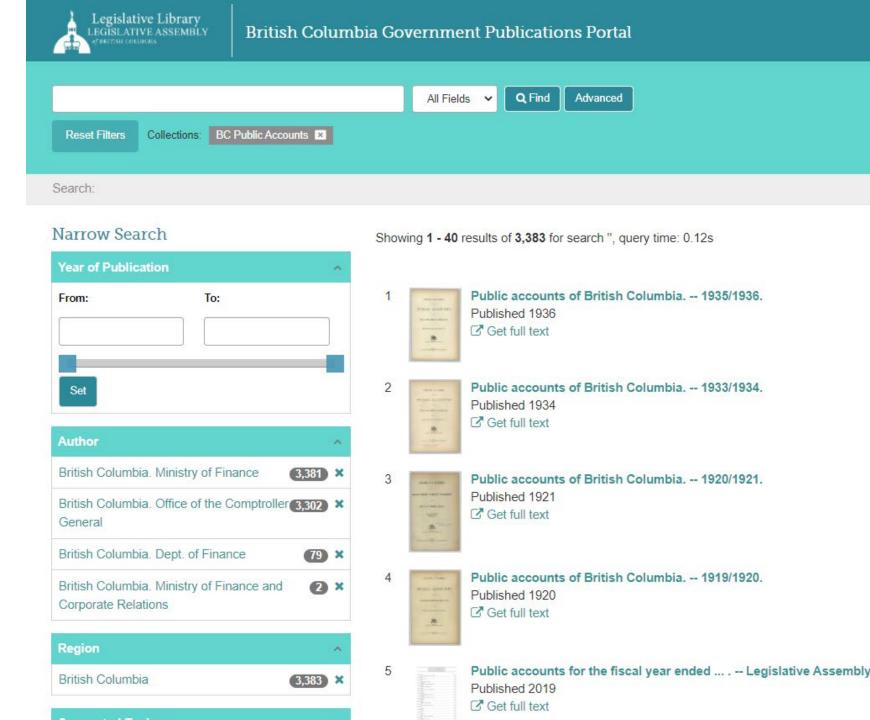
245 \$b add from href tag on index page 260 \$c replace with YYYY from href tag on index page 856 \$a replace with link on index page

- Extracts full text from the links
- Creates thumbnails



#### Results

https://vufind.llbc.leg.bc.ca/ 100,721 item records



# Next Steps

- Incremental Updates and Deletions
- Use Marcalizer for seed data to populate a new database table, which will generate new index pages and eliminate need for scraping
- Long term: move to a content management system/IR?



#### **Questions? Comments? Concerns?**

Louise Brittain Boisvert
Systems Librarian
Legislative Library of British Columbia

louise.boisvert@leg.bc.ca

