Aalborg Universitet



Bridging data gaps in a Research Information Management System with OpenAlex

Vidmar, Søren

DOI (link to publication from Publisher): 10.5278/AAUOPENALEXCRIS

Creative Commons License CC BY 4.0

Publication date: 2024

Document Version Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA): Vidmar, S. (2024). Bridging data gaps in a Research Information Management System with OpenAlex. Poster presented at Officiel Lanceringskonference – Danmarks Forskningsportal, Frederiksberg, Denmark. https://doi.org/10.5278/AAUOPENALEXCRIS

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
 You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Bridging data gaps in a Research Information Management System with OpenAlex

How can you discern the absence of something when you lack a comprehensive source of all necessary information?

Identifying data gaps within a research information management system presents a considerable challenge. It hinges on the available resources, licenses, organizational procedures, and of course time. Promising endeavors such as OpenAlex could offer us significant assistance.

Consequently, we have begun experimenting with new tools that could aid us in both detecting and bridging these gaps eventually.

Author

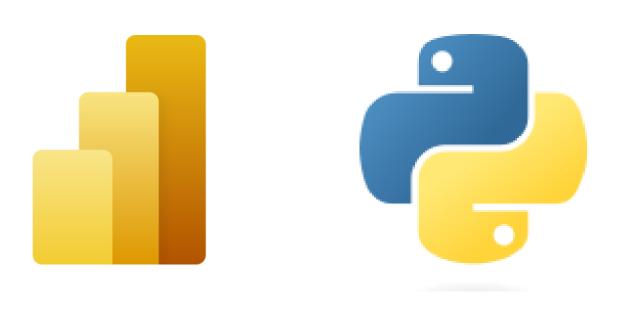
Søren Vidmar sv@aub.aau.dk 0000-0003-3055-6053

Affiliation

Aalborg University Library

AALBORG University





CRISearch and OpenAlex2RIS are experimental tools that can help you identify and close the data gap in your Current Research Information System (CRIS)

1 Introduction

Data within a CRIS must be as complete as possible to provide a dependable base for distributing, showcasing and evaluating an institution's research contributions. Despite solid institutional processes and access to multiple systems to help us out, some records inevitably slip through the cracks each year. OpenAlex aids us in identifying missing records and potentially enhancing the metadata quality in our CRIS.

2 **Objective**

To initiate the creation of an easy-to-use tool that facilitates the identification and closure of data gaps within our CRIS system.

Identifying missing content with CRISearch
 - A PowerBI tool for the OpenAlex API

CRISearch v0.1	OpenAlex	434		
		DOI's found in OpenAlex, not found in CRIS		

First step in closing the gap is identifying the publications missing in the CRIS.

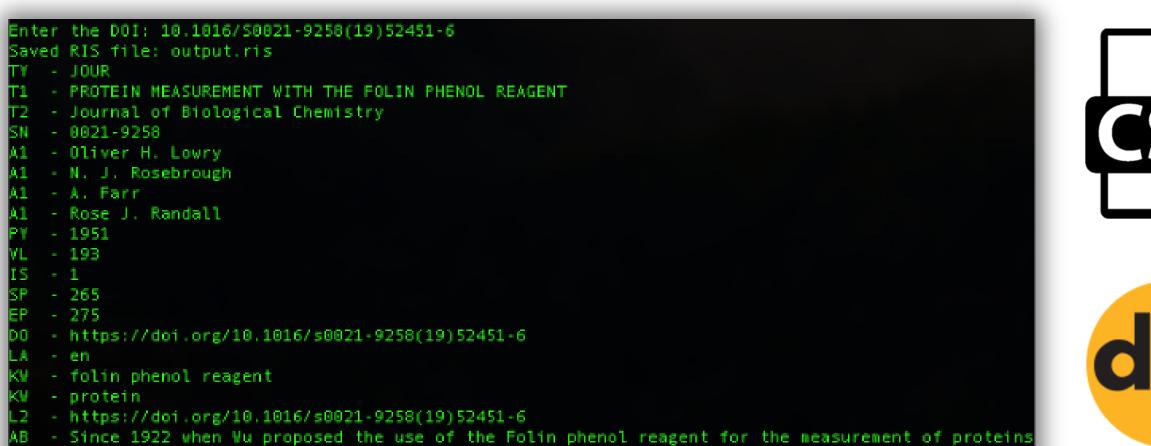
The primary function of CRISearch is to compare the publications from our institution in OpenAlex against those in our CRIS, providing us with a comprehensive overview of any research outputs our institution may be lacking.

We use Elsevier's Pure as our CRIS, but the tool is system-agnostic as such, relying on DOI matching and a supplementary Lucene search for title matching to account for missing or faulty DOIs. By utilizing the APIs of OpenAlex and our CRIS, we can systematically identify gaps in our database, by narrowing down a list of publications which belong to our institution, but which we don't have in our CRIS.

Possible match on title in CRIS?		Published in volume?		Publication type	~	Search affiliation string		
No	No		All		\sim	Search		۹ /
Yes		Yes						
001	Publica tion Year	Possible match on title in CRIS?	Title		Туре	Journal/Venue	Volume	Has volume
ttps://doi.org/10.1001/jamanetworkopen.2023.55716	2024	No	-	y of Midline vs Peripherally atheters Among Adults	journal-article	JAMA network open	7	Yes
ttps://doi.org/10.1002/9781394188789.ch1	2024	Yes	The Necessity for	Modernizing the Coupled gent Transportation Systems	other			No
ttps://doi.org/10.1002/9781394188789.ch6	2024	Yes		n of Power-To-X Energy ortation Networks	other			No
ttps://doi.org/10.1002/adfm.202313850	2024	No	Cyanoesterthioph	ene Based Low-Cost Polymer fficiency Organic Solar Cells	journal-article	Advanced Functional Materials		No
ttps://doi.org/10.1002/advs.202304834	2024	No	De Novo Atomisti	c Discovery of Disordered naterials by Machine Learning	journal-article	Advanced Science		No
ttps://doi.org/10.1002/alz.13681	2024	Yes	Mapping morbidi	ty 10 years prior to a g onset Alzheimer's disease	journal-article	Alzheimer's & Dementia		No
ttps://doi.org/10.1002/ctm2.1565	2024	No	Pericardial delta li ligand 1 (Dlk1) au	ke non-canonical NOTCH gments fibrosis in the heart to mesenchymal transition	journal-article	Clinical and translational medicine	14	Yes
ttps://doi.org/10.1002/dmrr.3775	2024	No	The impact of sod inhibitors on dem	ium-glucose co-transporter-2 entia and cardiovascular patients with atrial fibrillation	journal-article	Diabetes/Metabolism Research and Reviews	40	Yes
<u>ttps://doi.org/10.1002/ehf2.14688</u>	2024	Yes	Computed tomog	raphy or chest X-ray to assess stion in dyspnoeic patients	journal-article	Esc Heart Failure		No

4 Filling the gap with Python - OpenAlex2RIS

So identifying what's missing is one thing - filling the gap is another. To facilitate



modified analytical procedures ut ilizing this reagent have been reported for the

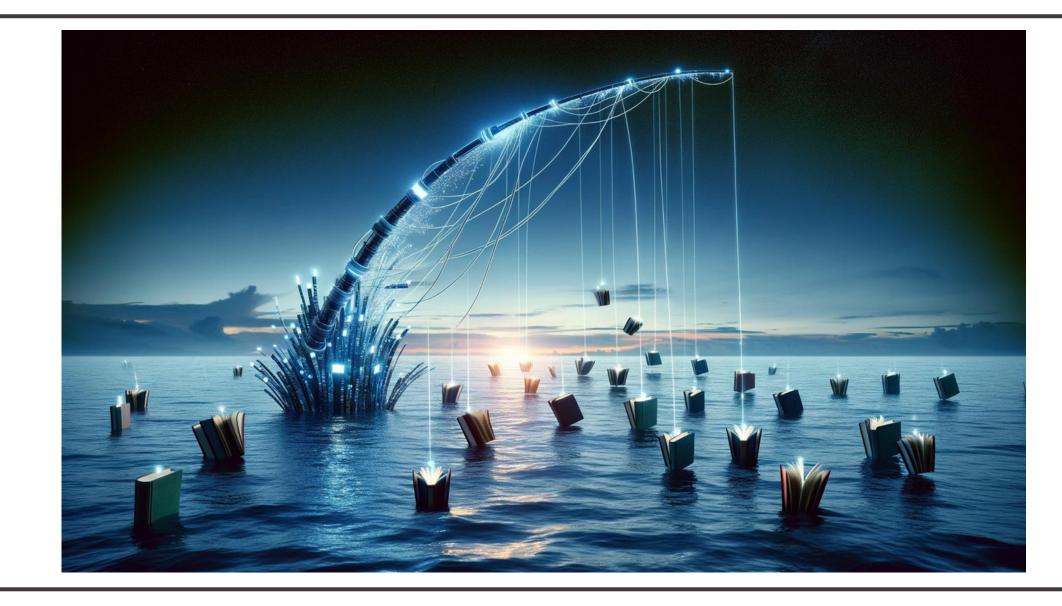
nation of proteins in serum (2-G), in antigen-antibody precipitates (7-9), and in insulin (1

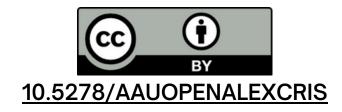
the integration of missing records, these scripts are capable of submitting either an individual DOI or a collection of DOIs in a CSV file to the OpenAlex API, producing RIS files in response. These RIS files can then be imported into our CRIS system.



5 Conclusion

Based on initial tests, it's advisable to approach publication import automation cautiously and ensure manual quality checks are in place. No database can encompass all sources, and while an open, comprehensive database by itself doesn't close the gap, it represents an important step in the right direction. Therefore, it's crucial that we explore OpenAlex more extensively in our work with information management. OpenAlex offers many opportunities we have yet to fully explore. Let's continue to delve into this resource and see what we can develop from it.





Related content Code available on GitHub: <u>https://github.com/svidmar</u> Article on LinkedIn: <u>https://www.linkedin.com/pulse/bridging-data-gaps-research-information-management-system-vidmar-wbonf</u>