New open source tools for collections analysis

Melissa Belvadi, UPEI
May 2024

The Tools

- Perpetual Access Rights - Ebooks tracker
- LC Call Number Harvester
- COUNTER 5.1 Usage Harvester

Common principles:

- open source (MIT license)
- python code, with executables for Windows, Mac (also runs in Linux)
  - no budget to "register" apps, scary warnings
- apps should NOT stray into your OS beyond your own account/app folders
- use sqlite local database to store data
  - what is sqlite?
  - recommend dbeaver to interact directly with them (Win/Mac/Linux)

Ebook rights tracker

- Purpose: provide a storage and search tool for checking if you have perpetual access rights for a given ebook on a given platform
  - Use case: you lose access to a book, and need to verify your rights to make the case to the platform's customer service to regain access
- KBESC work on standard spreadsheet format for CRKN-licensed ebooks
- Desktop app - actually two of them!
- Apps support your own local spreadsheets too (must follow CRKN format)
- Stores data in sqlite local database
- Entirely desktop - no server required
- Both versions written in python
- Both versions offer French or English interface
- Both versions have you choose your institution during first run

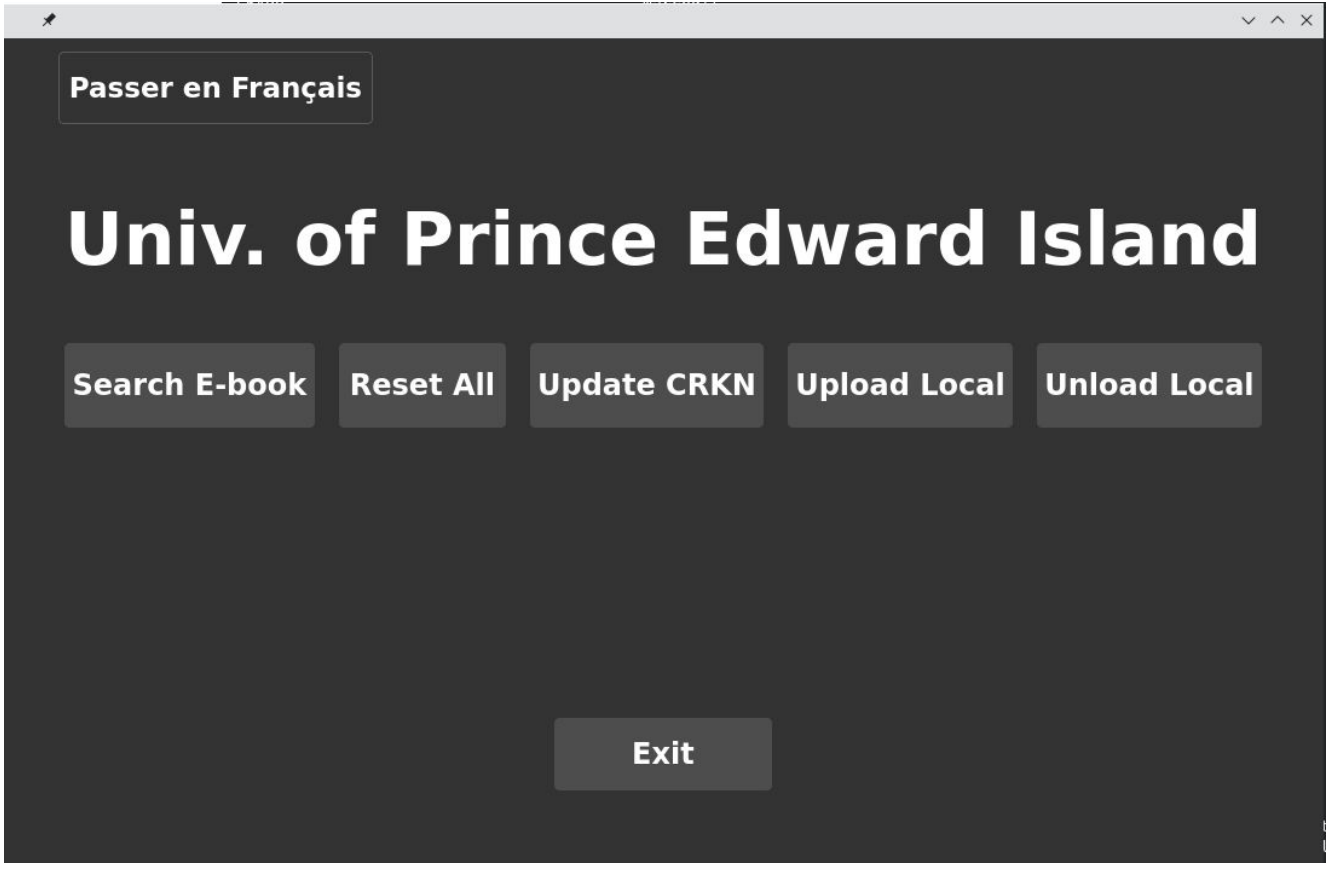
Ebook rights tracker

Spreadsheet format to ingest (KBESC)

	A	B	C	D	E	F	G	H	I
1	**Proquest**					Links to Worldcat: https://worldcat.org/title/OCN			
2									
3	**Title**	**Publisher**	**Platform_YOP**	**Platform_eISBN**	**OCN**	**agreement_code**	**collection_name**	**title_metadata_last_modified**	**Acadia Univ.**
4	Courts and Conflict in Twelfth-Century Tuscany	Oxford University Press	2003	9780199265862	252682095				
5	Strangers in the Night: Law and Medicine in the Managed Care Era	Oxford University Press	2002	9780195152715	654106476				
6	History of the English Language, A	Taylor & Francis Group	2002	9786610281671	1024278292				
7	Wisdom of Egypt: Changing Visions Through the Ages	Taylor & Francis Group	2003	9781843147633	437069447				
8	Mysterious Lands	Taylor & Francis Group	2003	9781843147626	51993921				
9	Views of Ancient Egypt since Napoleon Bonaparte: Imperialism Colonialism and Modern Appropriations	Taylor & Francis Group	2003	9781843147596	1060901571				
10	Ancient Egypt in Africa	Taylor & Francis Group	2003	9781843147589	171579072				
11	New Perspectives on Property Law: Obligations and Restitution	Taylor & Francis Group	2013	9781843145714	54045887				
12	Cases and Materials on the Carriage of Goods By Sea	Taylor & Francis Group	2004	9781843145622	17553834				
13	Statutes And Conventions On International Trade	Taylor & Francis Group	2003	9781843145431	227571318				

Ebook rights tracker app 1

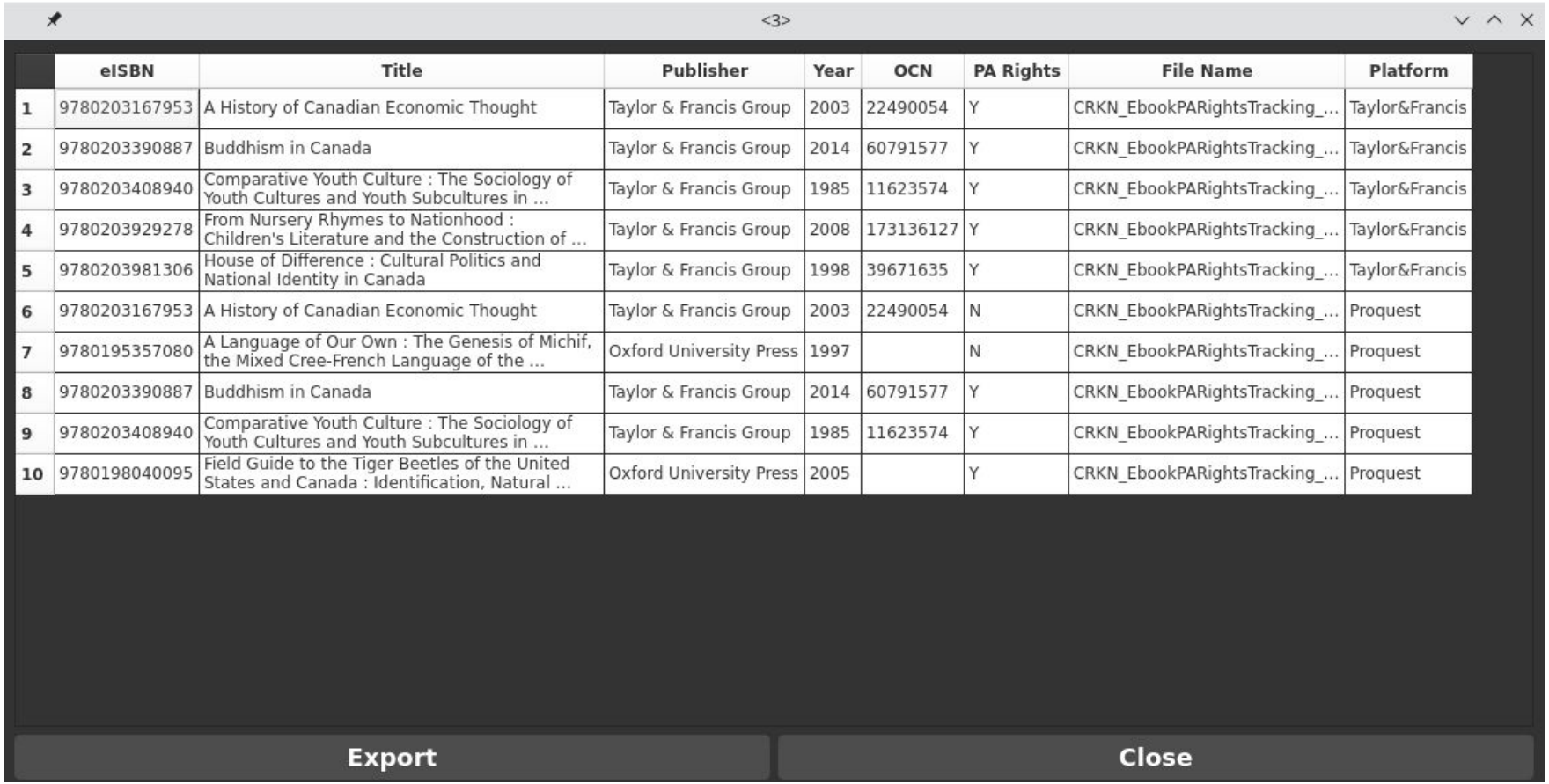
https://github.com/CS-4820-Library-Project/Ebook_Perpetual_Access_Tracking



Ebook rights tracker app 1

Ebook rights tracker app 1

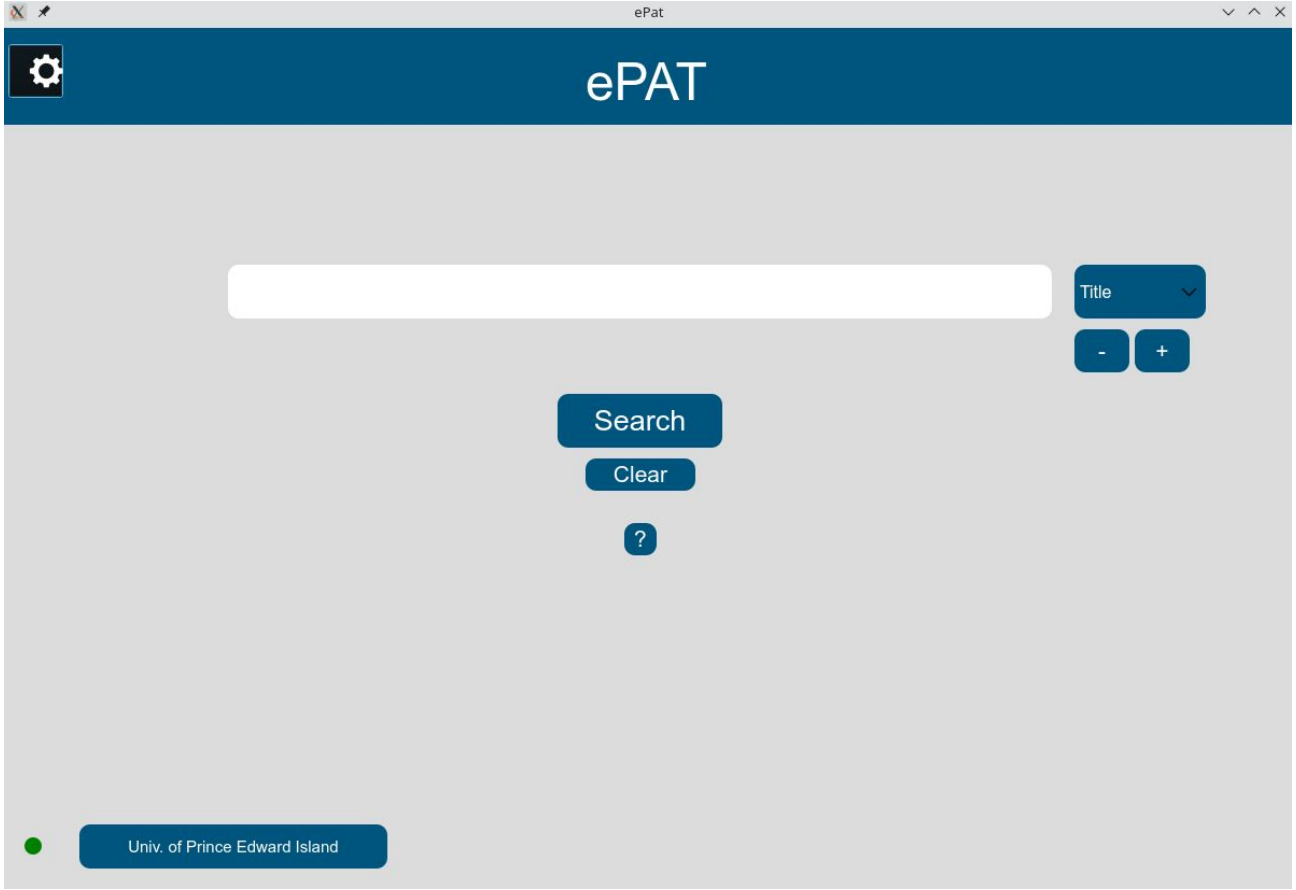
Ebook rights tracker app 1



	eISBN	Title	Publisher	Year	OCN	PA Rights	File Name	Platform
1	9780203167953	A History of Canadian Economic Thought	Taylor & Francis Group	2003	22490054	Y	CRKN_EbookPARightsTracking_...	Taylor&Francis
2	9780203390887	Buddhism in Canada	Taylor & Francis Group	2014	60791577	Y	CRKN_EbookPARightsTracking_...	Taylor&Francis
3	9780203408940	Comparative Youth Culture : The Sociology of Youth Cultures and Youth Subcultures in ...	Taylor & Francis Group	1985	11623574	Y	CRKN_EbookPARightsTracking_...	Taylor&Francis
4	9780203929278	From Nursery Rhymes to Nationhood : Children's Literature and the Construction of ...	Taylor & Francis Group	2008	173136127	Y	CRKN_EbookPARightsTracking_...	Taylor&Francis
5	9780203981306	House of Difference : Cultural Politics and National Identity in Canada	Taylor & Francis Group	1998	39671635	Y	CRKN_EbookPARightsTracking_...	Taylor&Francis
6	9780203167953	A History of Canadian Economic Thought	Taylor & Francis Group	2003	22490054	N	CRKN_EbookPARightsTracking_...	Proquest
7	9780195357080	A Language of Our Own : The Genesis of Michif, the Mixed Cree-French Language of the ...	Oxford University Press	1997		N	CRKN_EbookPARightsTracking_...	Proquest
8	9780203390887	Buddhism in Canada	Taylor & Francis Group	2014	60791577	Y	CRKN_EbookPARightsTracking_...	Proquest
9	9780203408940	Comparative Youth Culture : The Sociology of Youth Cultures and Youth Subcultures in ...	Taylor & Francis Group	1985	11623574	Y	CRKN_EbookPARightsTracking_...	Proquest
10	9780198040095	Field Guide to the Tiger Beetles of the United States and Canada : Identification, Natural ...	Oxford University Press	2005		Y	CRKN_EbookPARightsTracking_...	Proquest

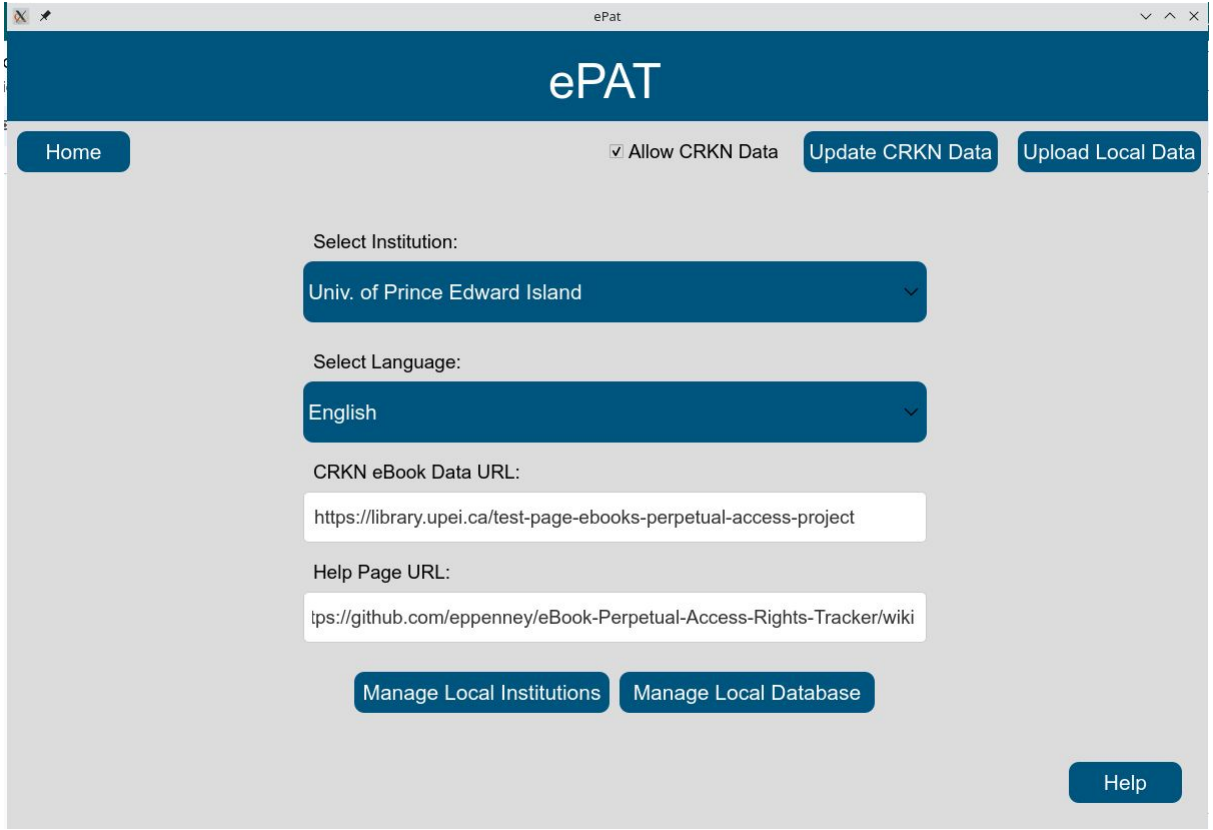
Ebook rights tracker app 2

https://github.com/CS-4820-Library-Project/EPAT_Team8

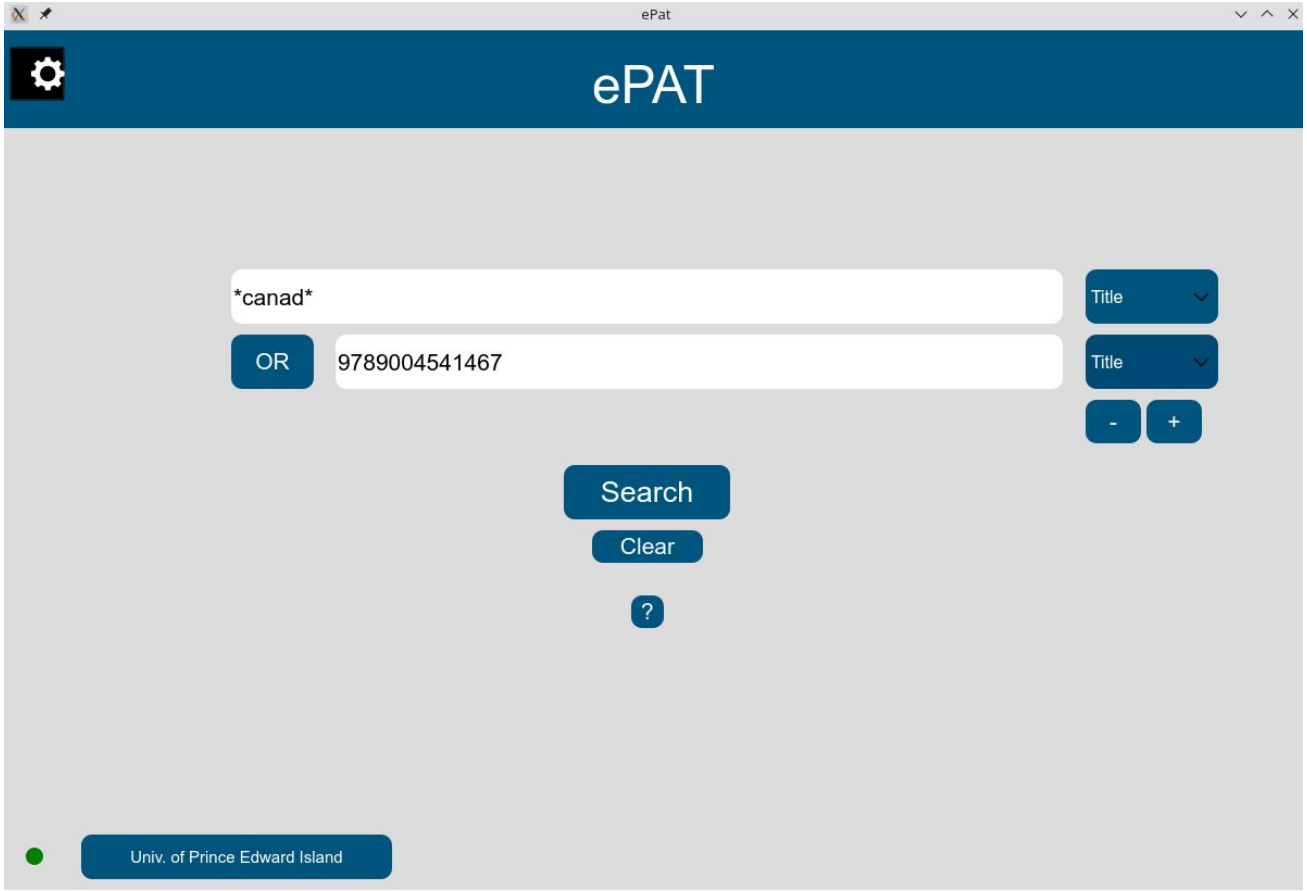


Ebook rights tracker app 2

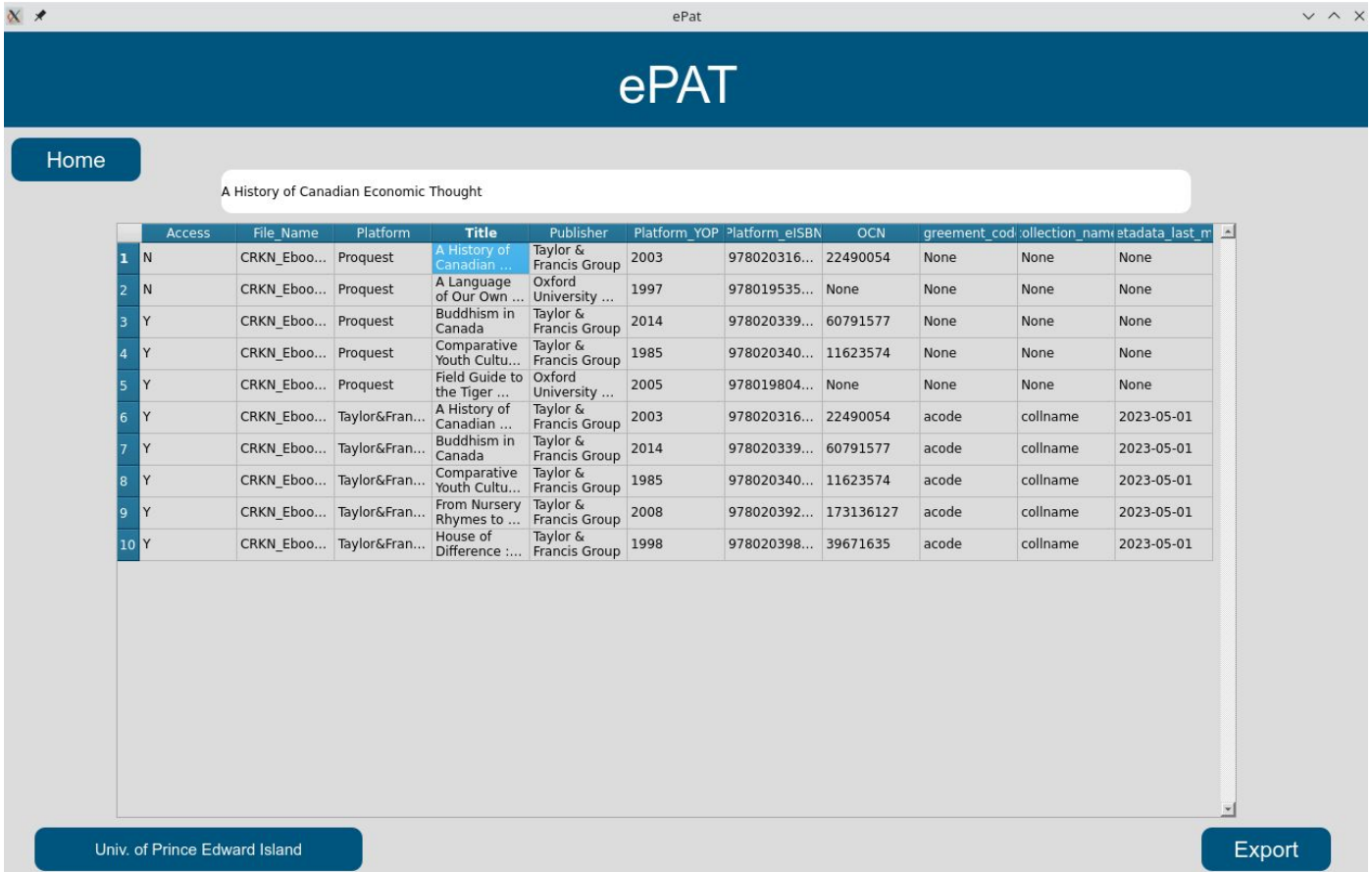
Settings screen



Ebook rights tracker app 2



Ebook rights tracker app 2



	Access	File_Name	Platform	Title	Publisher	Platform_YOP	Platform_eISBN	OCN	greement_cod	ollection_name	etadata_last_m
1	N	CRKN_Eboo...	Proquest	A History of Canadian ...	Taylor & Francis Group	2003	978020316...	22490054	None	None	None
2	N	CRKN_Eboo...	Proquest	A Language of Our Own ...	Oxford University ...	1997	978019535...	None	None	None	None
3	Y	CRKN_Eboo...	Proquest	Buddhism in Canada	Taylor & Francis Group	2014	978020339...	60791577	None	None	None
4	Y	CRKN_Eboo...	Proquest	Comparative Youth Cultu...	Taylor & Francis Group	1985	978020340...	11623574	None	None	None
5	Y	CRKN_Eboo...	Proquest	Field Guide to the Tiger ...	Oxford University ...	2005	978019804...	None	None	None	None
6	Y	CRKN_Eboo...	Taylor&Fran...	A History of Canadian ...	Taylor & Francis Group	2003	978020316...	22490054	acode	collname	2023-05-01
7	Y	CRKN_Eboo...	Taylor&Fran...	Buddhism in Canada	Taylor & Francis Group	2014	978020339...	60791577	acode	collname	2023-05-01
8	Y	CRKN_Eboo...	Taylor&Fran...	Comparative Youth Cultu...	Taylor & Francis Group	1985	978020340...	11623574	acode	collname	2023-05-01
9	Y	CRKN_Eboo...	Taylor&Fran...	From Nursery Rhymes to ...	Taylor & Francis Group	2008	978020392...	173136127	acode	collname	2023-05-01
10	Y	CRKN_Eboo...	Taylor&Fran...	House of Difference :...	Taylor & Francis Group	1998	978020398...	39671635	acode	collname	2023-05-01

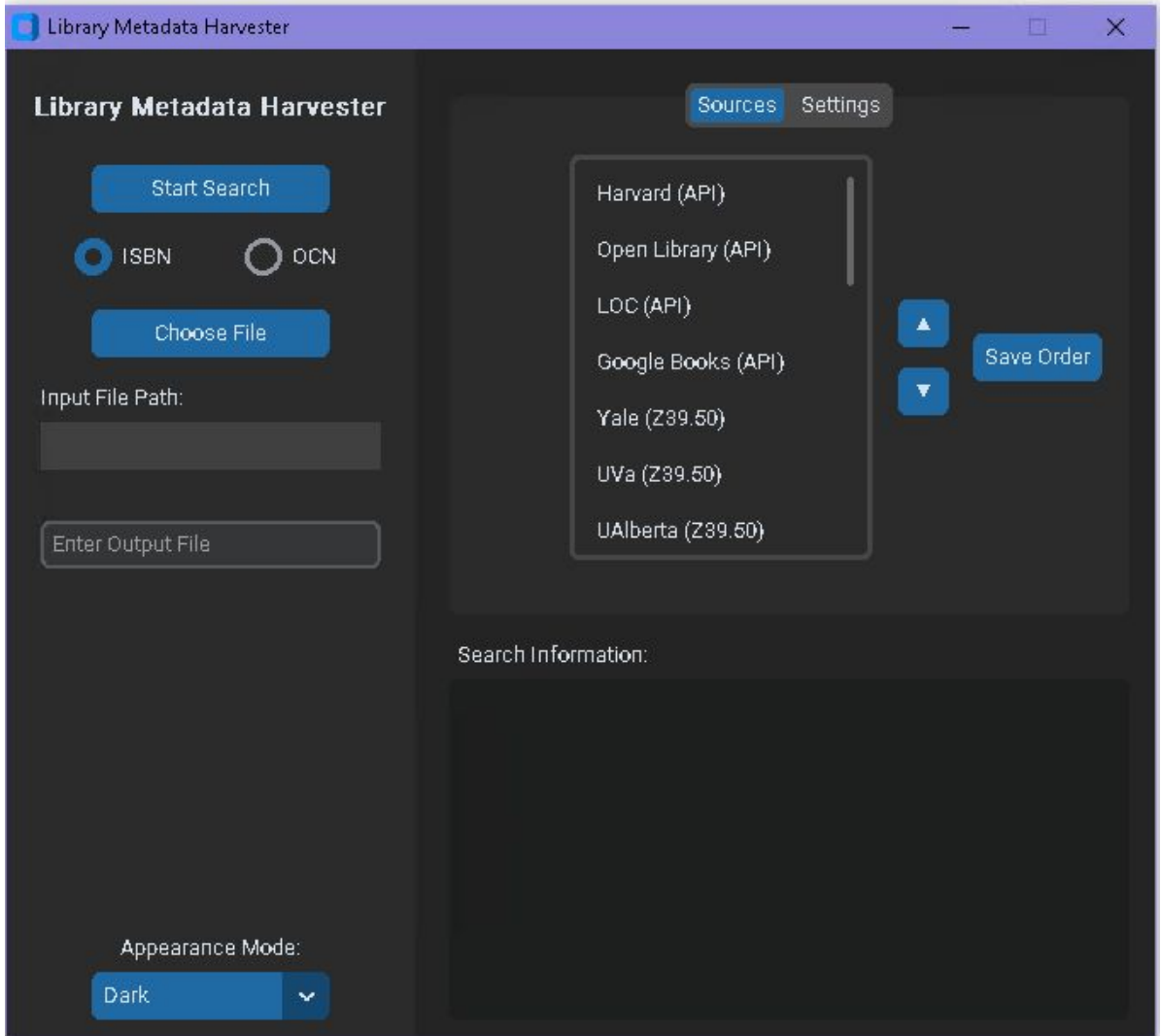
LCC Harvester

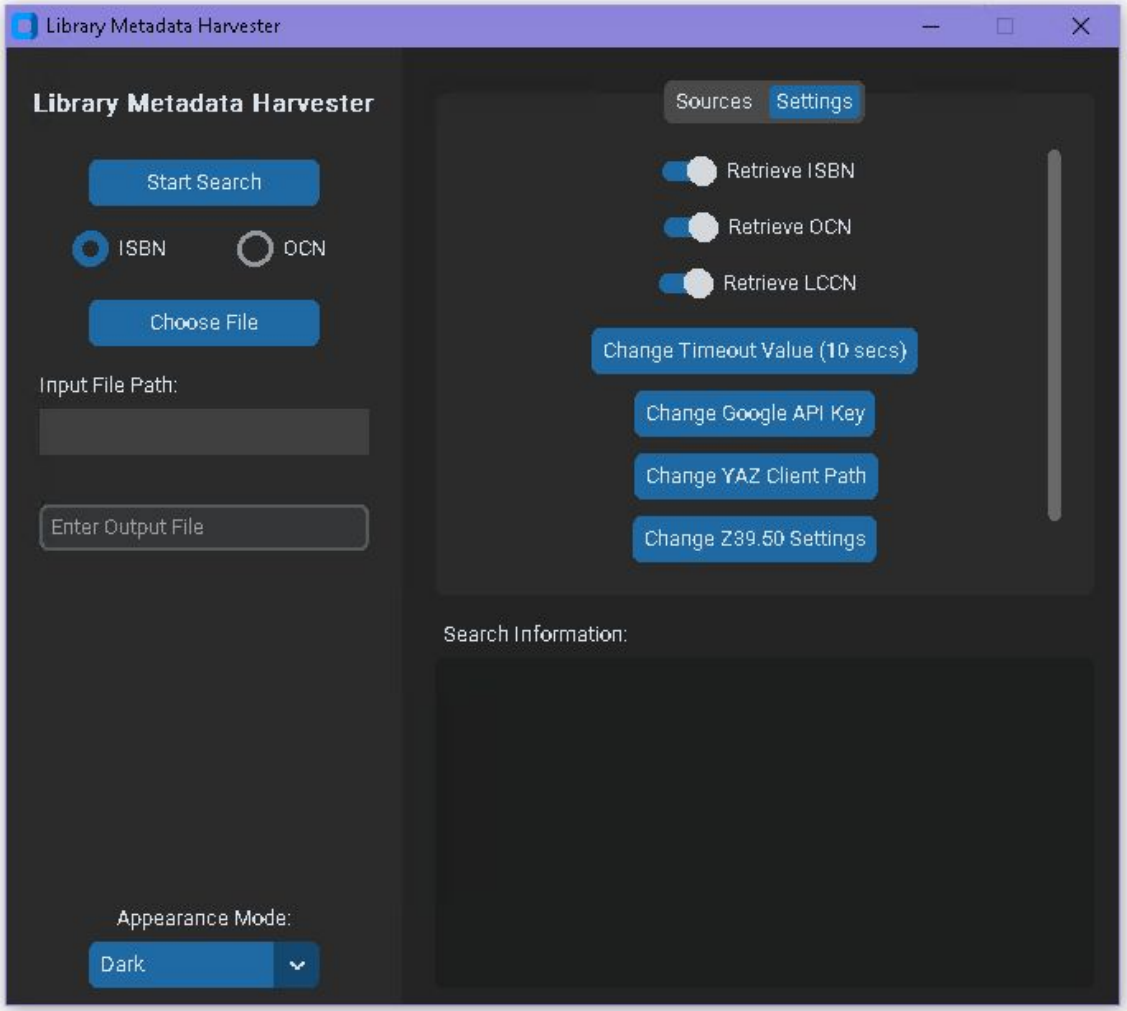
https://github.com/CS-4820-Library-Project/LibraryMetadataHarvester_Team5

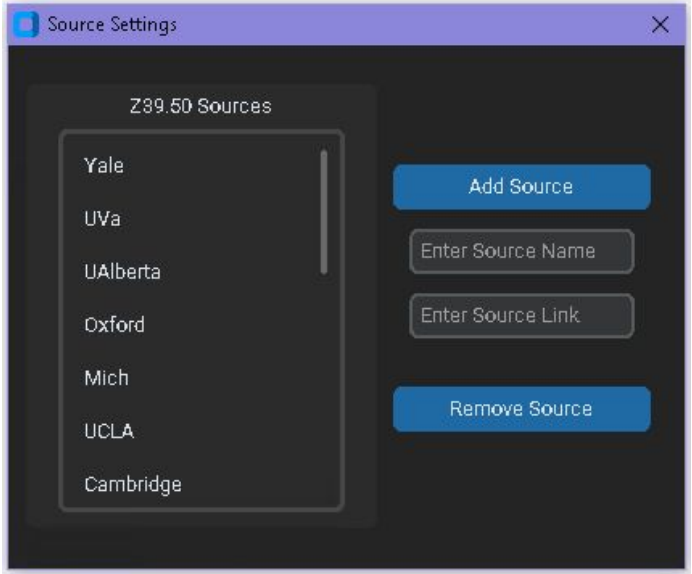
- Purpose: find LC Call Numbers (LCCs) for ebooks for collection analysis
- LCCs = much greater granularity than typical broad subject categories provided by vendors
- LCCs: different platforms offer different subject category schemas in their proprietary reports for you - this gives apples-to-apples comparison
- Given a list of ISBNs, use a combination of APIs and Z39.50 to find an appropriate LCC for use with that book
  - To use the Google API, you need to provide a free API key
  - To use Z39.50, you also need to install the open source "yaz-client"
  - Functionality to webscrape Blacklight catalogues but for legal reasons not populated by default

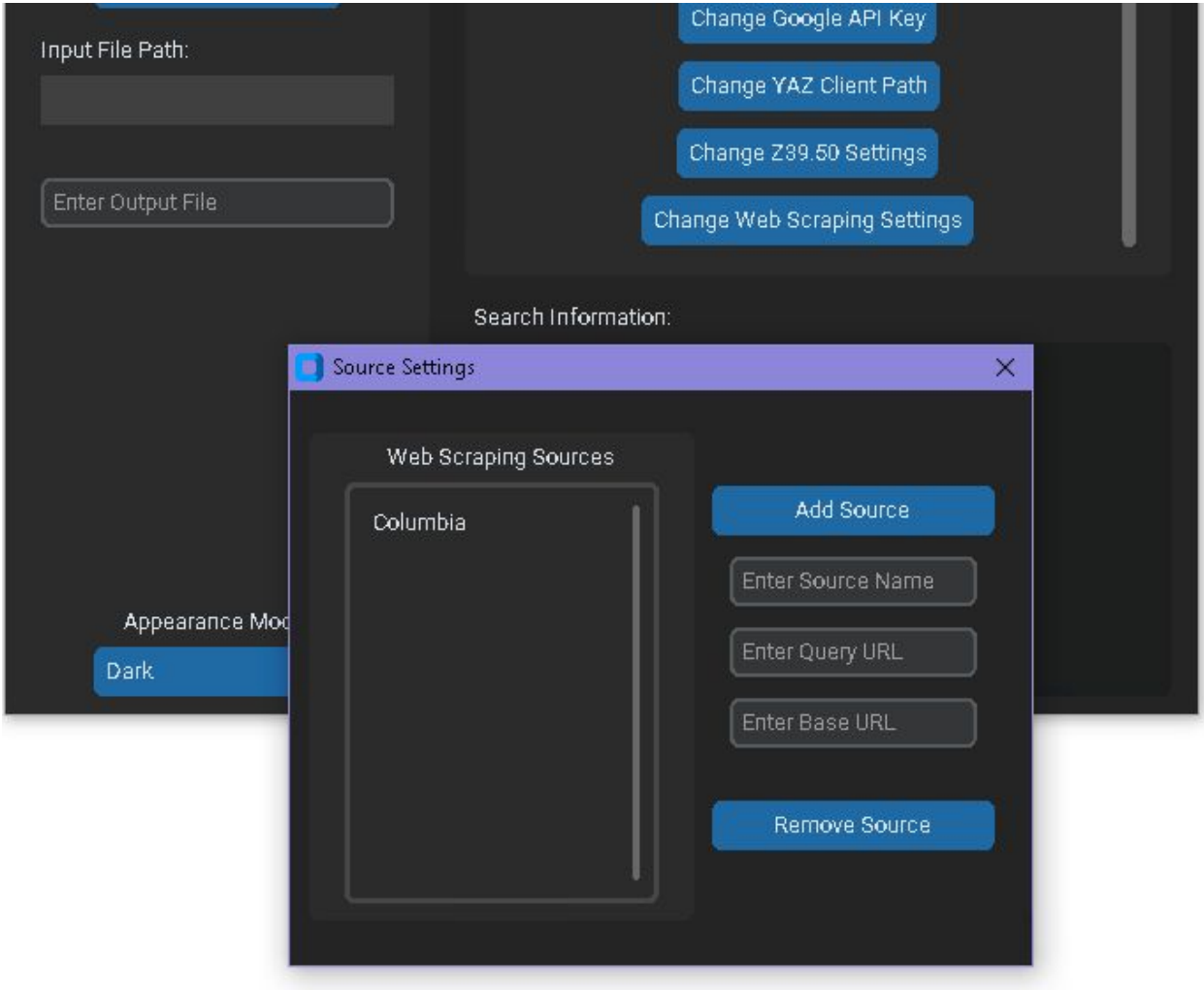
LCC Harvester

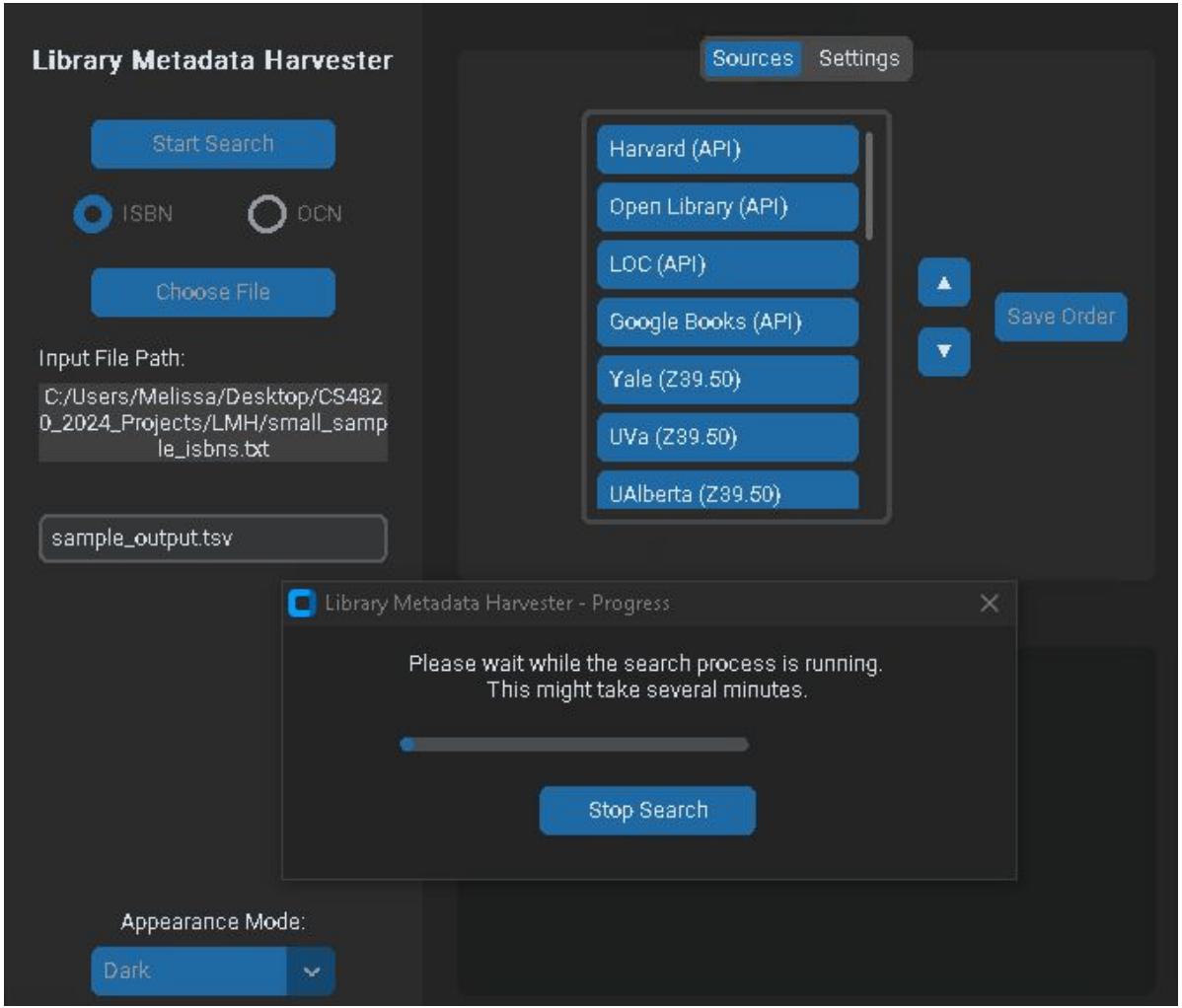
- Stores data in sqlite database, and uses first
- You choose which sources to use, prioritize
- Outputs the results into a tsv (tab-delimited) file suitable for combining with other spreadsheets (eg COUNTER usage reports)
- You can add additional Z39.50 targets and Blacklight OPAC webscrape targets
- Disclaimer - some ISBNs can't be matched to an LC from any of your "targets" - may have to find some manually
- Brief explanation of why OCLC API not included
    - App also offers OCN options (use, find) but limited because of that API problem
- Contact me for "seed" data from previous similar projects (too huge for github)


Library Metadata Harvester
Library Metadata Harvester
Start Search
ISBN
OCN
Choose File
Input File Path:
Enter Output File
Appearance Mode:
Dark
Sources
Settings
Harvard (API)
Open Library (API)
LOC (API)
Google Books (API)
Yale (Z39.50)
UVa (Z39.50)
UAlberta (Z39.50)
Save Order
Search Information:


Library Metadata Harvester
Library Metadata Harvester
Start Search
ISBN
OCN
Choose File
Input File Path:
Enter Output File
Appearance Mode:
Dark
Sources
Settings
Retrieve ISBN
Retrieve OCN
Retrieve LCCN
Change Timeout Value (10 secs)
Change Google API Key
Change YAZ Client Path
Change Z39.50 Settings
Search Information:


Source Settings
Z39.50 Sources
Yale
UVa
UAlberta
Oxford
Mich
UCLA
Cambridge
Add Source
Enter Source Name
Enter Source Link
Remove Source


Input File Path:
Enter Output File
Change Google API Key
Change YAZ Client Path
Change Z39.50 Settings
Change Web Scraping Settings
Search Information:
Appearance Mod
Dark
Source Settings
Web Scraping Sources
Columbia
Add Source
Enter Source Name
Enter Query URL
Enter Base URL
Remove Source



Output file example (tsv)

	A	B	C	D
1	ISBN	OCLC	LCCN	LCCN-Source
2	9780191513015		PJ7942.I27 A53 2020	Columbia
3	9780191513022		PJ7942.I27 A53 2020	Columbia
4	9780191513039		PJ7942.I27 A53 2020	Columbia
5	9780191513183		PJ7942.I27 A53 2020	Columbia
6	9780191513213		PJ7942.I27 A53 2020	Columbia
7	9780191513275		PN56.I465 B644 2005	Harvard
8	9780191513329		PJ7942.I27 A53 2020	Columbia
9	9780191513336		DS54.86 .H65 1998	Harvard
10	9780191513428		DD120.F8 H48 2000	Harvard
11	9780191513435		BX1492 .W66 2000	Harvard
12	9780191513459		PJ7942.I27 A53 2020	Columbia
13	9780191513466		PJ7942.I27 A53 2020	Columbia
14	9780191513473		KD671 .H63 2005	Harvard
15	9780191513480		PJ7942.I27 A53 2020	Columbia
16	9780191513503		BL1575 .H56 2005	Harvard
17	9780191513510		BR115.P7 S6324 1997	Harvard
18	9780191513572		SB975 .W3 2005	Harvard
19	9780191513589		PJ7942.I27 A53 2020	Columbia

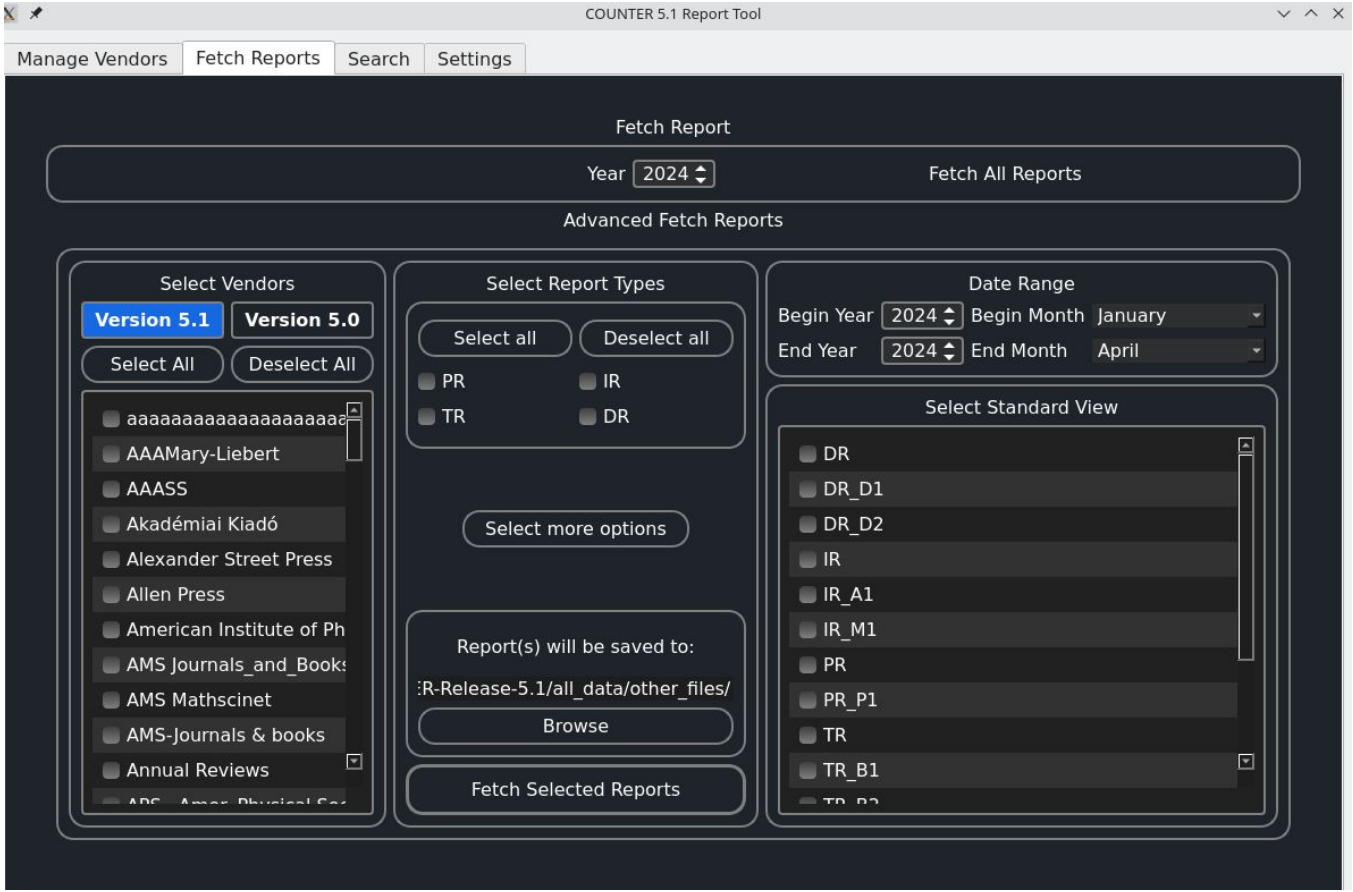
COUNTER 5.1 Usage Harvester

- Final version delayed - expected done by end of summer
  - 5.1 compatibility problem - no live vendor servers to test against yet
  - Expect 5.1 COUNTER reports starting January 2025
  - App handles 5.0 and 5.1 data
- Desktop app uses COUNTER Metrics API (formerly "SUSHI")
- Retrieves main reports, standard views, custom main reports (filters & attributes)
- Generates tsv files exactly according to COUNTER report standards
- Simple title/issn/isbn search included, outputs tsv of all matching data across all platforms
- Stores data in an sqlite database that combines across platforms for more complex queries
- Also retrieves Title Reports with the YOP (Year of Publication) attribute expanded
- Useful tools for managing your vendor credentials data, and basic password option for the vendor table
- Provides log file for troubleshooting, reporting issues to COUNTER/vendor

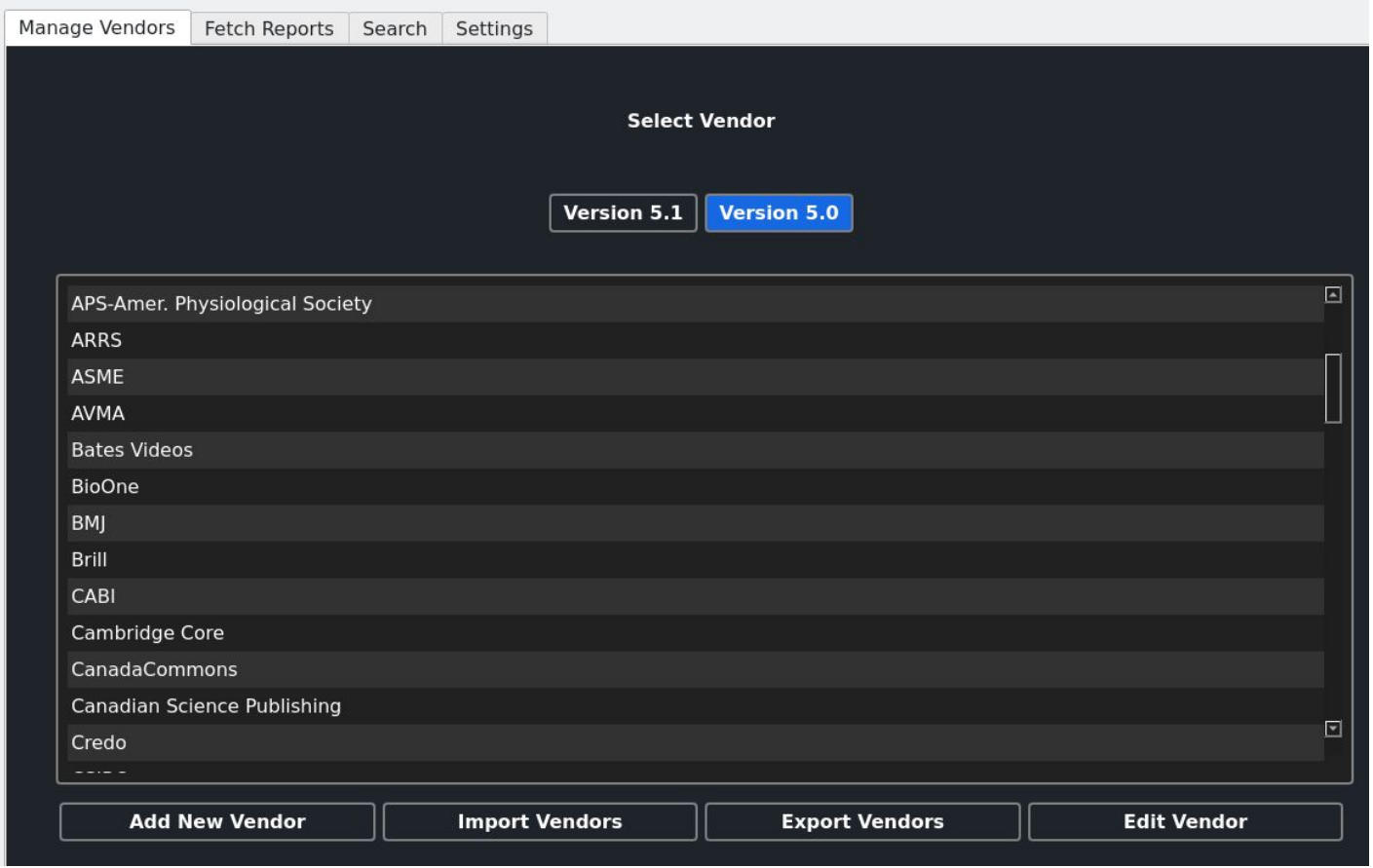
COUNTER 5.1 official docs: https://cop5.projectcounter.org/en/5.1/

COUNTER official website (new!): https://www.countermetrics.org/

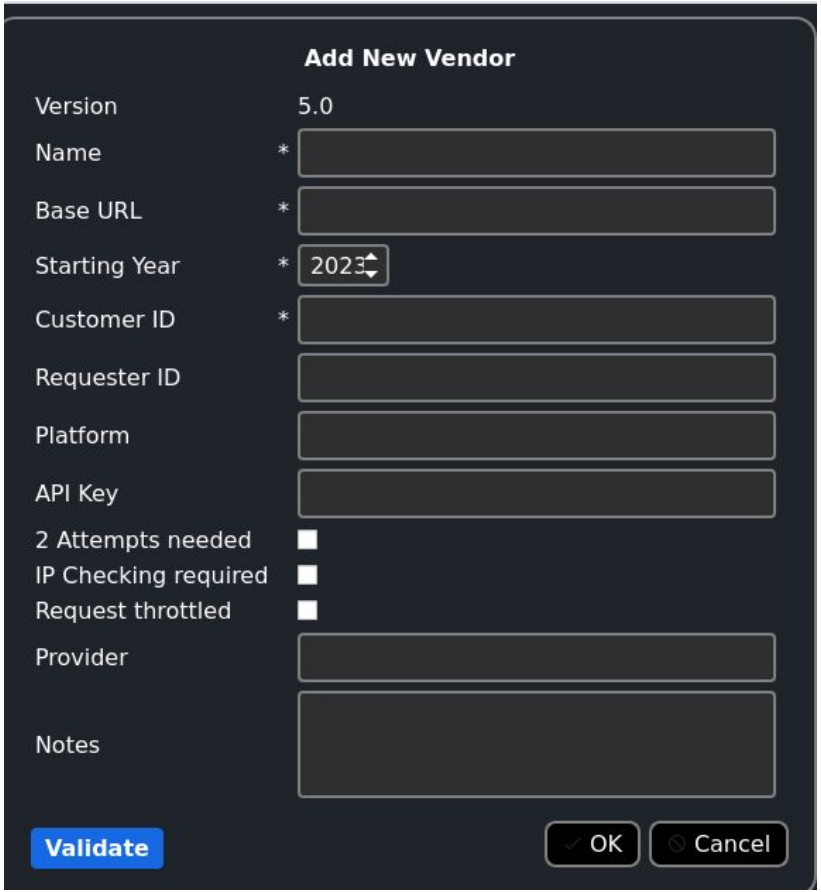
COUNTER 5.1 Usage Harvester



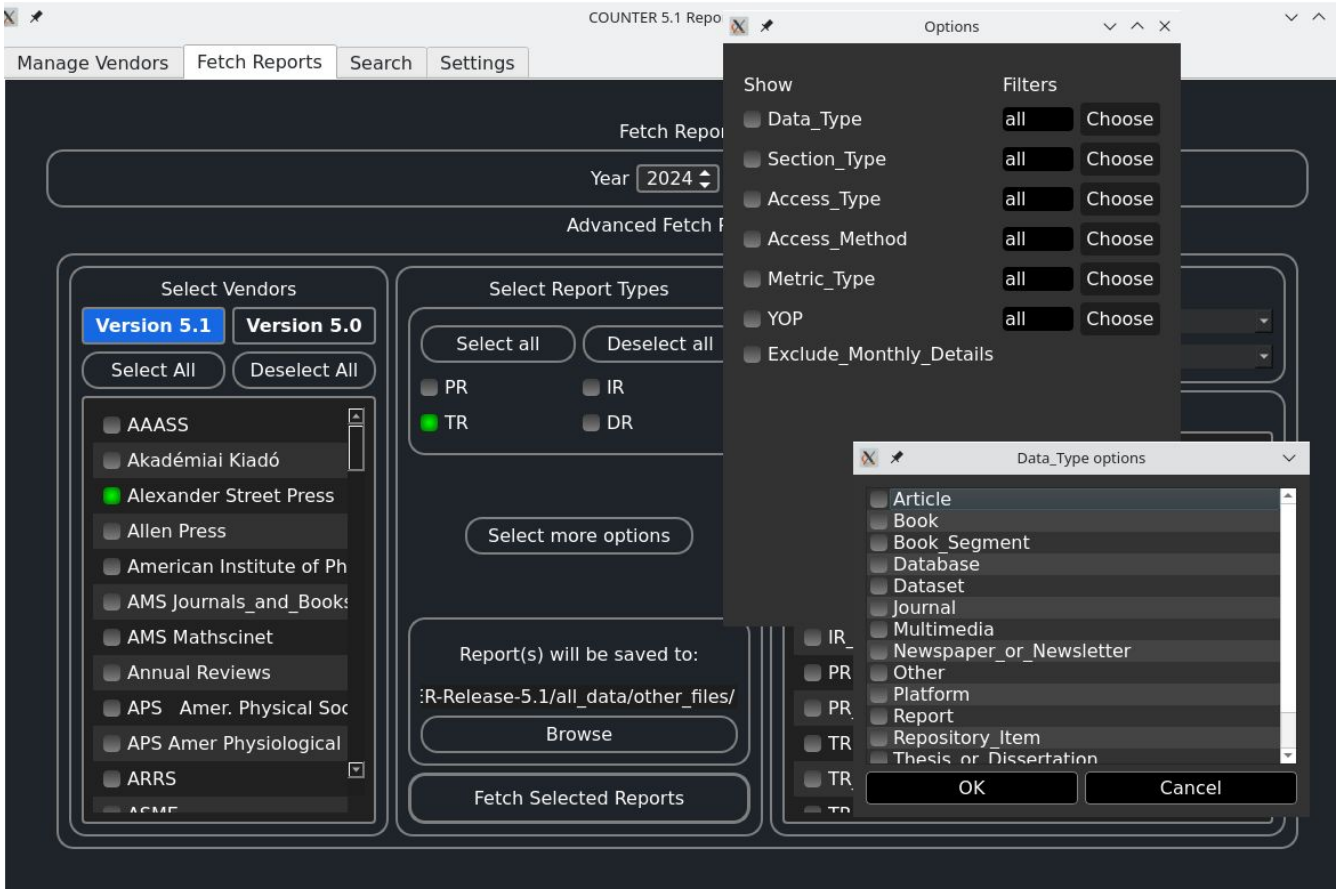
COUNTER 5.1 Usage Harvester



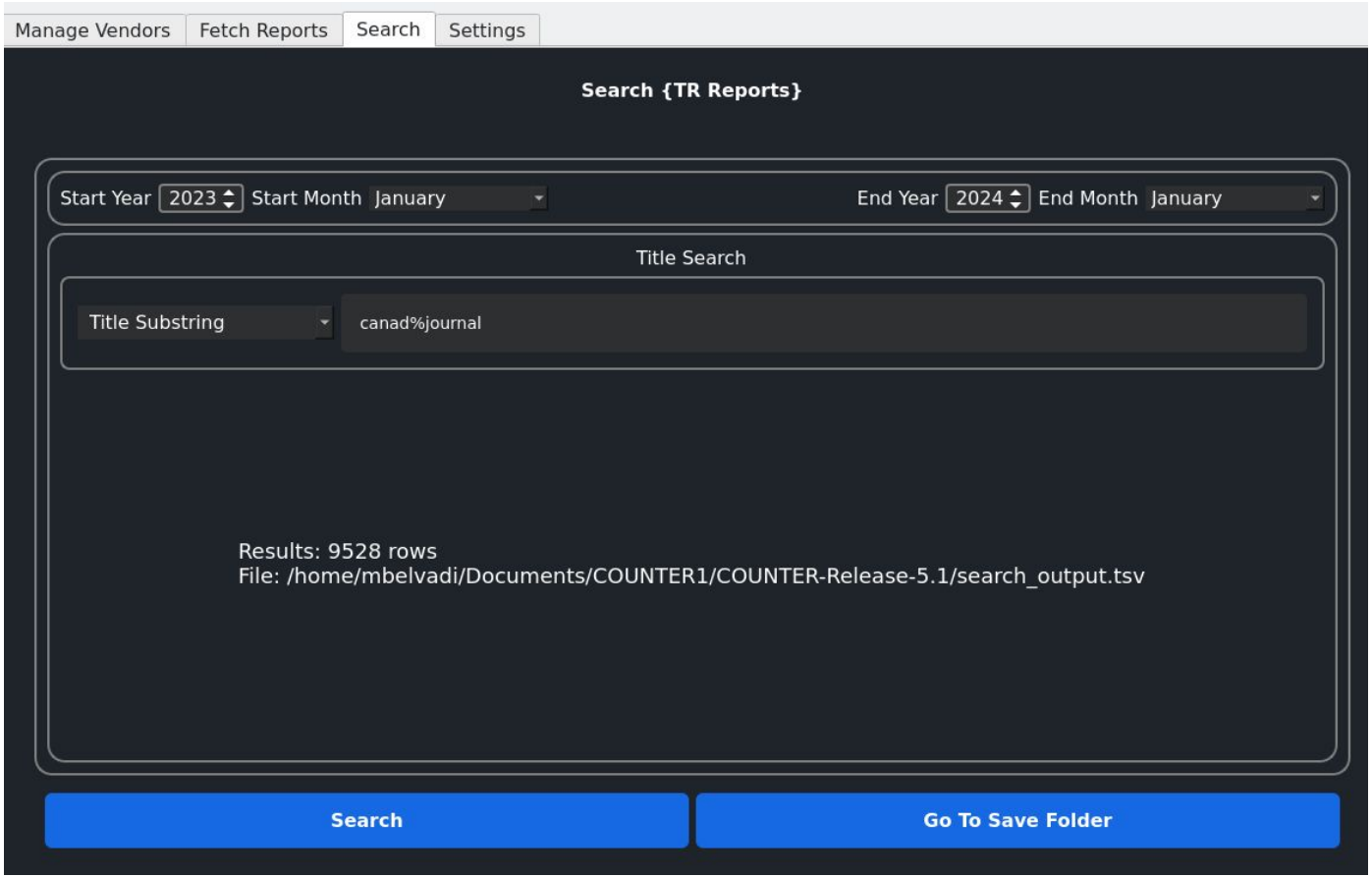
COUNTER 5.1 Usage Harvester



COUNTER 5.1 Usage Harvester



COUNTER 5.1 Usage Harvester



COUNTER 5.1 Usage Harvester

More use situations:

- EBA spendouts - when you have that book in an EBA and also other platforms, useful to see total use of that title
- Subscription title removals (EBSCO, Proquest), considering purchasing before lost access - can check usage across both subscriptions (esp. if book is being lost from both)
- DR data: Determine user preferences for platforms (databases) including discovery layer vs native database platforms
- Similar for journals in different sub packages, publisher sites

Using the tools together

How you might use combinations of these for collections analysis

Scenario 1: see if you have a lot of book turnaways in certain subject areas

- Use the COUNTER harvester to pull all of your TR_B2 "denials" data over whatever usage time period is of interest
- Use a simple sqlite query to pull the ISBNS into a simple text file (and do again to create a tsv file with additional info, eg titles, YOP, publisher, platform, etc.)
- Run the ISBN list against the LCC Harvester to create the ISBN-LCC output file
- Use Excel/etc. to put both output files into a spreadsheet then "vlookup" to join the denials data with the LCCs
- Use pivot table or your other favorite spreadsheet tools to produce a useful summary

Using the tools together

Scenario 2: analyze your perpetual access holdings by subject areas for gaps

- Use one of the Ebook tools to get all of your PAR data into its sqlite database across all platforms
- Use a simple sqlite query to pull the ISBNS into a simple text file
- Run the ISBN list against the LCC Harvester to create the ISBN-LCC output file
- Use Excel/Google Sheets/LibreOffice to put both output files into a spreadsheet then "vlookup" to join the PAR data with the LCCs
- Use pivot table or your other favorite spreadsheet tools to produce a useful summary

Using the tools together

How could this help your collections analysis (or even information literacy instruction)?

- If the data shows your collection leans on non-PAR titles (EBA, DDA, subscription) heavily in certain subject areas with high use, maybe increase budget for e-books in those areas? Or the reverse? Add more funds to purchase more PAR titles in areas heavily used?
- Are there patterns by subject and usage across different vendors/publishers that you purchase PAR on? Might reflect librarian choices OR patron choices. For instance, if psychology users tend to use the ebooks you purchase or license on Vendor 1 much more per-title than Vendor 2, might reflect information literacy training experiences emphasizing certain platforms, or faculty recommendations to students, or even quality of content for your program by publisher/vendor?

Using the tools together

How could this help your collections analysis (or even information literacy instruction)?

- Compare big ebook subscription packages by usage/subjects (for dept-budget cost-sharing?)
- COUNTER data includes publisher for books on aggregate platforms, so you can evaluate for a given publisher (again by subject if needed) if usage is better on the "native" publisher platform vs various aggregators.
    - eg. Should I buy my PAR for Springer books on EBSCO/Proquest or Springer's own?

Can you think of more uses? Let your imagination run wild! Share your ideas!

Question about future projects

Would it be useful to have an app that could ingest COUNTER data reports you retrieved manually into an sqlite database rather than using this API harvester?

Would it be useful to have a single desktop app that combined these three, or any two of them?

Would it be useful to look at combining Google Analytics or other proprietary vendor data (eg searches by title/issn in your A-Z list) with your COUNTER data into a single sqlite database? Specific ones, eg EBSCO and Proquest have detailed proprietary usage reports? KBART files?