

Nouveaux outils open source pour l'analyse des collections

Melissa Belvadi, UPEI
Mai 2024

Les outils

- Droits d'accès perpétuels - Ebooks tracker
- LC Numéro d'appel Moissonneur
- COMPTEUR 5.1 Moissonneur d'usage

Principes communs :

- open source (licence MIT)
- code python, avec des exécutables pour Windows, Mac (fonctionne également sous Linux)
 - pas de budget pour "enregistrer" les applications, avertissements effrayants
- les applications ne doivent PAS s'aventurer dans votre système d'exploitation au-delà de vos propres dossiers de compte/d'application
- utiliser la base de données locale sqlite pour stocker les données
 - qu'est-ce que sqlite ?
 - recommander dbeaver pour interagir directement avec eux (Win/Mac/Linux)

Suivi des droits sur

- Objectif : fournir un outil de stockage et de recherche pour vérifier si vous avez des droits d'accès perpétuels pour un livre électronique donné sur une plate-forme donnée.
 - Cas d'utilisation : vous perdez l'accès à un livre et vous avez besoin de vérifier vos droits pour faire valoir vos arguments auprès du service clientèle de la plateforme afin d'en récupérer l'accès.
- Travail du KBESC sur un format standard de feuille de calcul pour les livres électroniques sous licence du RCDR
- Application de bureau - en fait, il y en a deux !
- Les applications prennent également en charge vos propres feuilles de calcul locales (elles doivent respecter le format CRKN).
- Stocke les données dans une base de données locale sqlite
- Entièrement bureautique - aucun serveur n'est nécessaire

Suivi des droits sur

- Les deux versions sont écrites en python
- Les deux versions offrent une interface en français ou en anglais
- Dans les deux versions, vous pouvez choisir votre institution lors de la première manche.

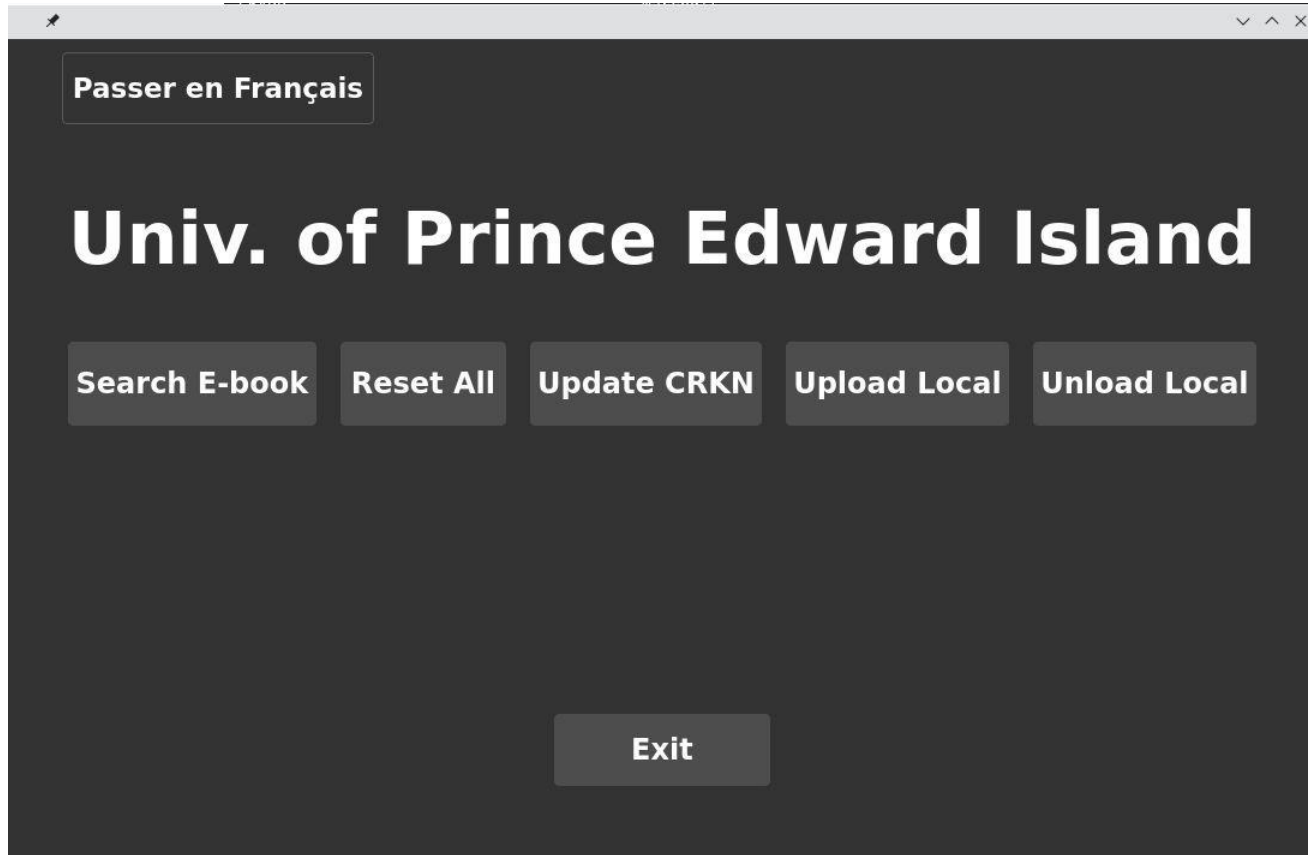
Suivi des droits sur

Format de feuille de calcul à ingérer (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_etSBN	OCN	agreement_code	collection_name	title_metadata_ ast_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

E-Book Search

canad

Search

Cancel

OCN

Title

eISBN

Ebook rights tracker app 1

E-Book Search

canad

Search

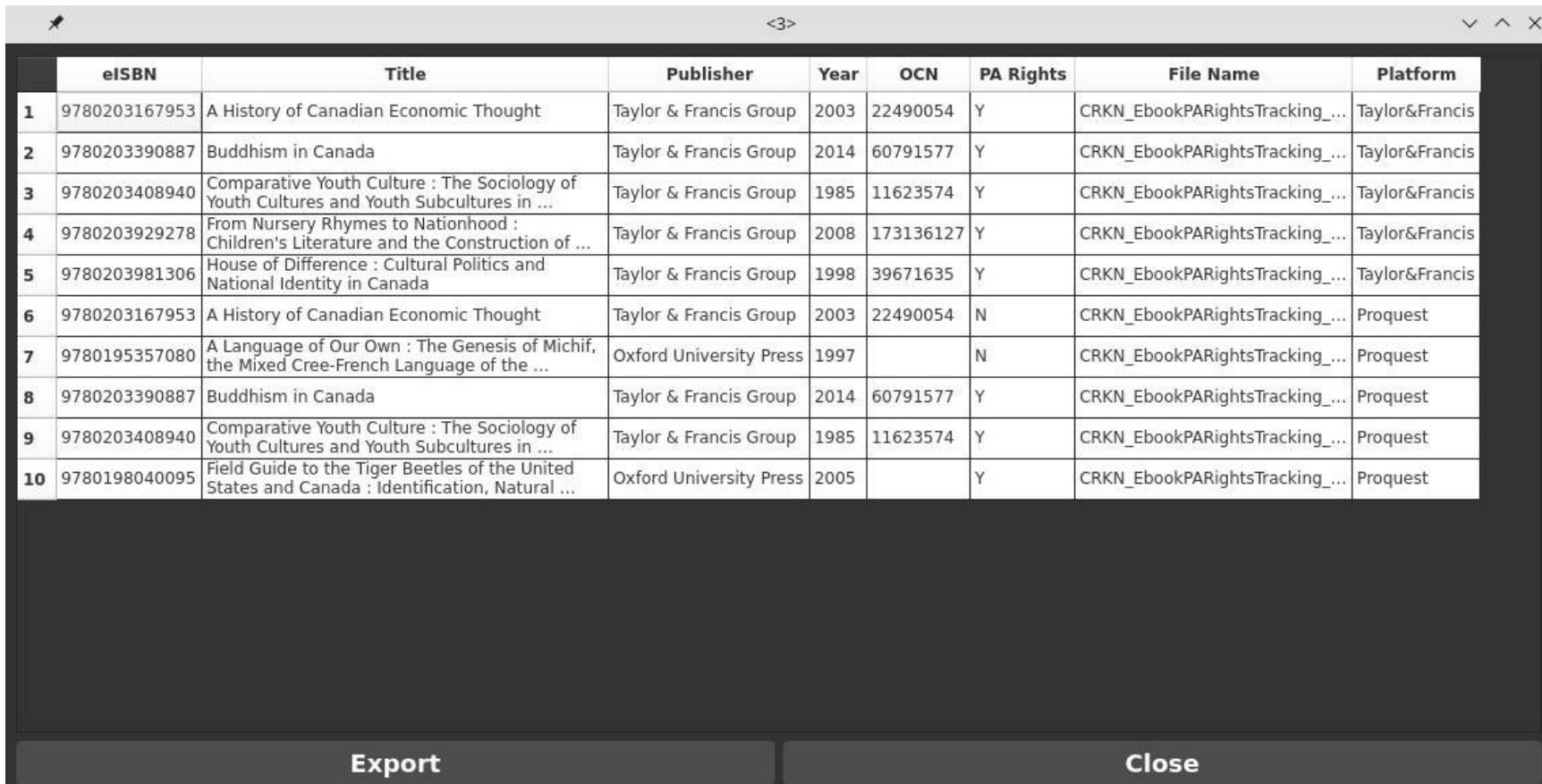
Cancel

OCN

Title

eISBN

Ebook rights tracker app 1

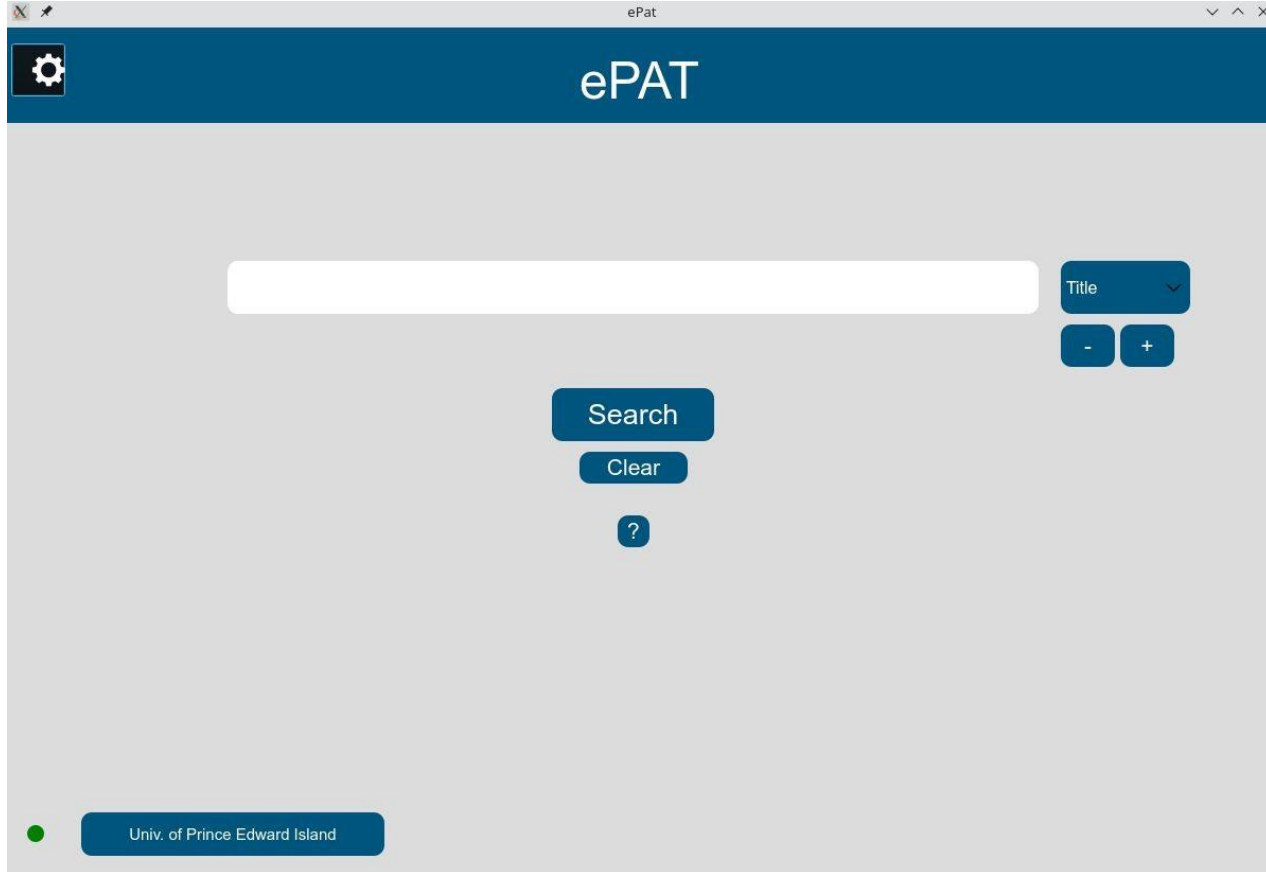


The screenshot shows a software window titled "Ebook rights tracker app 1". The window contains a table with 10 rows of book data. At the bottom of the window, there are two buttons: "Export" and "Close".

	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

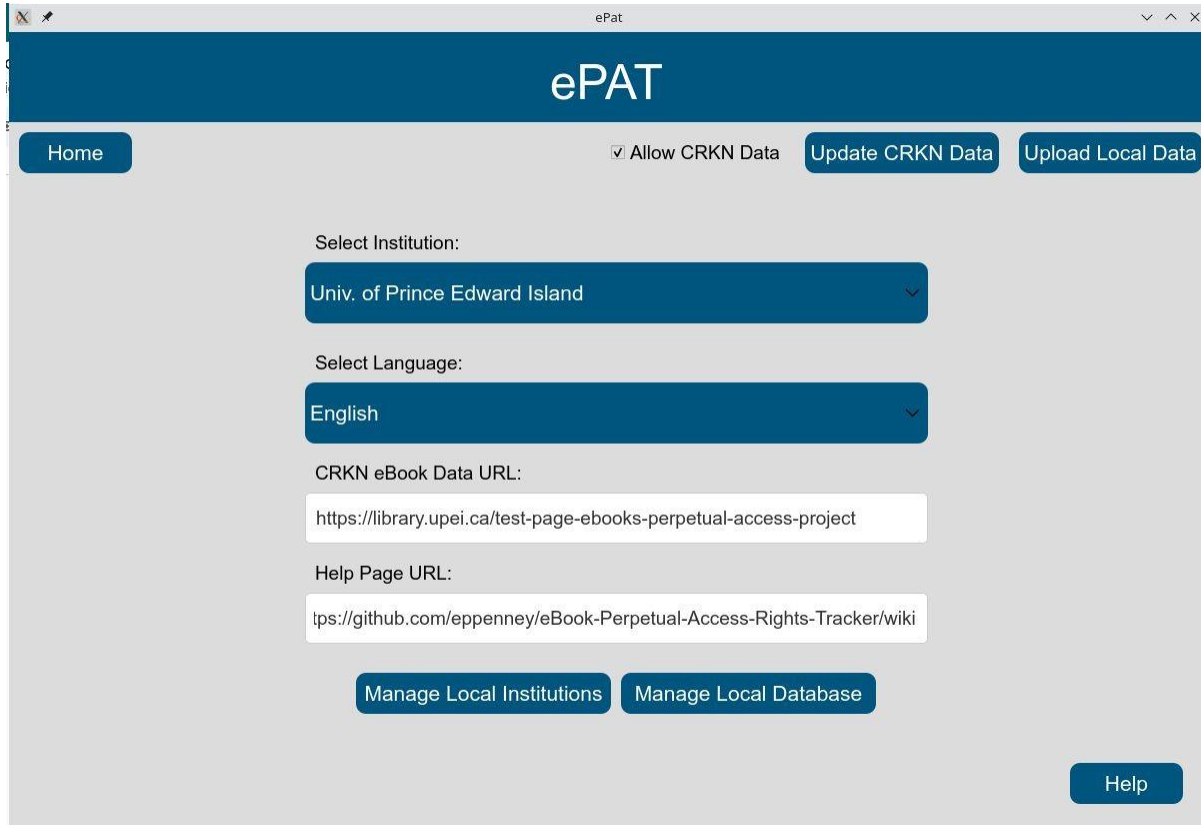
Ebook rights tracker app 2

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



Ebook rights tracker app 2

Écran des paramètres

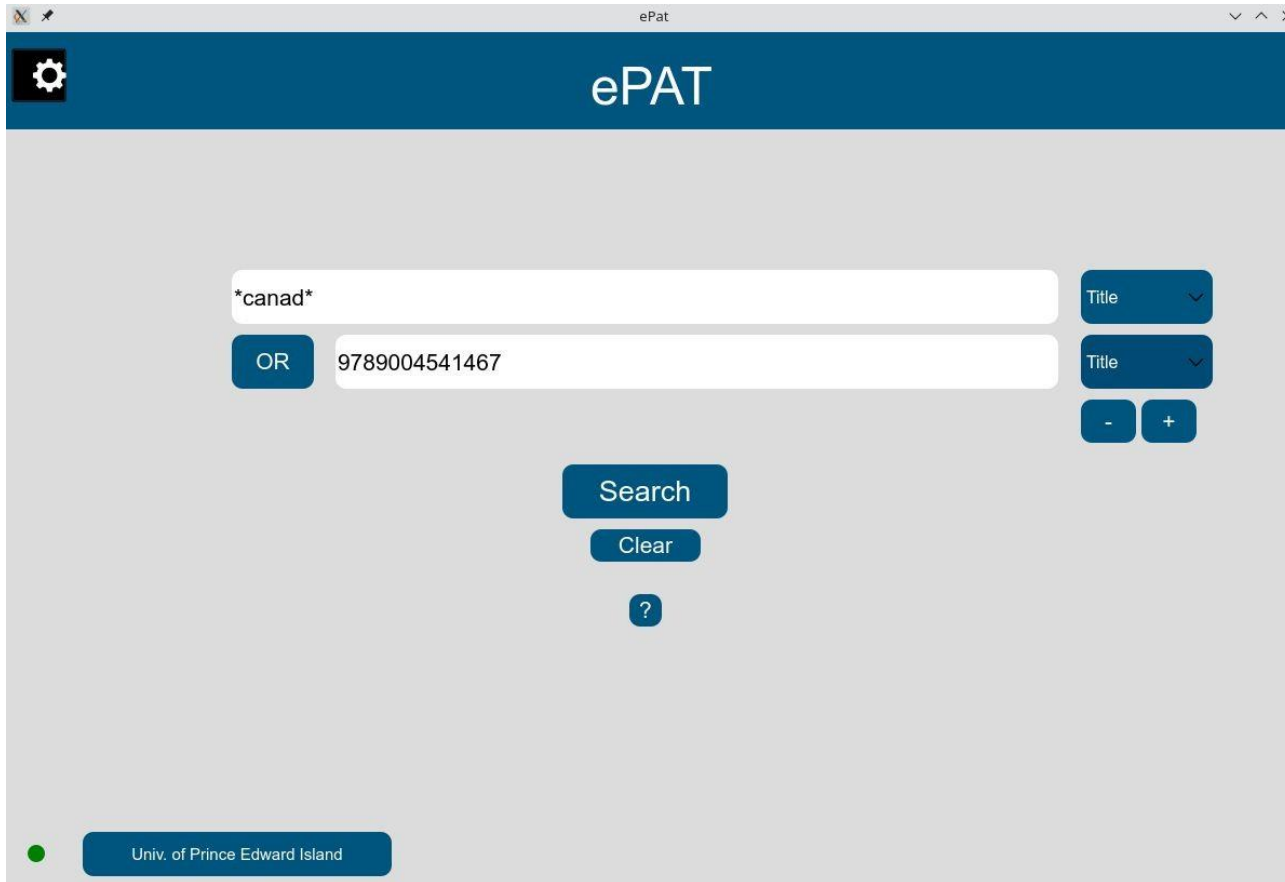


The screenshot shows the 'ePAT' application interface. At the top, there is a dark blue header with the 'ePAT' logo. Below the header, there is a navigation bar with a 'Home' button on the left and three buttons on the right: 'Allow CRKN Data' (with a checked checkbox), 'Update CRKN Data', and 'Upload Local Data'. The main content area is light gray and contains several settings:

- Select Institution:** A dropdown menu with 'Univ. of Prince Edward Island' selected.
- Select Language:** A dropdown menu with 'English' selected.
- CRKN eBook Data URL:** A text input field containing 'https://library.upei.ca/test-page-ebooks-perpetual-access-project'.
- Help Page URL:** A text input field containing 'tps://github.com/eppenney/eBook-Perpetual-Access-Rights-Tracker/wiki'.

At the bottom of the main content area, there are two buttons: 'Manage Local Institutions' and 'Manage Local Database'. In the bottom right corner, there is a 'Help' button.

Ebook rights tracker app 2



The screenshot shows the ePAT application interface. At the top, there is a dark blue header with a gear icon on the left and the text "ePAT" in the center. Below the header, the main content area is light gray. It features two search input fields: the first contains the text "*canad*" and the second contains "9789004541467". Between these fields is a blue button labeled "OR". To the right of the first field is a blue dropdown menu labeled "Title" with a downward arrow. To the right of the second field is another blue dropdown menu labeled "Title" with a downward arrow. Below these dropdowns are two small blue buttons, one with a minus sign "-" and one with a plus sign "+". In the center of the interface are three stacked blue buttons: "Search", "Clear", and a button with a question mark "?". At the bottom left, there is a small green dot and a blue button labeled "Univ. of Prince Edward Island".

Ebook rights tracker app 2

The screenshot shows the ePAT application interface. At the top, there is a search bar containing the text "A History of Canadian Economic Thought". Below the search bar is a table with 12 columns and 10 rows of data. The columns are: Access, File_Name, Platform, Title, Publisher, Platform_YOP, Platform_eISBN, OCN, greement_cod, ollection_nam, etadata_last_m. The table lists various books and their associated metadata. At the bottom of the interface, there are two buttons: "Univ. of Prince Edward Island" on the left and "Export" on the right.

Home

A History of Canadian Economic Thought

	Access	File_Name	Platform	Title	Publisher	Platform_YOP	Platform_eISBN	OCN	greement_cod	ollection_nam	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

Univ. of Prince Edward Island

Export

Moissonneuse

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

- Objectif : trouver les numéros d'appel LC (LCC) des ebooks pour l'analyse de la collection.
- LCC = granularité beaucoup plus grande que les grandes catégories de sujets typiques fournies par les vendeurs
- LCC : différentes plates-formes proposent différents schémas de catégories de sujets dans leurs rapports exclusifs, ce qui permet de comparer des pommes avec des pommes.
- Étant donné une liste d'ISBN, utiliser une combinaison d'API et de Z39.50 pour trouver un LCC approprié à utiliser avec ce livre.
 - Pour utiliser l'API de Google, vous devez fournir une clé d'API gratuite.
 - Pour utiliser Z39.50, vous devez également installer le logiciel libre "yaz-client"
 - Fonctionnalité de récupération des catalogues Blacklight sur le web, mais pour des raisons juridiques, ces catalogues ne sont pas remplis par défaut.

Moissonneuse

- Stocke les données dans une base de données sqlite, et utilise la première
- Vous choisissez les sources à utiliser, vous les classez par ordre de priorité.
- Produit les résultats dans un fichier tsv (délimité par des tabulations) qui peut être combiné avec d'autres feuilles de calcul (par exemple, les rapports d'utilisation de COUNTER).
- Vous pouvez ajouter des cibles Z39.50 et des cibles Blacklight OPAC webscrape supplémentaires.
- Avertissement - certains ISBN ne peuvent être associés à une lettre de crédit d'aucune de vos "cibles" - vous devrez peut-être en trouver manuellement.
- Brève explication de la raison pour laquelle OCLC API n'est pas inclus
 - L'application propose également des options OCN (utiliser, trouver) mais elles sont limitées à cause de ce problème d'API.
- Contactez-moi pour obtenir des données de base provenant de projets similaires antérieurs (trop volumineuses pour github).

Library Metadata Harvester

Start Search

 ISBN OCN

Choose File

Input File Path:

Enter Output File

Sources

Settings

Harvard (API)

Open Library (API)

LOC (API)

Google Books (API)

Yale (Z39.50)

UVa (Z39.50)

UAlberta (Z39.50)



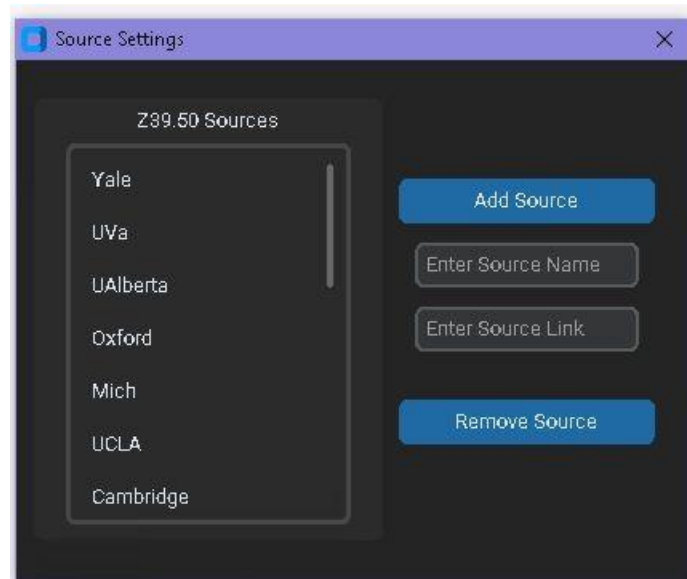
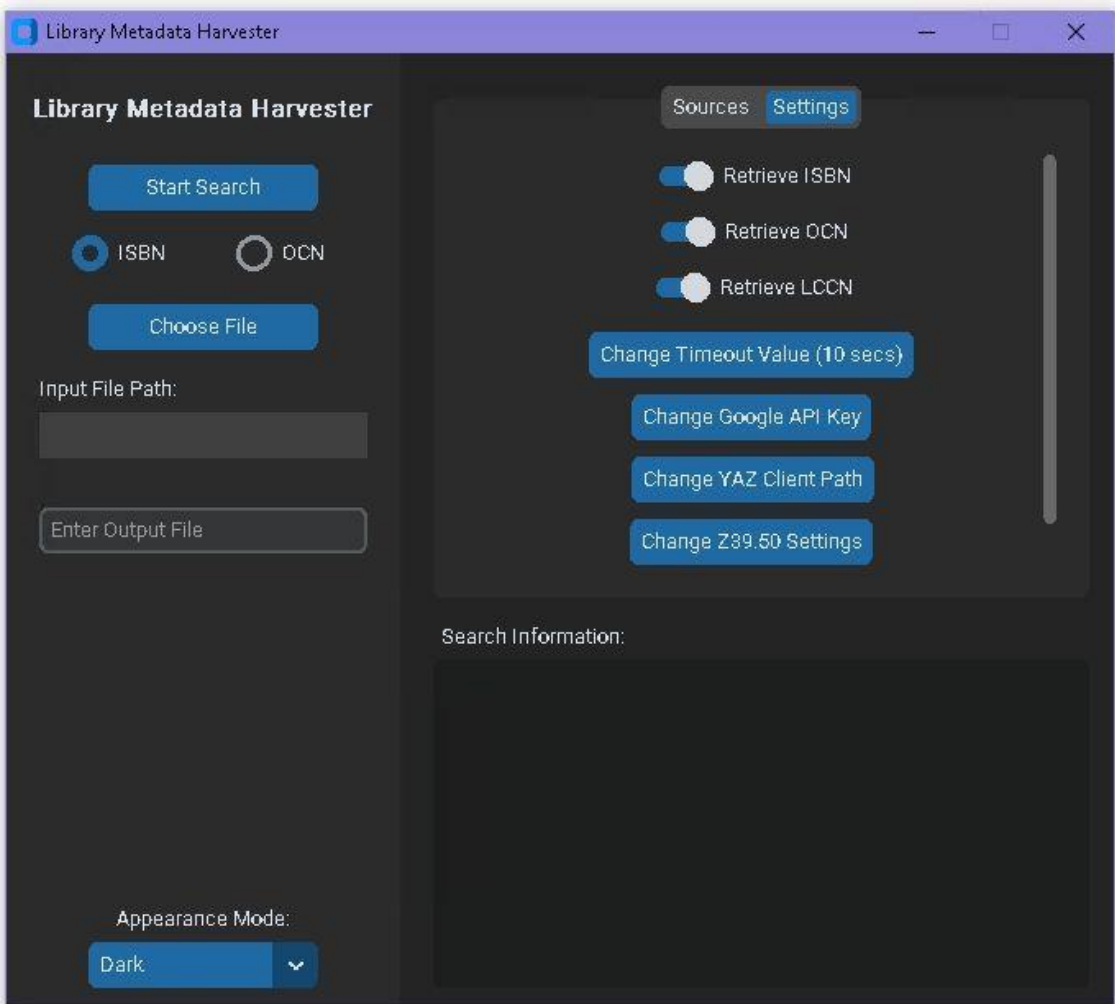
Save Order

Search Information:

Appearance Mode:

Dark





[Modifier l'API Google Ke}](#)

Chemin du fichier
d'entrée :

[Modifier le chemin d'accès du](#)

[client YAZ](#) [Modifier les](#)

[paramètres Z39.EO](#) [Modifier](#)

[les paramètres de Web](#)

[Scraping](#)

3Les informations sur la recherche :

Sources d'extraction de données sur le Web

[Apparition Mo'](#)

[Colombie](#)

[Ajouter 3source](#)

[Sombre](#)

[Supprimer la source](#)

Library Metadata Harvester

Start Search

ISBN OCN

Choose File

Input File Path:

C:/Users/Melissa/Desktop/CS482
0_2024_Projects/LMH/small_samp
le_isbns.txt

sample_output.tsv

Sources Settings

Harvard (API)

Open Library (API)

LOC (API)

Google Books (API)

Yale (Z39.50)

UVa (Z39.50)

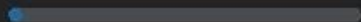
UAlberta (Z39.50)



Save Order

Library Metadata Harvester - Progress

Please wait while the search process is running.
This might take several minutes.



Stop Search

Appearance Mode:

Dark



Exemple de fichier de sortie
(tsv)

	A	B	C	D
1	SBN	OCLC	LCCN	LCCN-Source
2	9780191513015		PJ7942.127 A53 2020	Columbia
3	9780191513022		PJ7942.127 A53 2020	Columbia
4	9780191513039		PJ7942.127 A53 2020	Columbia
5	9780191513183		PJ7942.127 A53 2020	Columbia
6	9780191513213		PJ7942.127 A53 2020	Columbia
7	9780191513275		PN56.1465 B644 2005	Harvard
8	9780191513329		PJ7942.127 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		PJ7942.127 A53 2020	Columbia
13	9780191513466		PJ7942.127 A53 2020	Columbia
14	9780191513473		KD671 .H63 2005	Harvard
15	9780191513480		PJ7942.127 A53 2020	Columbia
16	9780191513503		BL1575 .H56 2005	Harvard
17	9780191513510		BR115.P7 S6324 1997	Harvard
18	9780191513572		SB975 .W3 2005	Harvard
19	9780191513589		PJ7942.127 A53 2020	Columbia
20				

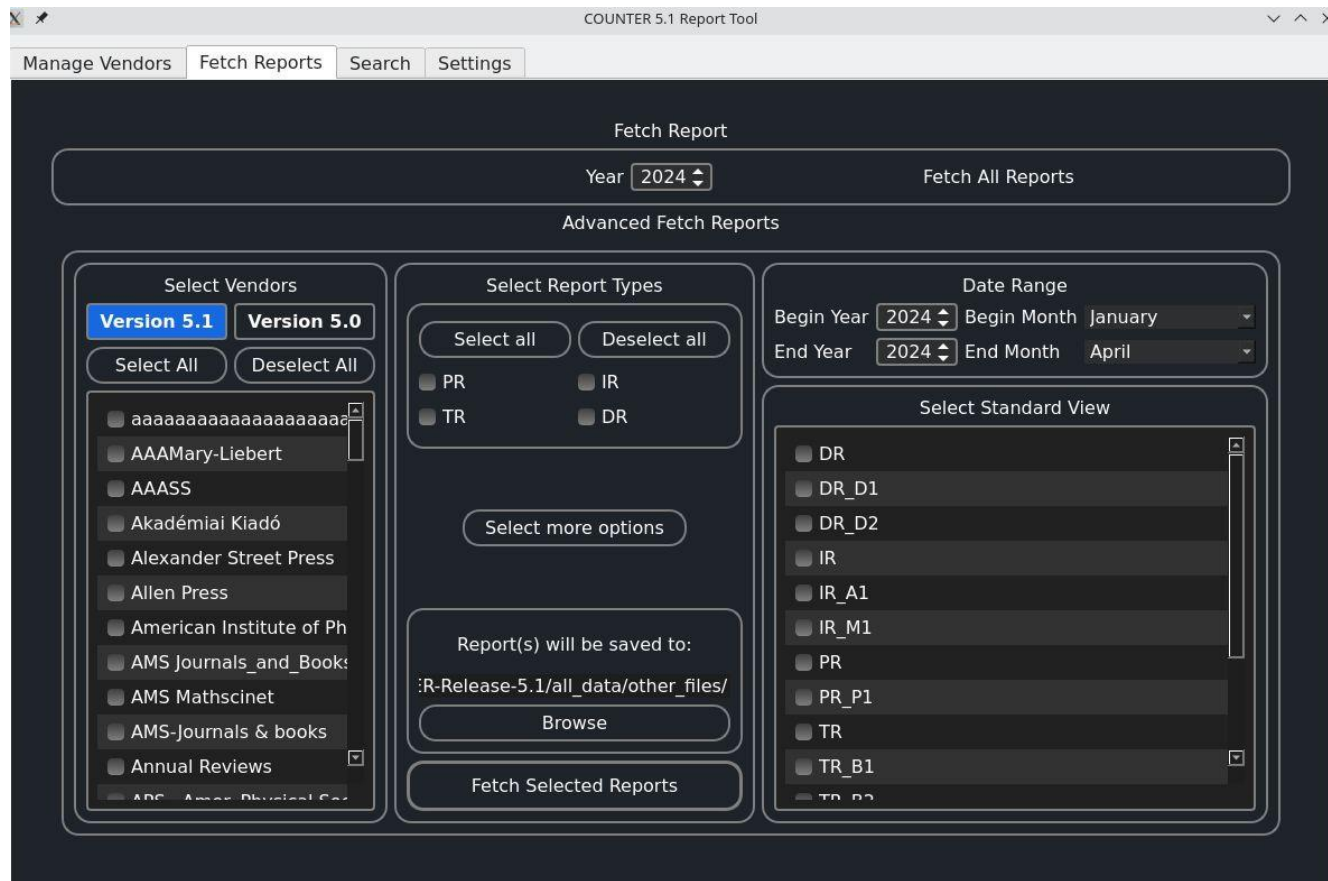
COMPTEUR 5.1 Moissonneur

- La version finale est retardée - elle devrait être prête à la fin de l'été
 - Problème de compatibilité avec la version 5.1 - il n'y a pas encore de serveurs de fournisseurs réels pour les tester.
 - Attendre 5.1 Rapports COUNTER à partir de janvier 2025
 - L'application gère les données 5.0 et 5.1
- L'application de bureau utilise l'API COUNTER Metrics (anciennement "SUSHI")
- Récupère les rapports principaux, les vues standard, les rapports principaux personnalisés (filtres et attributs).
- Génère des fichiers tsv exactement conformes aux normes du rapport COUNTER
- Recherche simple par titre/issn/isbn incluse, produit des fichiers tsv de toutes les données correspondantes sur toutes les plates-formes.
- Les données sont stockées dans une base de données sqlite qui est combinée entre les plates-formes pour des requêtes plus complexes.
- Permet également d'extraire les rapports sur les titres avec l'attribut YOP (Year of Publication) étendu.
- Outils utiles pour gérer les données d'identification de vos fournisseurs, et option de mot de passe de base pour la table des fournisseurs
- Fournit un fichier journal pour le dépannage, en signalant les problèmes à COUNTER/vendeur.

Documentation officielle de COUNTER 5.1 : <https://cop5.projectcounter.org/en/5.1/>

Site officiel de COUNTER (nouveau !): <https://www.countermetrics.org/>

COMPTEUR 5.1 Moissonneur



COMPTEUR 5.1 Moissonneur

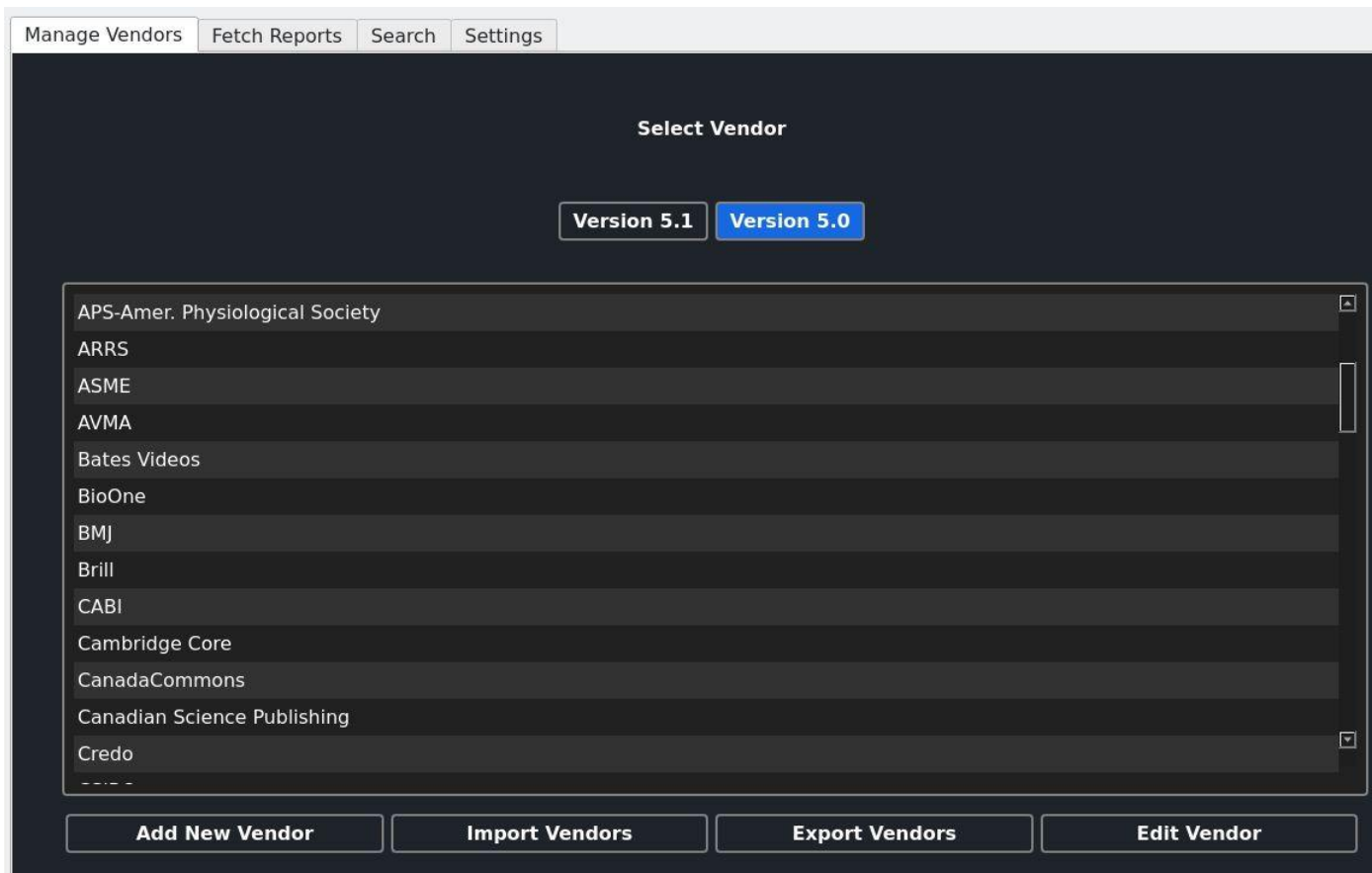
Manage Vendors Fetch Reports Search Settings

Select Vendor

Version 5.1 Version 5.0

- AP5-Amer. Physiological Society
- ARRS
- ASME
- AVMA
- Bates Videos
- BioOne
- BMJ
- Brill
- CABI
- Cambridge Core
- CanadaCommons
- Canadian Science Publishing
- Credo

Add New Vendor Import Vendors Export Vendors Edit Vendor



COMPTEUR 5.1 Moissonneur

Add New Vendor

Version 5.0

Name *

Base URL *

Starting Year *

Customer ID *

Requester ID

Platform

API Key

2 Attempts needed

IP Checking required

Request throttled

Provider

Notes

COMPTEUR 5.1 Moissonneur

The screenshot displays the COMPTEUR 5.1 Moissonneur software interface. The main window has a title bar "COUNTER 5.1 Repo" and a menu bar with "Manage Vendors", "Fetch Reports", "Search", and "Settings". The "Fetch Reports" tab is active, showing a "Year" dropdown set to "2024" and an "Advanced Fetch" section with "Select Vendors" and "Select Report Types" panels. The "Select Vendors" panel shows a list of vendors with "Version 5.1" selected. The "Select Report Types" panel shows "TR" selected. An "Options" dialog box is open, listing filters for "Data_Type", "Section_Type", "Access_Type", "Access_Method", "Metric_Type", "YOP", and "Exclude_Monthly_Details", each with a radio button and a "Choose" button. A "Data_Type options" dialog box is also open, showing a list of report types: "Article", "Book", "Book_Segment", "Database", "Dataset", "Journal", "Multimedia", "Newspaper_or_Newsletter", "Other", "Platform", "Report", "Repository_Item", and "Thesis_or_Dissertation". The "Data_Type options" dialog has "OK" and "Cancel" buttons.

Manage Vendors | Fetch Reports | Search | Settings

COUNTER 5.1 Repo | Options

Fetch Reports

Year: 2024

Advanced Fetch

Select Vendors

Version 5.1 | Version 5.0

Select All | Deselect All

- AAASS
- Akadémiai Kiadó
- Alexander Street Press
- Allen Press
- American Institute of Ph
- AMS Journals_and_Book
- AMS Mathscinet
- Annual Reviews
- APS Amer. Physical Soc
- APS Amer Physiological
- ARRS
- ACME

Select Report Types

Select all | Deselect all

- PR
- TR
- IR
- DR

Select more options

Report(s) will be saved to:
:R-Release-5.1/all_data/other_files/
Browse

Fetch Selected Reports

Options

Show

- Data_Type
- Section_Type
- Access_Type
- Access_Method
- Metric_Type
- YOP
- Exclude_Monthly_Details

Filters

- all Choose
- all Choose
- all Choose
- all Choose
- all Choose
- all Choose

Data_Type options

- Article
- Book
- Book_Segment
- Database
- Dataset
- Journal
- Multimedia
- Newspaper_or_Newsletter
- Other
- Platform
- Report
- Repository_Item
- Thesis_or_Dissertation

OK | Cancel

COMPTEUR 5.1 Moissonneur

Manage Vendors

Fetch Reports

Search

Settings

Search {TR Reports}

Start Year 2023

Start Month January

End Year 2024

End Month January

Title Search

Title Substring

canad%journal

Results: 9528 rows

File: /home/mbelvadi/Documents/COUNTER1/COUNTER-Release-5.1/search_output.tsv

Search

Go To Save Folder

COMPTEUR 5.1 Moissonneur

Plus de situations d'utilisation :

- Dépenses TSA - lorsque vous avez un livre dans un TSA et également sur d'autres plateformes, il est utile de voir l'utilisation totale de ce titre.
- Suppression de titres d'abonnement (EBSCO, Proquest), envisager l'achat avant la perte d'accès - peut vérifier l'utilisation sur les deux abonnements (en particulier si le livre est perdu sur les deux abonnements).
- Données DR : Déterminer les préférences des utilisateurs pour les plates-formes (bases de données), y compris la couche de découverte par rapport aux plates-formes de base de données natives.
- Semblable pour les revues dans différents sous-paquets, sites d'éditeurs

Utiliser les outils

Comment utiliser des combinaisons de ces éléments pour l'analyse des collections ?

Scénario 1 : vérifiez si vous avez beaucoup de livres en rupture dans certains domaines.

- Utilisez le moissonneur COUNTER pour extraire toutes vos données TR_B2 "refus" sur la période d'utilisation qui vous intéresse.
- Utiliser une simple requête sqlite pour extraire l'ISBN dans un simple fichier texte (et recommencer pour créer un fichier tsv avec des informations supplémentaires, par exemple les titres, le numéro de série, l'éditeur, la plateforme, etc.)
- Comparer la liste ISBN avec le collecteur LCC pour créer le fichier de sortie ISBN-LCC.
- Utiliser Excel/etc. pour mettre les deux fichiers de sortie dans une feuille de calcul puis "vlookup" pour joindre les données sur les refus aux LCC
- Utilisez un tableau croisé dynamique ou vos autres outils de feuille de calcul préférés pour produire un résumé utile.

Utiliser les outils

Scénario 2 : analyser vos fonds en accès perpétuel par domaine pour détecter les lacunes

- Utilisez l'un des outils Ebook pour récupérer toutes les données de votre PAR dans sa base de données sqlite sur toutes les plateformes.
- Utiliser une simple requête sqlite pour extraire l'ISBN dans un simple fichier texte.
- Comparer la liste ISBN avec le collecteur LCC pour créer le fichier de sortie ISBN-LCC.
- Utiliser Excel/Google Sheets/LibreOffice pour mettre les deux fichiers de sortie dans une feuille de calcul, puis "vlookup" pour joindre les données PAR aux LCC.
- Utilisez un tableau croisé dynamique ou vos autres outils de feuille de calcul préférés pour produire un résumé utile.

Utiliser les outils

Comment cela pourrait-il aider votre analyse des collections (ou même l'enseignement de la maîtrise de l'information) ?

- Si les données montrent que votre collection s'appuie fortement sur des titres non-PAR (EBA, DDA, abonnement) dans certaines matières très utilisées, vous pourriez peut-être augmenter le budget pour les livres électroniques dans ces domaines ? Ou l'inverse ? Augmenter les fonds pour acheter plus de titres PAR dans les domaines fortement utilisés ?
- Existe-t-il des modèles par sujet et par utilisation entre les différents fournisseurs/éditeurs auprès desquels vous achetez des RAP ? Cela peut refléter les choix des bibliothécaires OU les choix des utilisateurs. Par exemple, si les utilisateurs en psychologie ont tendance à utiliser les ebooks que vous achetez ou licenciez auprès du vendeur 1 beaucoup plus par titre que le vendeur 2, cela pourrait refléter des expériences de formation à la maîtrise de l'information mettant l'accent sur certaines plateformes, ou des recommandations de la faculté aux étudiants, ou même la qualité du contenu pour votre programme par éditeur/fournisseur ?

Utiliser les outils

Comment cela pourrait-il aider votre analyse des collections (ou même l'enseignement de la maîtrise de l'information) ?

- Comparer les formules d'abonnement aux grands livres électroniques en fonction de l'utilisation et des sujets (pour le partage des coûts entre les départements et les budgets ?)
- Les données COUNTER incluent l'éditeur pour les livres sur les plateformes agrégées, de sorte que vous pouvez évaluer pour un éditeur donné (à nouveau par sujet si nécessaire) si l'utilisation est meilleure sur la plateforme "native" de l'éditeur par rapport à divers agrégateurs.
 - ex. Dois-je acheter mon PAR pour les livres de Springer sur EBSCO/Proquest ou sur le site de Springer ?

Pouvez-vous imaginer d'autres utilisations ? Laissez libre cours à votre imagination ! Partagez vos idées !

Question sur les projets futurs

Serait-il utile d'avoir une application qui puisse ingérer les rapports de données COUNTER que vous récupérez manuellement dans une base de données sqlite plutôt que d'utiliser cet API harvester ?

Serait-il utile de disposer d'une application de bureau unique combinant ces trois éléments, ou deux d'entre eux ?

Serait-il utile d'envisager de combiner les données de Google Analytics ou d'autres fournisseurs propriétaires (par exemple les recherches par titre/issn dans votre liste A-Z) avec vos données COUNTER dans une seule base de données sqlite ? Certains fournisseurs, par exemple EBSCO et Proquest, ont des rapports d'utilisation détaillés et exclusifs. Fichiers KBART ?