Beyond Bibliometrics

*Using Scopus to assess collections and support faculty*

Lori Ostapowicz-Critz, Worcester Polytechnic Institute
Julie Morris, University of New Brunswick
Eric Livingston, Elsevier
Emily Singley, Elsevier

May 9, 2023
Helping Faculty Tell Their Story at WPI

Lori Ostapowicz-Critz
Associate Director - Scholarly Communication and Open Strategies
Worcester Polytechnic Institute
Worcester Polytechnic Institute

- “A Distinctive Approach to STEM Education and Research”
- Global Leader in Project-Based Education
- Top 20% U.S. college rankings (Times Higher Education, World University Rankings 2022)
- ~6300 FTE
- 70+ PhD, master's, bachelor’s, and certificate programs are offered both on campus and online
- 600+ Incredible faculty dedicated to both teaching & research
A Universal Issue …

- Faculty prepare promotion/tenure packets
- Increasing need to demonstrate “impact”
- Citation/bibliometric analysis required
  - Direct data mining in Scopus interface
  - Easy to pull traditional citation analysis data
  - H-index automatically calculated
- Additional data needed to “tell the story”
Additional Data for “Impact”

- Social media data is a critical component of impact!
- Scopus provides PlumX metrics at record level
  - provide insights into the ways people interact with individual pieces of research output in the online environment (e.g. tweets, favorites, shares, likes ...)
- A great addition to the bibliometric analysis in faculty packet
  - NOT easy to capture
  - Copy & paste for each record
- Then, I discovered the API!
The Data Fetcher Comes to the Rescue!

- What took HOURS of work now produced in minutes!
- Required components:
  - API Key
  - Institutional Key/Token
  - Data Fetcher client
- Output is easily integrated into spreadsheets, Word documents, etc.

"Key 3" by -Brenda-Starr- is licensed under [CC BY 2.0](https://creativecommons.org/licenses/by/2.0/).
Using the Data Fetcher for PlumX Metrics and more

1. Enter your API Key and Institutional Token (if needed) under CREDENTIALS

2. Enter your QUERY
   - Use EDIT button to enter changes
   - For specific researchers you need their Scopus Author ID
Using the Data Fetcher for PlumX Metrics and more

3. Select your OUTPUT parameters (under FIELDS)
   - Click on EDIT
   - In EDIT window use ADD button
   - SELECT fields to add and click OK

4. SAVE the parameters when done
Using the Data Fetcher for PlumX Metrics and more

5. Click BEGIN/RUN
6. OUTPUT file will be in the Output folder for the client when query is completed
Output from Data Fetcher

Some data clean-up needed (re-organization, renaming of fields, etc.)
The Final Product

A 29 page report including the PlumX metrics!

And, a less stressed librarian plus appreciative faculty!

PlumX Metrics

<table>
<thead>
<tr>
<th>Pub Date</th>
<th>Title</th>
<th>Captured Mentions</th>
<th>Captured Reader Count</th>
<th>Captured Impact Scores</th>
<th>Citations</th>
<th>Shares</th>
<th>Mentions</th>
<th>Social Media Mentions</th>
<th>Social Media Uses</th>
<th>Social Media Comments</th>
<th>Social Media Force Score</th>
<th>Image Use</th>
<th>Image Downloads</th>
<th>Image Links</th>
<th>Image Visits</th>
<th>Unique Visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>Moore et al. supported an imaginative multi-disciplinary research on optimizing energy recovery during catalytic hydrogenation of biodiesel</td>
<td>7</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Difference as a Model of Relevance for Predicting Interactions in Macromolecular Assemblies</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>Integration of Statistical Analysis and Recursion in Vanilla and Oblique Decision Trees from Samples</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Multiple structural determinations from electron flow models analysis</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Subcritical water hydrothermal pretreatment of sugarcane bagasse to produce second generation ethanol</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Experimental and conceptual evaluation of heavy water extraction, water reaching dangerous levels of contaminated water</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Hydrothermal conversion of other wastes to produce renewable biofuels: pellets for use in renewable energy</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Two-scale analytical simulations of inorganic-organic composites: Experiments, applications and applicability to Sc-PHSCs</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Chemicals from heavy crude in the EOR industry, multicomponent environmental engineering</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Six steps for extending platinum durability in sulfidized systems using thermodynamic coatings</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Sequential solids-water process applied to orange peel for the recovery of flavonoids and sugars</td>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td>Rapid Desorption of Desorbed Radionuclides to Soluble Products via Physisorption under H2 Conditions</td>
<td>5</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
By Our Powers Combined: Bridging collections analysis and bibliometrics, using researcher article citations to assess collection impact

Julie Morris (they/them)
Collections Analysis/Bibliometrics Librarian
University of New Brunswick
AGENDA

1. Introduction
2. Collections Analysis
3. Lit Review
4. Citation Analysis Project
5. Limitations & Lessons Learned
6. Applications
7. Conclusion & Next Steps
INTRODUCTION

Intersections of bibliometrics and collections analysis

Focus on citation analysis

Analysis is needed in the time of rising serials costs

CA can be used to help reevaluate Big Deals
THE QUESTIONS

- How will this project help us unbundle our Big Deals?
- To what extent does our collection meet our authors’ research needs?
- Do gaps in our collections holdings adversely impact our authors’ research?
WHAT IS COLLECTIONS ANALYSIS?

DATA-INFORMED

COLLECTION MANAGEMENT

ORGANIZATIONAL PRIORITIES
<table>
<thead>
<tr>
<th>Database</th>
<th>Publisher</th>
<th>Publisher_ID</th>
<th>Reporting_Period_Total</th>
<th>Proprietary_ID</th>
<th>Metric_Type</th>
<th>Reporting_Period_Total</th>
<th>Jan-2017</th>
<th>Feb-2017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>2822</td>
<td>pubx:dbA</td>
<td>Searches_Automated; Searches_Federated; Searches.Regular; Total_Item_Investigations; Total_Item_Requests</td>
<td>234</td>
<td>531</td>
<td></td>
</tr>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>2372</td>
<td>pubx:dbA</td>
<td>Searches_Federated</td>
<td>159</td>
<td>456</td>
<td></td>
</tr>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>3410</td>
<td>pubx:dbA</td>
<td>Searches.Regular</td>
<td>332</td>
<td>629</td>
<td></td>
</tr>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>17208</td>
<td>pubx:dbA</td>
<td>Total_Item_Investigations</td>
<td>1450</td>
<td>3232</td>
<td></td>
</tr>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>8484</td>
<td>pubx:dbA</td>
<td>Total_Item_Requests</td>
<td>705</td>
<td>1596</td>
<td></td>
</tr>
<tr>
<td>Database A</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>3008</td>
<td>pubx:dbB</td>
<td>Searches_Automated; Searches_Federated</td>
<td>229</td>
<td>543</td>
<td></td>
</tr>
<tr>
<td>Database B</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>2558</td>
<td>pubx:dbB</td>
<td>Searches_Federated</td>
<td>154</td>
<td>468</td>
<td></td>
</tr>
<tr>
<td>Database B</td>
<td>DB Publisher X</td>
<td>PlatformX</td>
<td>3596</td>
<td>pubx:dbB</td>
<td>Searches.Regular</td>
<td>327</td>
<td>641</td>
<td></td>
</tr>
</tbody>
</table>
LIT REVIEW

- “Citation analysis is a branch of bibliometrics” (Hoffman & Doucette, 2012)
- No set of guidelines for citation analysis (Hoffman & Doucette, 2012); not much automation (White, 2019)
- Many studies focus on single disciplines and the results
UNB STATS

75 Programs
10,000+ Students
1,200+ Graduate students
900+ Faculty
70% Sponsored research in NB
$55.67M External research funding
THE PROJECT

- Scopus data source
- 3 years of data (2019-2021)
- N = 120,982 journal publications
- 68,919 had DOIs
- Includes graduate students and faculty research
- Work in progress
STEP 1: EXTRACT YOUR PUBLICATION DATA
STEP 2: RUN IT BACK THROUGH THE DATA FETCHER
**STEP 3: CLEAN YOUR PUBLICATION DATA**

### Using facets and filters

Use facets and filters to select subsets of your data to act on. Choose facet and filter methods from the menus at the top of each data column.

Not sure how to get started? Watch these screencasts.

---

<table>
<thead>
<tr>
<th>Record</th>
<th>Format</th>
<th>EID</th>
<th>SD</th>
<th>First Letter (LM)</th>
<th>Source Name</th>
<th>Alternate Source Name</th>
<th>Verified Title Spell? (LM)</th>
<th>Part of KBART file? (LM)</th>
<th>Part of Scopus Pub List? (LM)</th>
<th>Volume</th>
<th>First Page</th>
<th>Last Page</th>
<th>First Year</th>
<th>Last Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>85065338671</td>
<td>3550761</td>
<td>A Natural History Bulletin of the New York Metropolitan Society</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>15</td>
<td>9</td>
<td>35</td>
<td>1978</td>
<td>1979</td>
</tr>
<tr>
<td>2</td>
<td>85065236352</td>
<td>35506749208</td>
<td>Can GIS Adult Geò Reg Data</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2012</td>
<td>2012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>85107053020</td>
<td>351070523655</td>
<td>Food-Based Dietary Guidelines for India 2021</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2016</td>
<td>2016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>85106490209</td>
<td>35106553993</td>
<td>Revised Selected Papers, Lecture Notes in Computer Science</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2016</td>
<td>2015</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>85110051269</td>
<td>35506512599</td>
<td>WHO Guidelines on Physical Activity and Sedentary Behaviour</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2020</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>85114252933</td>
<td>35505728547</td>
<td>Journal of American Studies</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2014</td>
<td>2014</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>85065364705</td>
<td>35506570027</td>
<td><em>Lazy, Improvise People</em>: Myth and Reality in the Making of Spanish History</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2056</td>
<td>2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>85114509267</td>
<td>35114535865</td>
<td>*Documentary Cultural Landscapes of Movement in the Uskoks of Makarska: First Nation Territory in New Brunswick, Canada, and Maine, United States</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2020</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>85065319515</td>
<td>35506384254</td>
<td><em>“Palin, as the Fifth Vital Sign,” and Dependence on the “Nautical Pain Case” Is Being Abandoned in the US Navy?</em></td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2016</td>
<td>2016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>85128519973</td>
<td>35125077803</td>
<td>The Networks Infection Contagious Diseases: Publishing System (NCDP System): The Case of</td>
<td>MLA</td>
<td>MLA</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>#N/A</td>
<td>2020</td>
<td>2020</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
STEP 4: KEEP CLEANING YOUR DATA

Web of Science

Journal

Abbreviations List
STEP 5: WORK WITH DOIS

- Extract DOIs
- Extract DOI prefixes
- Determine publisher name
STEP 6: EXTRACT HOLDINGS AND CLEAN THE DATA
STEP 7: COMBINE YOUR DATA

| B   | C   | D    | E     | F     | G     | H     | I     | J     | K     | L     | M     | N     | O     | P     | Q     | R     | S     | T     | U     | V     | W     | X     | Y     | Z     | AA   | AB   | AC   |
|-----|-----|------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| EID | SID | Source | Part of | Volum | FirstPage | LastPage | FirstYear | LastYear | Title | Text | DOI | DOI Pre | Publish | Alternate | ISSN | e-ISSN | Open A | Source | Author | Genera | Science | r Biology | 11 | 12 | 13 |
| 8.51E+10 | 8.51E+10 | 10th Euro | no | no | 2013 | 2013 | Tidal turb, Fleming, C., McIntosh, S., Willden, R.J., Tidal turbine performance in sheared flow. 10th Euro | Fleming C., McIntosh S., Willden R.J. |
| 8.51E+10 | 3.18E+10 | Forest E | yes | yes | 12 | 19 | 2005 | 2005 | Turkelnbck Sidle RC, 10.1016/j. | 10.1016 Elsevier | Elsevier |
| 8.51E+10 | 8.8E+10 | 11th IEEE | no | no | 2013 | 2013 | Conceptualizing BLOCKCHAINs: CHARACTERISTICS & APP | Conceptualizing BLOCKCHAINs: CHARACTERISTICS & APP |
| 8.51E+10 | 8.8E+10 | 11th IEEE | no | no | 2018 | 2018 | Conceptualizing BLOCKCHAINs: CHARACTERISTICS & APP | Conceptualizing BLOCKCHAINs: CHARACTERISTICS & APP |
| 8.51E+10 | 8.13E+12 | 12th USE | no | no | 2016 | 2016 | Don't Get David Lion, Adrian Chiu, Hailong Sun, Xin Zhuang, Nikola Groveski, and Ding Yuan. Don't Get C | David Lion," Adrian Chiu, "Hailong Sun, "Xin Zhuang, "Nikola Groveski, "Ding Yuan |
STEP 8: RUN ANALYSIS USING POWER BI
WHAT PUBLISHERS ARE WE USING?

2007 # of unique publishers cited

Count of Publisher Name by Number of Citations

- Elsevier: 34.8K
- Springer: 11.8K
- Wiley: 9.9K
- IEEE: 5.8K
- Taylor & Francis: 5.1K
- American Chemical Society: 5.1K
- The Korean Society of Plant Taxonomists: 4.2K
- SAGE: 2.3K
- Oxford University Press: 2.3K
- Nature Publishing Group: 2.0K
- MDPI: 1.5K
- American Geophysical Union: 1.4K
- Royal Society of Chemistry: 1.3K
WHAT JOURNAL PACKAGES ARE WE USING?
WHAT JOURNALS ARE WE USING?

Count of Journal Title by Number of Citations

- Ceramics International: 1261
- SCIENCE: 887
- Carbohydrate Polymers: 840
- PLoS One: 797
- ACS applied materials & interfaces: 723
- NATURE: 707
- Bioresource Technology: 688
- Cellulose: 635
- CHEMICAL ENGINEERING JOURNAL: 619
- Proceedings of the National Acad...: 539
- Materials Science and Engineering...: 534
- Canadian Journal of Fisheries and ...: 476
- Scientific Reports: 465
- ADVANCED MATERIALS: 455
- ACS sustainable chemistry & engi...: 454
- Geophysical Research Letters: 442
- IEEE Transactions on Industry Appl...: 403
- ACTA MATERIALIA: 401
- ECOLOGY: 388
- Chemical Geology: 381
- MARINE ECOLOGY PROGRESS SER...: 375
- Journal of Materials Chemistry A: 361
- Geochimica et cosmochimica acta: 357
- Science of the Total Environment: 352
- Aquaculture: 350
- Journal of Chemical Physics: 343

7617 # of unique publishers cited
WHAT YEAR WERE MOST CITED JOURNAL ARTICLES PUBLISHED?

20% Share of publications from 2018-2020
WHICH OA JOURNALS ARE WE USING MOST FREQUENTLY?
WHICH JOURNALS ARE BEING CITED THAT WE DON’T HAVE IN THE COLLECTION?
WHAT FIELDS ARE WE CITING MOST FREQUENTLY?
FEEDBACK
LIMITATIONS

- Older publications (pre-2000) may not have DOIs
- Source is based on who owns the publication today
- Errors with how authors cite papers
- Trouble with technology
- Reproducibility in the Arts & Humanities
POSSIBLE APPLICATIONS

- Use citations to build out research guides
- Fill gaps in collections
- Look for trends in EDI research and strengthen our commitment to diverse collections
- Determine which backfiles we wish to purchase
- Analyze citations in conjunction with COUNTER reports to get a broader view of what is being used
- Cost per cited reference (CPCR) figures (Pastva et al., 2018)
- Author reading vs publishing trends
CONCLUSION

1. How will this project help us unbundle our Big Deals?
2. To what extent does our collection meet our authors’ research needs?
3. Do gaps in our collections holdings adversely impact our authors’ research?
WHAT'S NEXT?

1. Compare findings to COUNTER reports and ILL reports (Ramamoorthu & Jeyshankar, 2016)

2. Input results as scenarios in Unsub

3. Interview faculty members and liaison librarians about journal usage and citation patterns

4. Determine which journals are cited together, in clusters (Belter & Kaske, 2016)

5. Code: Investigate automation and analysis using Python scripts (White, 2019)
REFERENCES


Data Fetcher: API Data without Coding

Elsevier API Data Fetcher: graphical user interface program free for Elsevier RI subscribers

Easy access to API benefits, like:

- Alleviated record export limits
- Additional metadata fields
- Expanded export options
- Combining products (SD, Scopus, SciVal, PlumX, Data Monitor, etc)
- Network exports for authors, institutions, countries, and more
- Topic modelling
- And more!
Data Fetcher Support Webpage

https://dev.elsevier.com/data_fetcher.html

One-stop shop that includes:

- Download link for latest version
- User's Guide
- Data Fetcher Newsletter sign-up
- Dedicated support e-mail for Data Fetcher users

Data Fetcher

The API query tool, affectionately known as the Data Fetcher for its ability to fetch data from Scopus, PlumX, SciVal, and other APIs, was built in-house by an inspired Elsevier employee who wanted to help his clients more easily generate and analyze vast arrays of API data. In addition, the tool supports and continues to support analytical and intelligence use cases with needs beyond what’s available in the API interfaces of Elsevier’s Research Products, but that depend on the data available from those products. The menu-driven interface of the tool makes it easy to use and serves both casual users without the time or background to code as well as advanced coders and scripters not wishing to reinvent the wheel.

On the technical side of things, the Data Fetcher requires no installation. It is a self-extracting .exe file run on your device, and you simply need to download the tool archive, unzip the file into a local folder (e.g. on your Desktop, etc.), ensure your personal API key is placed into the tool’s configuration file, and then run the executable in place.

Why use the Data Fetcher?

- You are a data analyst but not a programmer
- You do not have the time to code
- Your data needs exceed what is possible or practical using the built-in export facilities of Scopus.com and other Elsevier products; the tool accesses (e.g. Plum Metrics, the Elsevier Fingerprint Engine, etc.)
- You need to mix data from several sources into one cohesive output (i.e. Scopus, SciVal, PlumX, etc.)
- The Scopus.com interface isn’t optimized for your specific use case, but Scopus data is

Who is eligible to use the Data Fetcher?

- Academic and government users associated with an institution that subscribes to Scopus and/or SciVal. The Data Fetcher ZIP file, available below, is password-protected. Please contact datafetcher.support@elsevier.com to acquire the password.
- The tool is currently Windows-only, though it has an optional web-based interface that can be accessed from other platforms (e.g. MacOS, Linux). We are currently working on a MacOS (M1) version as well.

Disclaimer

This is pre-production software provided "as is" for development use only. In no event shall Elsevier be liable for any claims, damages, or other liability arising from, out of, or in connection with the software or its use.

Is there documentation for the Data Fetcher?

Yes there is! You can view the Getting Started Guide for the Data Fetcher here.

Download the Data Fetcher Tool

NOTE: The Data Fetcher ZIP file is password-protected. Please contact datafetcher.support@elsevier.com to acquire the password.

Sign up for the Data Fetcher Newsletter

You can opt out of the newsletter at any time here.
Setting up Data Fetcher

Download Data Fetcher folder and extract onto your computer:
https://dev.elsevier.com/data_fetcher.html

In the `.env` file, add your API key* and institutional token**

* API keys can be obtained from https://dev.elsevier.com/; you must sign into your account to get access to your subscribed products
** Institutional tokens are required for some API services and can be requested from datasupportRD@elsevier.com
Data Fetcher Input Options: Query

Use any Scopus query as an input.

Pro Tip: make sure that it runs on Scopus.com first and then copy & paste.
Data Fetcher Input Options

Upload a file with a list of PIDs to retrieve information from. 
*Not (easily) replicated in Scopus.com*

Combine with query if desired!
Data Fetcher Output Options

- .csv, .txt (tab delimited), and .xlsx formats
- Publications
- Expand authors
- Expand affiliations
- Expand author affiliations
- Expand citations

Also options for .ris & Scopus.com export formats

Browse menu of metadata fields for each export options
Example Data Fetcher exports

Flexible to include the data you need in the format you need for any bibliometric analysis

Perform analysis on authorship position

- Export format with each publication-author combination on a separate row
- Exporting another column with author sequence on paper (e.g., first)
- Allows you to incorporate author position into bibliometric analyses
**Example Data Fetcher exports**

Flexible to include the data you need in the format you need for any bibliometric analysis

<table>
<thead>
<tr>
<th>AU-ID</th>
<th>First Name</th>
<th>Last Name</th>
<th>Initials</th>
<th>CitationCount</th>
<th>CitedByCount</th>
<th>CoAuthCount</th>
<th>DocCount</th>
<th>HIndex</th>
<th>ORCID</th>
<th>PubRange EndYear</th>
</tr>
</thead>
<tbody>
<tr>
<td>5721867247</td>
<td>Oh</td>
<td>B.</td>
<td></td>
<td>27</td>
<td>23</td>
<td>11</td>
<td>5</td>
<td>3</td>
<td></td>
<td>2020</td>
</tr>
<tr>
<td>55596380000</td>
<td>J. A.</td>
<td>Peacock</td>
<td>J.A.</td>
<td>89972</td>
<td>67338</td>
<td>2273</td>
<td>308</td>
<td>102</td>
<td>1983</td>
<td>2022</td>
</tr>
<tr>
<td>12142820200</td>
<td>Sadegh</td>
<td>Khochfar</td>
<td>S.</td>
<td>12806</td>
<td>6375</td>
<td>362</td>
<td>159</td>
<td>61</td>
<td>2003</td>
<td>2022</td>
</tr>
<tr>
<td>16432990400</td>
<td>Britton D.</td>
<td>Smith</td>
<td>B.D.</td>
<td>4499</td>
<td>3187</td>
<td>187</td>
<td>62</td>
<td>30</td>
<td>2006</td>
<td>2022</td>
</tr>
<tr>
<td>14056654600</td>
<td>A. B.</td>
<td>McDonald</td>
<td>A.B.</td>
<td>11631</td>
<td>7795</td>
<td>1184</td>
<td>160</td>
<td>39</td>
<td>0000-0002</td>
<td>1967</td>
</tr>
<tr>
<td>55958113500</td>
<td>J. R.</td>
<td>Klein</td>
<td>J.R.</td>
<td>9115</td>
<td>5710</td>
<td>2262</td>
<td>90</td>
<td>29</td>
<td>1994</td>
<td>2022</td>
</tr>
<tr>
<td>6701511183</td>
<td>David L.</td>
<td>Work</td>
<td>D.L.</td>
<td>7</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2003</td>
<td>2003</td>
</tr>
<tr>
<td>57203217381</td>
<td>Valeri P.</td>
<td>Frolov</td>
<td>V.P.</td>
<td>7308</td>
<td>4183</td>
<td>92</td>
<td>223</td>
<td>49</td>
<td>1971</td>
<td>2022</td>
</tr>
<tr>
<td>20433653700</td>
<td>Kyungmin</td>
<td>Kim</td>
<td>K.</td>
<td>45342</td>
<td>15835</td>
<td>6391</td>
<td>170</td>
<td>70</td>
<td>2007</td>
<td>2022</td>
</tr>
<tr>
<td>9249595100</td>
<td>Hyun-kyu</td>
<td>Lee</td>
<td>H.</td>
<td>46577</td>
<td>16156</td>
<td>6357</td>
<td>271</td>
<td>69</td>
<td>1989</td>
<td>2022</td>
</tr>
<tr>
<td>7006397641</td>
<td>F. C.</td>
<td>Khanna</td>
<td>F.C.</td>
<td>3147</td>
<td>2002</td>
<td>147</td>
<td>244</td>
<td>29</td>
<td>1960</td>
<td>2022</td>
</tr>
</tbody>
</table>

**Author profile information**

- Export a list of author profiles and the data you need from author profiles
Example Data Fetcher exports

Flexible to include the data you need in the format you need for any bibliometric analysis

Author-affiliation linkages

- Export duplicated publication list, with a separate row for each affiliation used by each author
- Can limit to just affiliations on the publications or their full affiliation history
- Enables mobility analyses and filtering by publications when an author used a specific affiliation
Other DataFetcher Features - network maps

Create network maps for authors, affiliations, countries, and more
Specify minimum number of co-authored publications
Filter networks by list of nodes of interest (e.g., author list)
Aggregate vs. disaggregated adjacency lists
Enhancing VOSviewer analyses

Replace raw data in Scopus exports with disambiguated data for accurate analyses

Kyle Demes’ co-authors (exported 11 May 2022) mapped in VoSViewer based on VoSViewer default (on left) and API-fueled map that disambiguates author name variants
Recent Data Fetcher Use Cases

- Large exports for research on research
- Social media & policy citation reports
- Customized metrics (e.g., time-limited H’index)
- Identifying institutions and countries citing work of an institution (large scale)
- Demonstrating collaboration among researchers
- Tracking affiliations of grad students post degree
- Finding researchers who have previously worked at another institution
Thank you!

Questions?

Lori Ostapowicz-Critz, lostapowiczcritz@wpi.edu
Julie Morris, julie.morris@unb.ca
Eric Livingston, e.livingston@elsevier.com
Emily Singley, e.singley@elsevier.com

For more information and Data Fetcher support:
- [https://dev.elsevier.com/data_fetcher.html](https://dev.elsevier.com/data_fetcher.html)
- DataFetcherSupport@elsevier.com