By Our Powers Combined: Bridging collections analysis and bibliometrics, using researcher article citations to assess collection impact

Julie Morris (they/she)

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

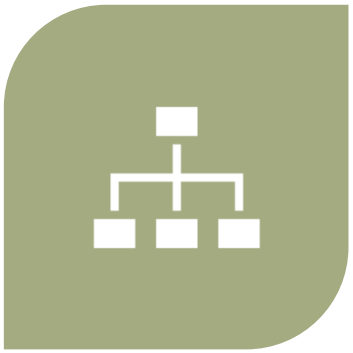
THE QUESTIONS

- 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?

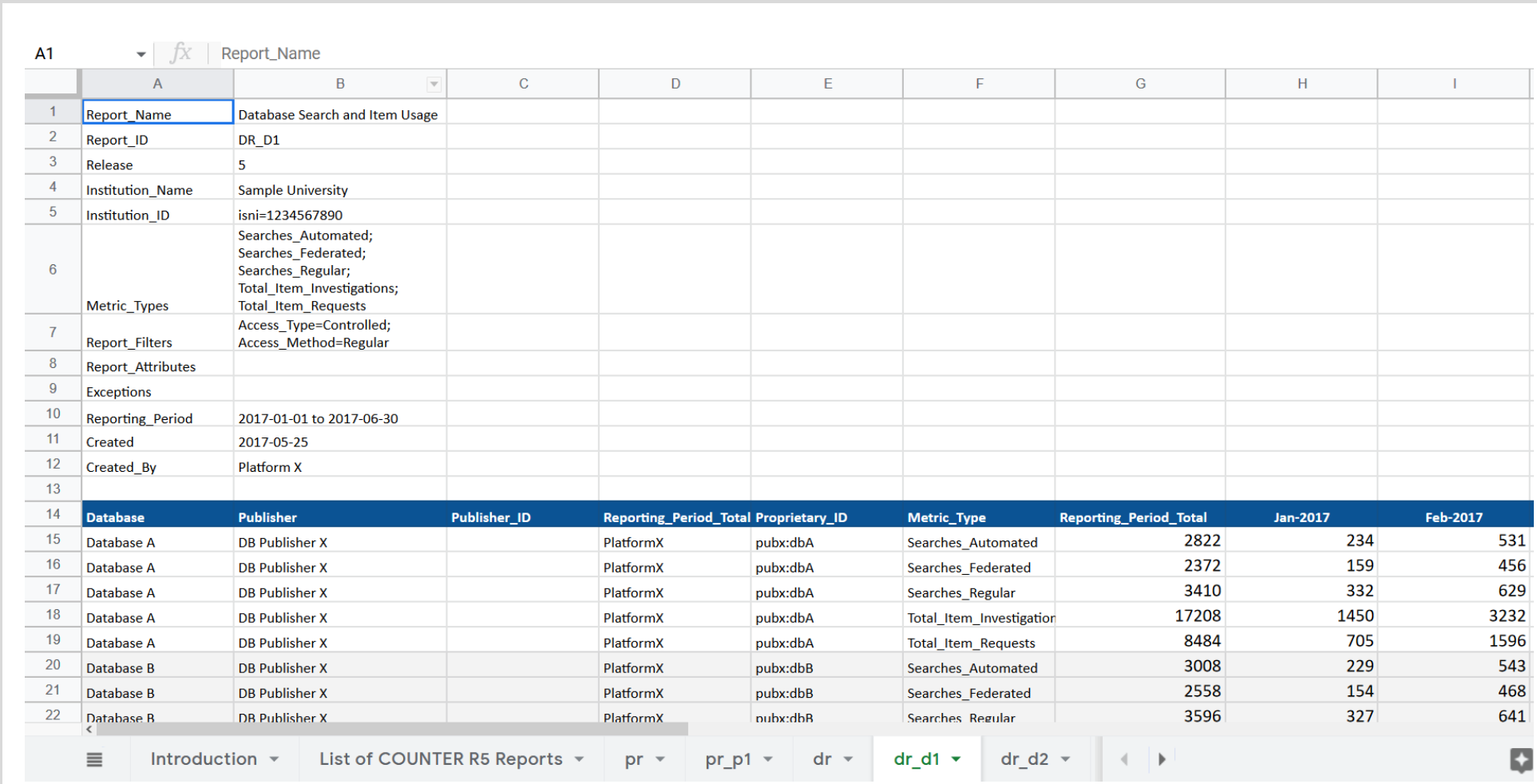
WHAT IS COLLECTIONS ANALYSIS?

DATA-INFORMED

COLLECTION MANAGEMENT



ORGANIZATIONAL PRIORITIES



A1 | fx | Report_Name

	A	B	C	D	E	F	G	H	I
1	Report_Name	Database Search and Item Usage							
2	Report_ID	DR_D1							
3	Release	5							
4	Institution_Name	Sample University							
5	Institution_ID	isni=1234567890							
6	Metric_Types	Searches_Automated; Searches_Federated; Searches_Regular; Total_Item_Investigations; Total_Item_Requests							
7	Report_Filters	Access_Type=Controlled; Access_Method=Regular							
8	Report_Attributes								
9	Exceptions								
10	Reporting_Period	2017-01-01 to 2017-06-30							
11	Created	2017-05-25							
12	Created_By	Platform X							
13									
14	**Database**	**Publisher**	**Publisher_ID**	**Reporting_Period_Total**	**Proprietary_ID**	**Metric_Type**	**Reporting_Period_Total**	**Jan-2017**	**Feb-2017**
15	Database A	DB Publisher X		PlatformX	pubx:dbA	Searches_Automated	2822	234	531
16	Database A	DB Publisher X		PlatformX	pubx:dbA	Searches_Federated	2372	159	456
17	Database A	DB Publisher X		PlatformX	pubx:dbA	Searches_Regular	3410	332	629
18	Database A	DB Publisher X		PlatformX	pubx:dbA	Total_Item_Investigation	17208	1450	3232
19	Database A	DB Publisher X		PlatformX	pubx:dbA	Total_Item_Requests	8484	705	1596
20	Database B	DB Publisher X		PlatformX	pubx:dbB	Searches_Automated	3008	229	543
21	Database B	DB Publisher X		PlatformX	pubx:dbB	Searches_Federated	2558	154	468
22	Database B	DB Publisher X		PlatformX	pubx:dbB	Searches_Regular	3596	327	641

Introduction | List of COUNTER R5 Reports | pr | pr_p1 | dr | dr_d1 | dr_d2

THE PROJECT

CITATION ANALYSIS

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

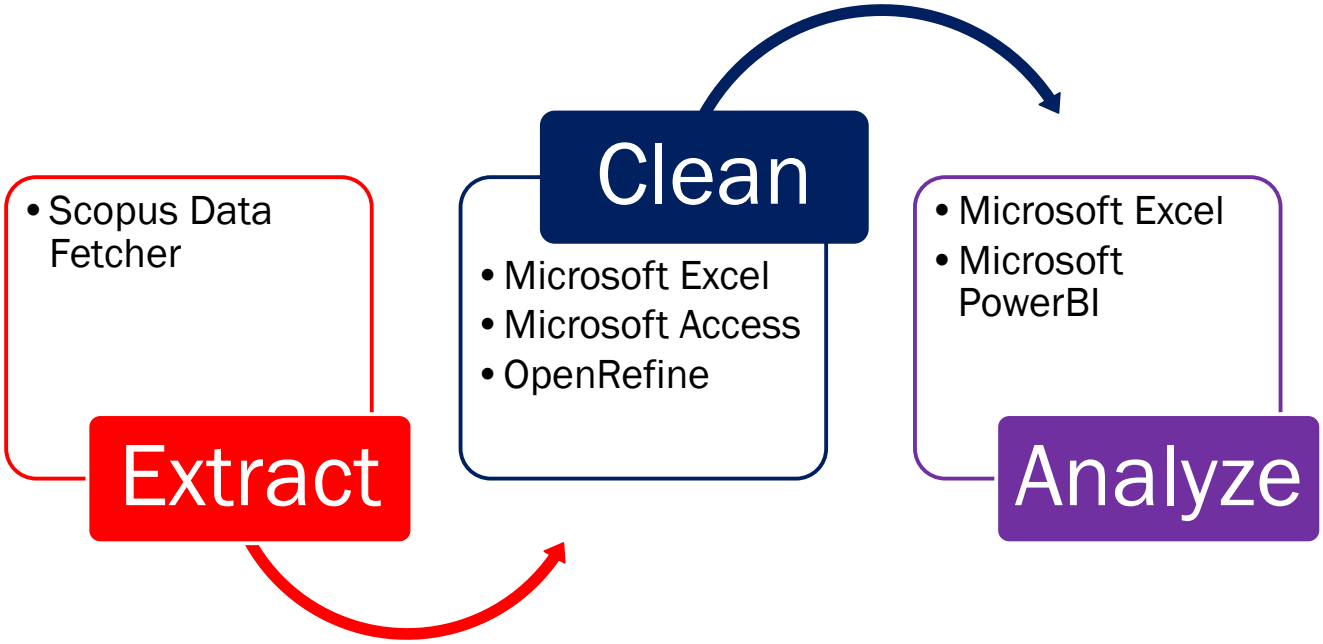
UNB STATS

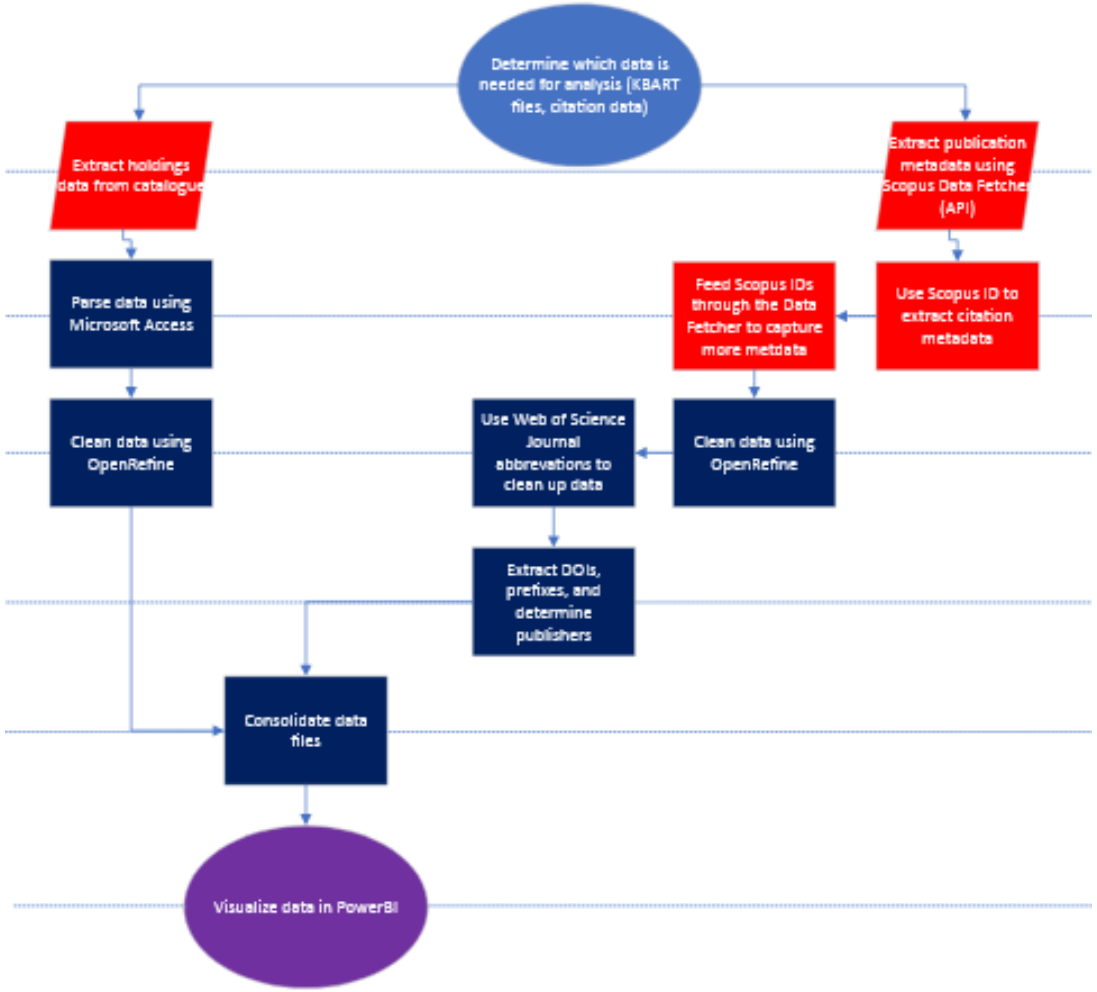
75	Programs
10,000+	Students
1,200+	Graduate students
900+	Faculty
70%	Sponsored research in NB
$55.67M	External research funding

THE PROJECT

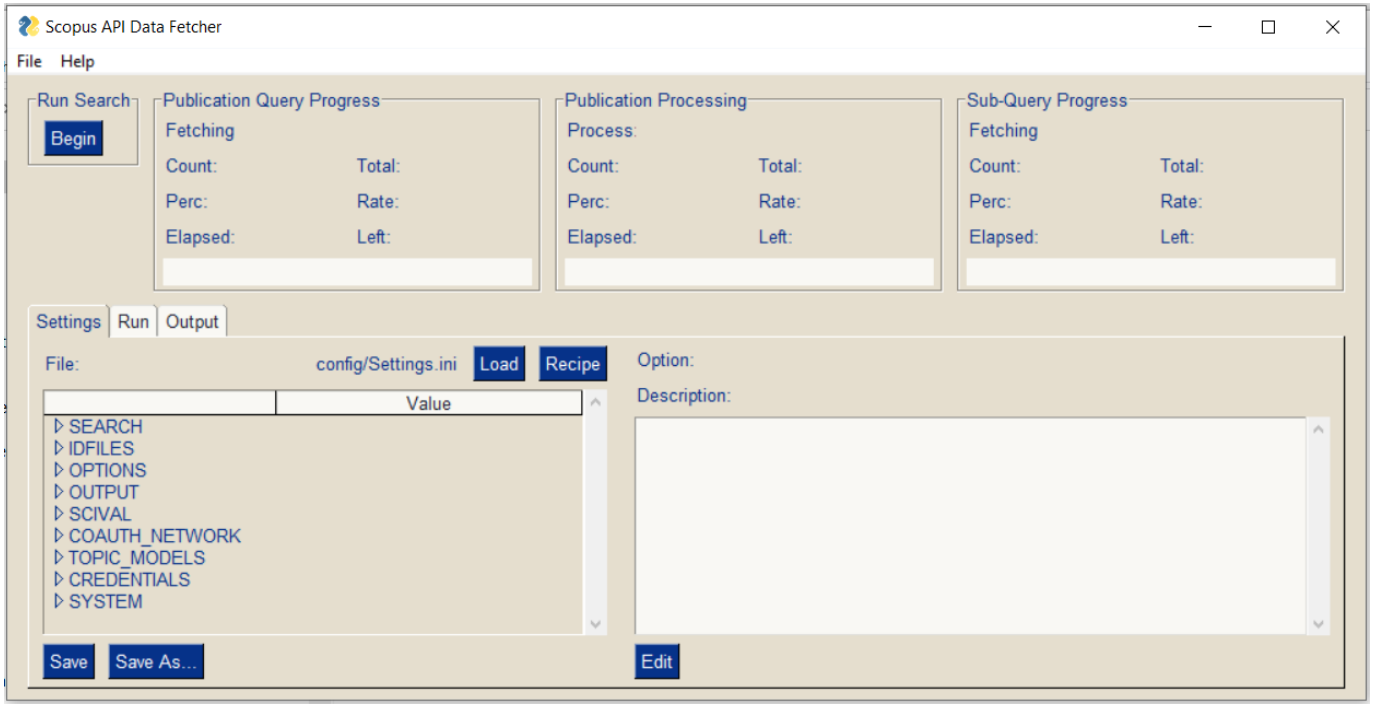
- Scopus data source
- 3 years of data (2019-2021)
- N = 120,982 journal publications
- 68,919 had DOIs
- Includes graduate students and faculty research
- Work in progress

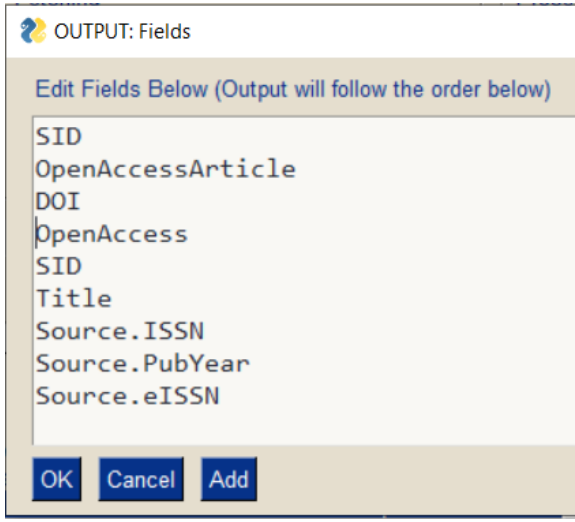
PROCESS


• Scopus Data Fetcher
Extract
Clean
• Microsoft Excel
• Microsoft Access
• OpenRefine
• Microsoft Excel
• Microsoft PowerBI
Analyze

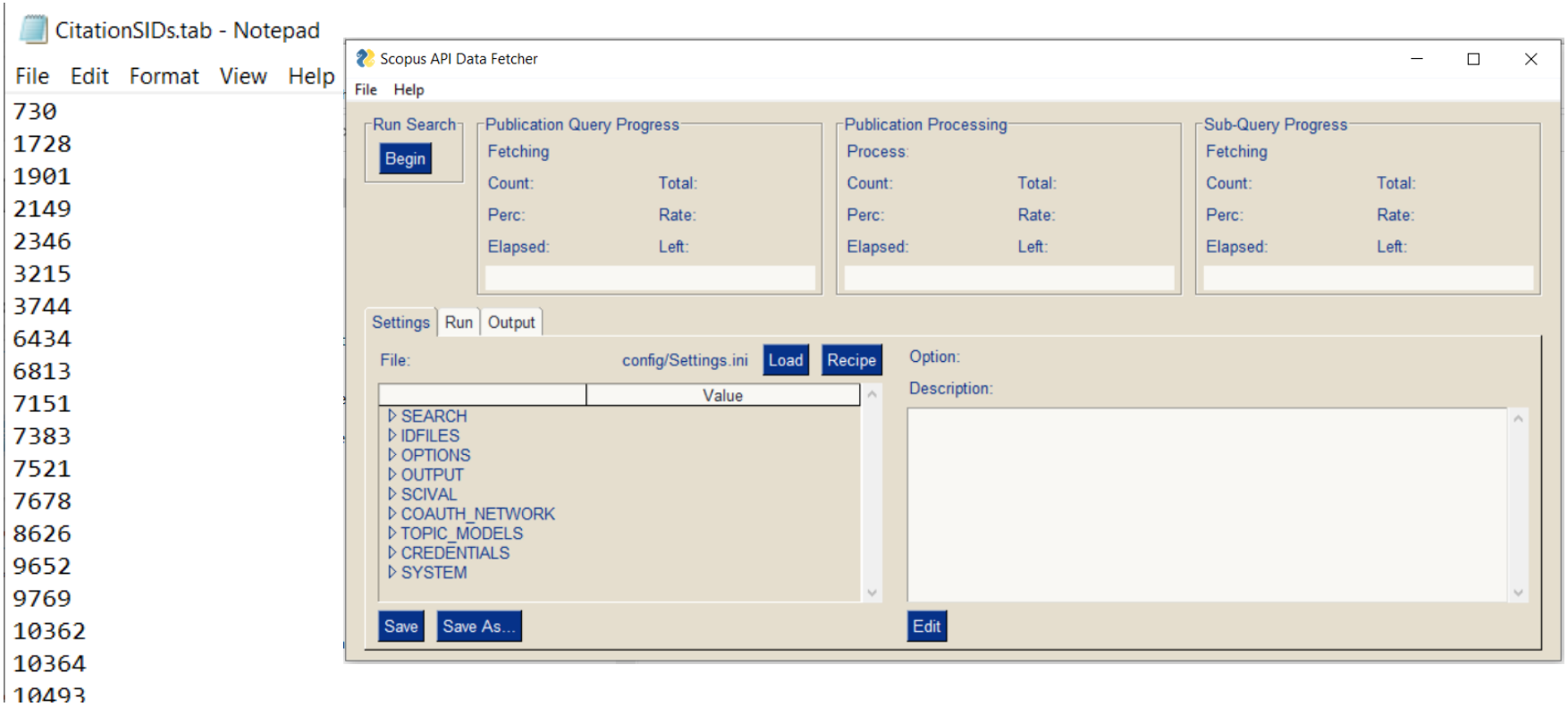

Determine which data is needed for analysis (KBART files, citation data)
Extract holdings data from catalogue
Extract publication metadata using Scopus Data Fetcher (API)
Parse data using Microsoft Access
Feed Scopus IDs through the Data Fetcher to capture more metdata
Use Scopus ID to extract citation metadata
Clean data using OpenRefine
Use Web of Science Journal abbrevations to clean up data
Clean data using OpenRefine
Extract DOIs, prefixes, and determine publishers
Consolidate data files
Visualize data in PowerBI

STEP 1: EXTRACT YOUR PUBLICATION DATA

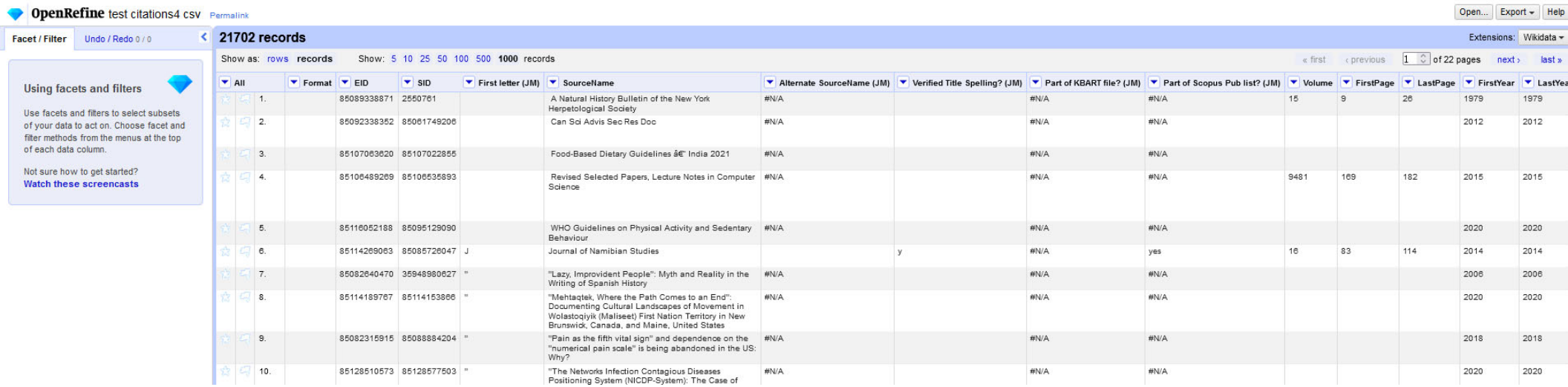




STEP 2: RUN IT BACK THROUGH THE DATA FETCHER



STEP 3: CLEAN YOUR PUBLICATION DATA



All		Format	EID	SID	First letter (JM)	SourceName	Alternate SourceName (JM)	Verified Title Spelling? (JM)	Part of KBART file? (JM)	Part of Scopus Pub list? (JM)	Volume	FirstPage	LastPage	FirstYear	LastYea
1.			85089338871	2550761		A Natural History Bulletin of the New York Herpetological Society	#N/A		#N/A	#N/A	15	9	26	1979	1979
2.			85092338352	85061749206		Can Sci Advis Sec Res Doc	#N/A		#N/A	#N/A				2012	2012
3.			85107063620	85107022855		Food-Based Dietary Guidelines â€“ India 2021	#N/A		#N/A	#N/A					
4.			85106489269	85106535893		Revised Selected Papers, Lecture Notes in Computer Science	#N/A		#N/A	#N/A	9481	169	182	2015	2015
5.			85116052188	85095129090		WHO Guidelines on Physical Activity and Sedentary Behaviour	#N/A		#N/A	#N/A				2020	2020
6.			85114269063	85085726047	J	Journal of Namibian Studies		y	#N/A	yes	16	83	114	2014	2014
7.			85082640470	35948980627	"	"Lazy, Improvident People": Myth and Reality in the Writing of Spanish History	#N/A		#N/A	#N/A				2006	2006
8.			85114189767	85114153866	"	"Mehtaqtek, Where the Path Comes to an End": Documenting Cultural Landscapes of Movement in Wolastoqiyik (Maliseet) First Nation Territory in New Brunswick, Canada, and Maine, United States	#N/A		#N/A	#N/A				2020	2020
9.			85082315915	85088884204	"	"Pain as the fifth vital sign" and dependence on the "numerical pain scale" is being abandoned in the US: Why?	#N/A		#N/A	#N/A				2018	2018
10.			85128510573	85128577503	"	"The Networks Infection Contagious Diseases Positioning System (NICDP-System): The Case of	#N/A		#N/A	#N/A				2020	2020

STEP 4: KEEP CLEANING YOUR DATA

Web of Science Journal Abbreviations List

Journal List

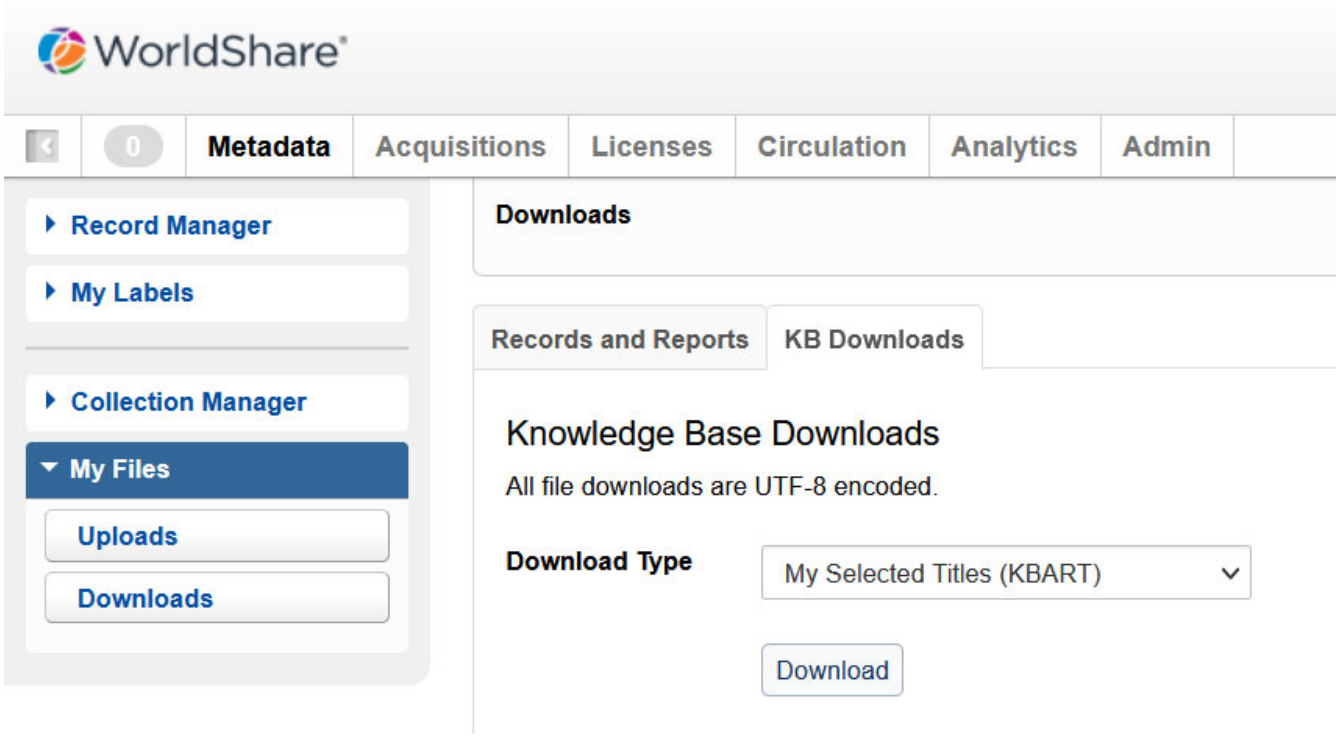
J-FOR-JOURNAL OF SCIENCE & TECHNOLOGY FOR FOREST PRODUCTS AND PROCESSES
J-FOR
JAAGSIEKTE SHEEP RETROVIRUS AND LUNG CANCER
CURR TOP MICROBIOL
JAAPA-JOURNAL OF THE AMERICAN ACADEMY OF PHYSICIAN ASSISTANTS
JAAPA-J AM ACAD PHYS
JACC-CARDIOVASCULAR IMAGING
JACC-CARDIOVASC IMAG
JACC-CARDIOVASCULAR INTERVENTIONS
JACC-CARDIOVASC INTE
JACC-HEART FAILURE
JACC-HEART FAIL
JACEP-JOURNAL OF THE AMERICAN COLLEGE OF EMERGENCY PHYSICIANS
JACEP-J AM COLL EMER
JACKED TUNNEL DESIGN AND CONSTRUCTION
GEOTECH SP
JACKIE ROBINSON: RACE, SPORTS AND THE AMERICAN DREAM

JACKSON SCHOOL PUBLICATIONS IN INTERNATIONAL STUDIES
J S PUB INT
JACOBS FOUNDATION SERIES ON ADOLESCENCE
JACOBS FOUND SER ADO
JACQUES LEFEVRE D'ETAPLES 1450?-1536
CCC RENAISS
JACQUES MARITAIN AND THE MODERN ERA

JADERNA ENERGIE
JAD ENERG
JAHN-TELLER EFFECT: FUNDAMENTALS AND IMPLICATIONS FOR PHYSICS AND CHEMISTRY
SPRINGER SER CHEM PH
JAHRBUCH DER BERLINER MUSEEN
JAHRB BERLIN MUSEEN
JAHRBUCH DER DEUTSCHEN SCHILLERGESELLSCHAFT
JAHRB DEUT SCHILLER
JAHRBUCH DES FREIEN DEUTSCHEN HOCHSTIFTS
JAHRB FREI DEUT HOCH
JAHRBUCH FUR INTERNATIONALE GERMANISTIK

STEP 5: WORK WITH DOIS

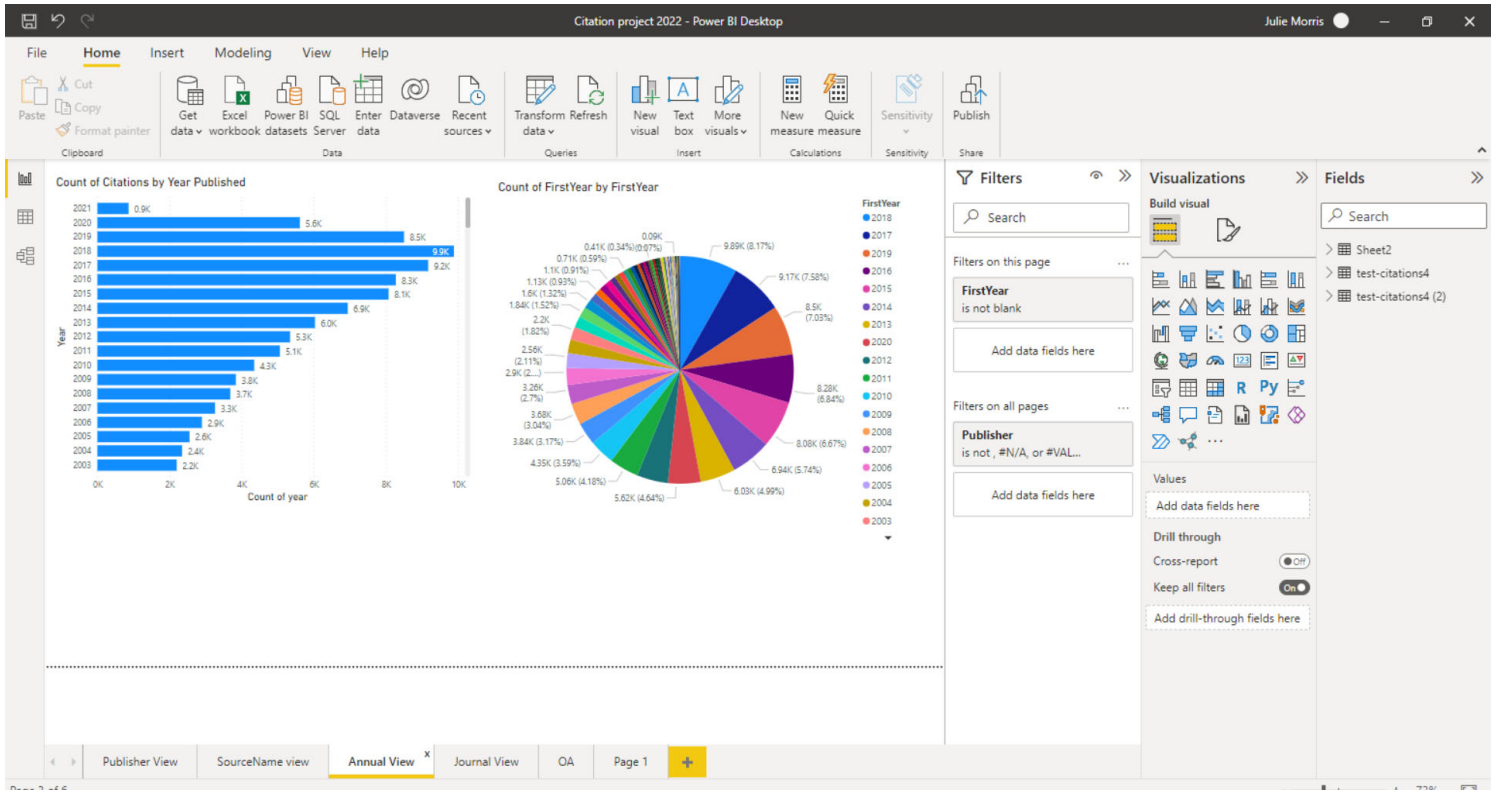
STEP 6: EXTRACT HOLDINGS AND CLEAN THE DATA



STEP 7: COMBINE YOUR DATA

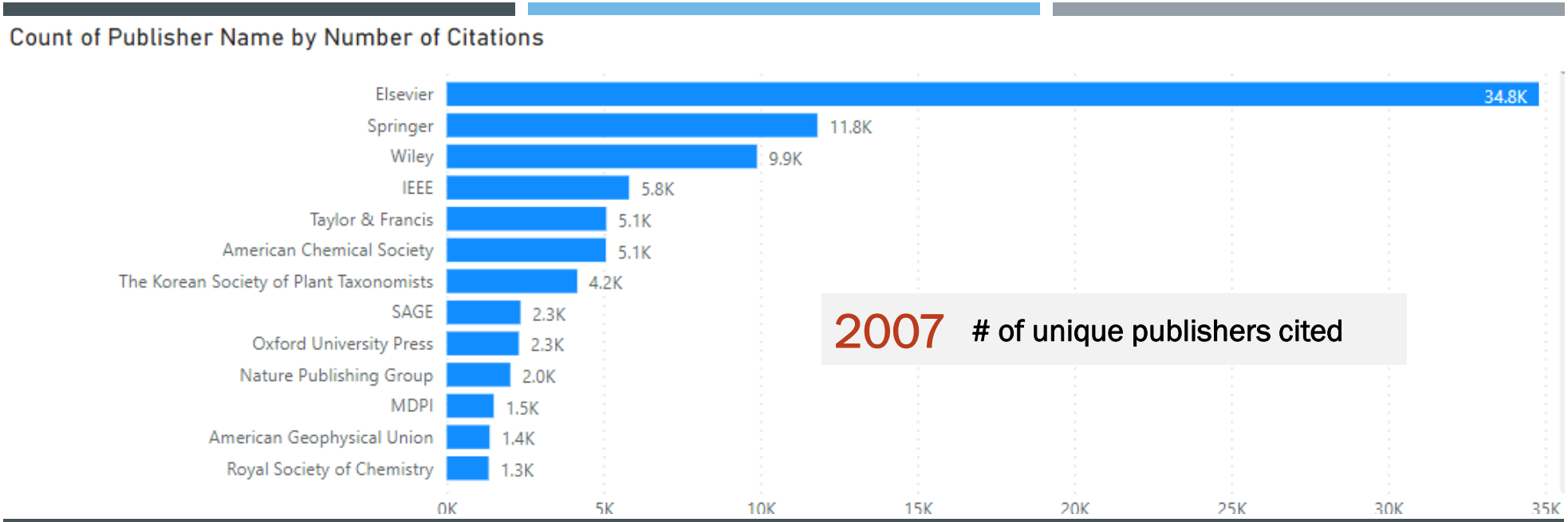
B	C	E	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
EID	SID	Source	Part of	Part of	Volume	FirstPa	LastPag	FirstYe	LastYea	Title	Text	DOI	DOI Pre	Publish	Alterna	ISSN	e-ISSN	Open A	Source	Author	1 Genera	11 Agricultural and Biological Science	12 Arts and Humaniti es	13 Biochemistry, Genetic and Molecular Biolog
8.51E+10	8.49E+10	100E Ecole	no	no		149	197	2015	2015		Burgess, C	10.1093/a	10.1093	Oxford Un	Oxford University Press				France, Ju	["Burgess C.P."]				
8.51E+10	8.51E+10	101 gresel	no	no				2012	2012		Nedelcu, Isabela. 2012. 101 greseligramaticale. Bucuresti: Humanitas.								Bucuresti:	["Nedelcu I."]				
8.51E+10	8.51E+10	10th Europ	no	no				2013	2013	Computat	Lloyd, T., Turnock, S., Humphrey, V., Computation of inflow turbulence noise of a tidal turbine									["Lloyd T.", "Turnock S.", "Humphrey V."]				
8.51E+10	8.5E+10	10th Europ	no	no				2013	2013	Tidal turbi	Fleming, C., McIntosh, S., Willden, R.J., Tidal turbine performance in sheared flow. 10th Europ									["Fleming C.", "McIntosh S.", "Willden R.J."]				
8.51E+10	8.5E+10	10th Europ	no	no		1	9	2013	2013	Turbulenc	Blackmore, T., Atten, W., Bahaj, A., Turbulence generation and its effect in LES approximation									["Blackmore T.", "Atten W.", "Bahaj A."]				
8.51E+10	3.18E+10	Forest Ecc	yes	yes	12	19		2005	2005	Turkelboc	Sidle RC, Z	10.1016/j.	10.1016	Elsevier	Elsevier			0		["Sidle R.(	0	Agricultur	0	
8.51E+10	36346410	10th IEEE I	no	no	3	1135	1139	2001	2001	Fuzzy entr	Al-sharhan, S., Karray, F., Gueaieb, W., and Basir, O. (2001). "Fuzzy entropy: A brief								(Melbourr	["Al-Sharhan S.", "Karray F.", "Gueaieb W.", "Basir				
8.51E+10	34770509	Agricultur	yes	no				2001	2001	Influence	Smolikow	10.1016/S	10.1016	Elsevier	Elsevier					["Smolikowski B.", "Puig H.", "Roose E."]				
8.51E+10	8.49E+10	Journal of	yes	yes				2012	2012	A review (	Tang X, Zh	10.1016/S	10.1016	Elsevier	Elsevier	10010742	0	0		["Tang X."	0	0	0	
8.51E+10	1.64E+09	Water Scie	yes	yes		462		2004	2004	Effectiven	Spruill TB	10.2134/je	10.2134	American	American Society of Agronomy			DOAJ	Open Access	["Spruill T	0	0	0	
8.51E+10	3.37E+10	10th Inter	no	no				2001	2001	Mechanisi	Andresen, P.L., Angeliu, T.M., Young, L.M., Catlin, W.R., Horn, R.M., Mechanisms an								NACE Lake	["Andresen P.L.", "Angeliu T.M.", "Young L.M.", "C				
8.51E+10	8.5E+10	10th Inter	no	no		277	284	2012	2012	Synthesis	Md. Mazder Rahman and Gerhard W. Dueck. Synthesis of linear nearest neighbor quantum cir									["Mazder Rahman Md.", "Dueck G.W."]				
8.51E+10	8.5E+10	10th Inter	no	no		277	284	2012	2012	Synthesis	M. M. Rahman and G. W. Dueck, Synthesis of linear nearest neighbor quantum circuits, 10th Ir									["Mazder Rahman Md.", "Dueck G.W.", "Rahman M				
8.51E+10	8.49E+10	10th Worl	no	no				2013	2013	Comparis	Brevault, L., Balesdent, M., BÃ©rend, N., & Le Riche, R. (2013). Comparison of different globa									["Brevault L.", "Balesdent M.", "Berend N.", "Le Ri				
8.51E+10	8.51E+10	10-year ec	no	no				2016	2016		Province of New Brunswick. (2016). 10-year education plan: Everyone at their best (								Retrieved	[]				
8.51E+10	8.51E+10	113th CBC	no	no				2013	2013		PRANTY, B. 2013. 113th CBC Florida Regional Summary. New York, USA: National Au								New York,	["PRANTY B."]				
8.51E+10	8E+10	11th Annu	no	no		1045	1048	2010	2010	Recurrent	Mikolov, T., Karafiat, M., Burger, L., Cernocky, J., and Khudanpur, S. (2010). Recurre								Chiba, Jap	["Mikolov T.", "Karafiat M.", "Burget L.", "ernocky				
8.51E+10	8E+10	11th Annu	no	no		1045	1048	2010	2010	Recurrent	Mikolov, T., Karafiat, M., Burger, L., Cernocky, J., and Khudanpur, S. (2010). Recurre								Chiba, Jap	["Mikolov T.", "Karafiat M.", "Burget L.", "ernocky				
8.51E+10	8.51E+10	11th IADIS	no	no				2018	2018	Conceptu	K. Sultan, O. Ruhi, and R. Lakhani. CONCEPTUALIZING BLOCKCHAINS: CHARACTERISTICS & APP									["Sultan K.", "Ruhi O.", "Lakhani R."]				
8.51E+10	8.51E+10	11Th IEEE	no	no				2020	2020	Using bert	Heidari, M., Jones, J.H.: Using bert to extract topic-independent sentiment features for social									["Heidari M.", "Jones J.H."]				
8.51E+10	8.51E+10	12th {USEI	no	no		383	400	2016	2016	Don't Get	David Lion, Adrian Chiu, Hailong Sun, Xin Zhuang, Nikola Grcevski, and Ding Yuan. Don't Get Ca									["Lion David", "Chiu Adrian", "Sun Hailong", "Zhua				
8.51E+10	8.5E+10	12th Annu	no	no		1	6	2015	2015	Help your	Hassan, M.A., Xiao, M., Wei, Q., Chen, S., Help your mobile applications with fog co								IEEE	["Hassan M.A.", "Xiao M.", "Wei Q.", "Chen S."]				
8.51E+10	8.5E+10	12th ICITS	no	no		498	500	2017	2017	Ethereum	W. Chan and A. Olmsted, "Ethereum transaction graph analysis," in 12th ICITST. IEEE								In 12th Int	["Chan W.", "Olmsted A."]				

STEP 8: RUN ANALYSIS USING POWER BI



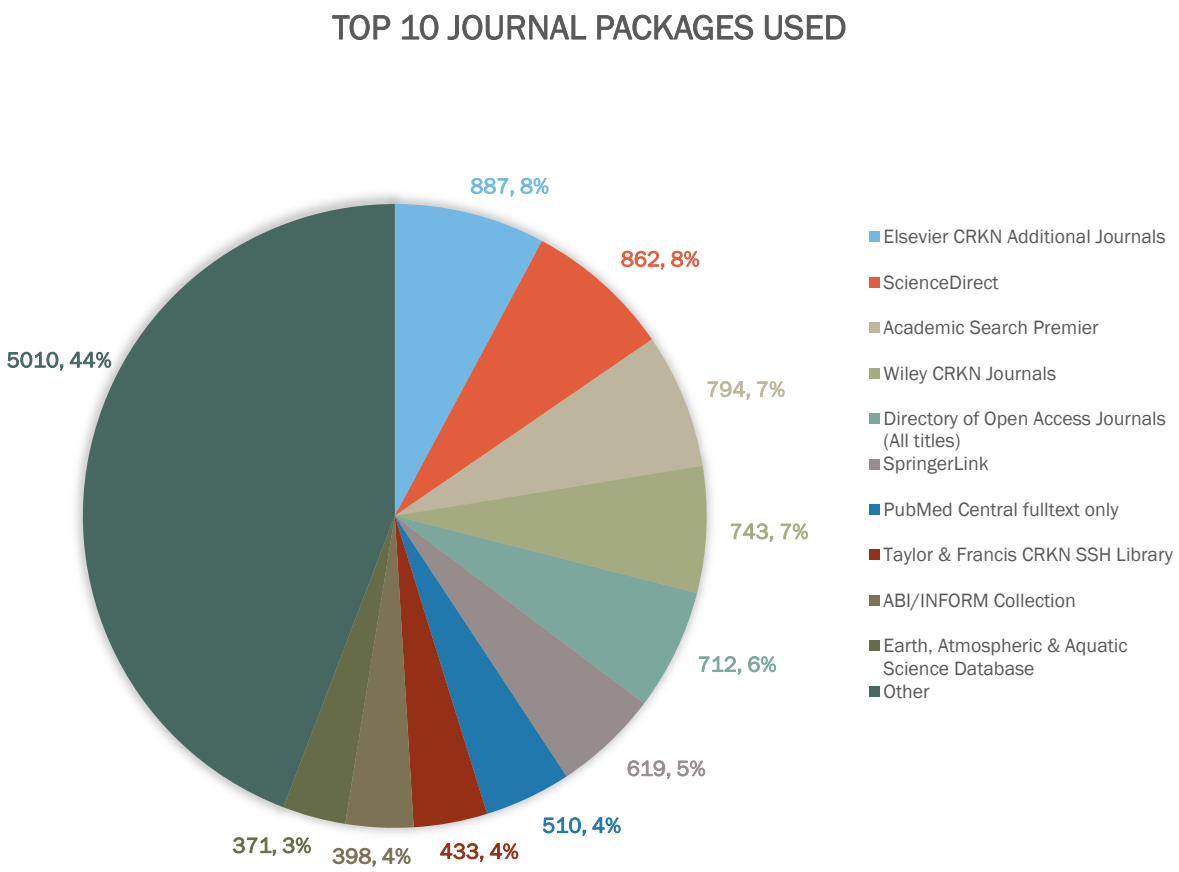
RESULTS

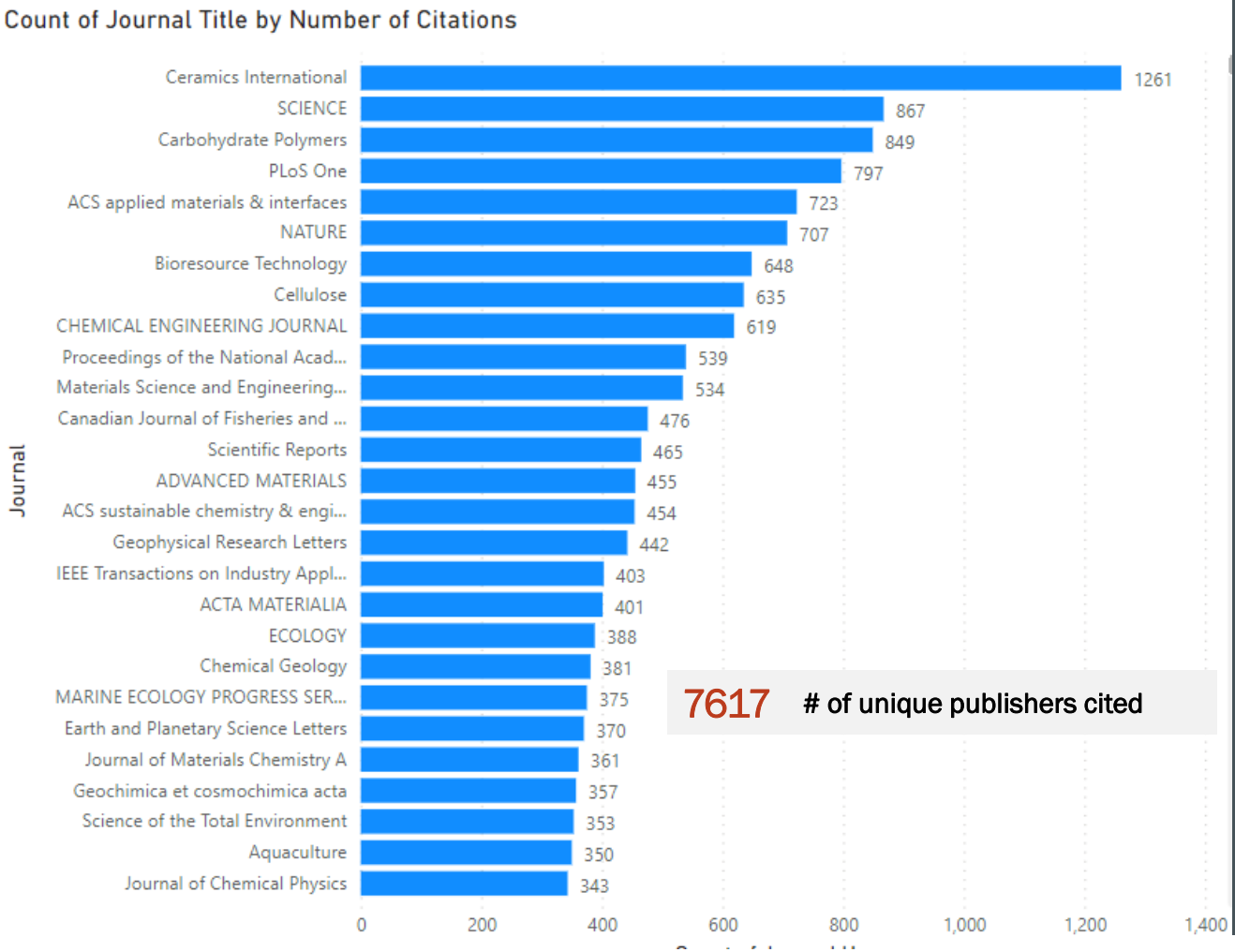
UNB CITATION ANALYSIS PROJECT



WHAT PUBLISHERS ARE WE USING?

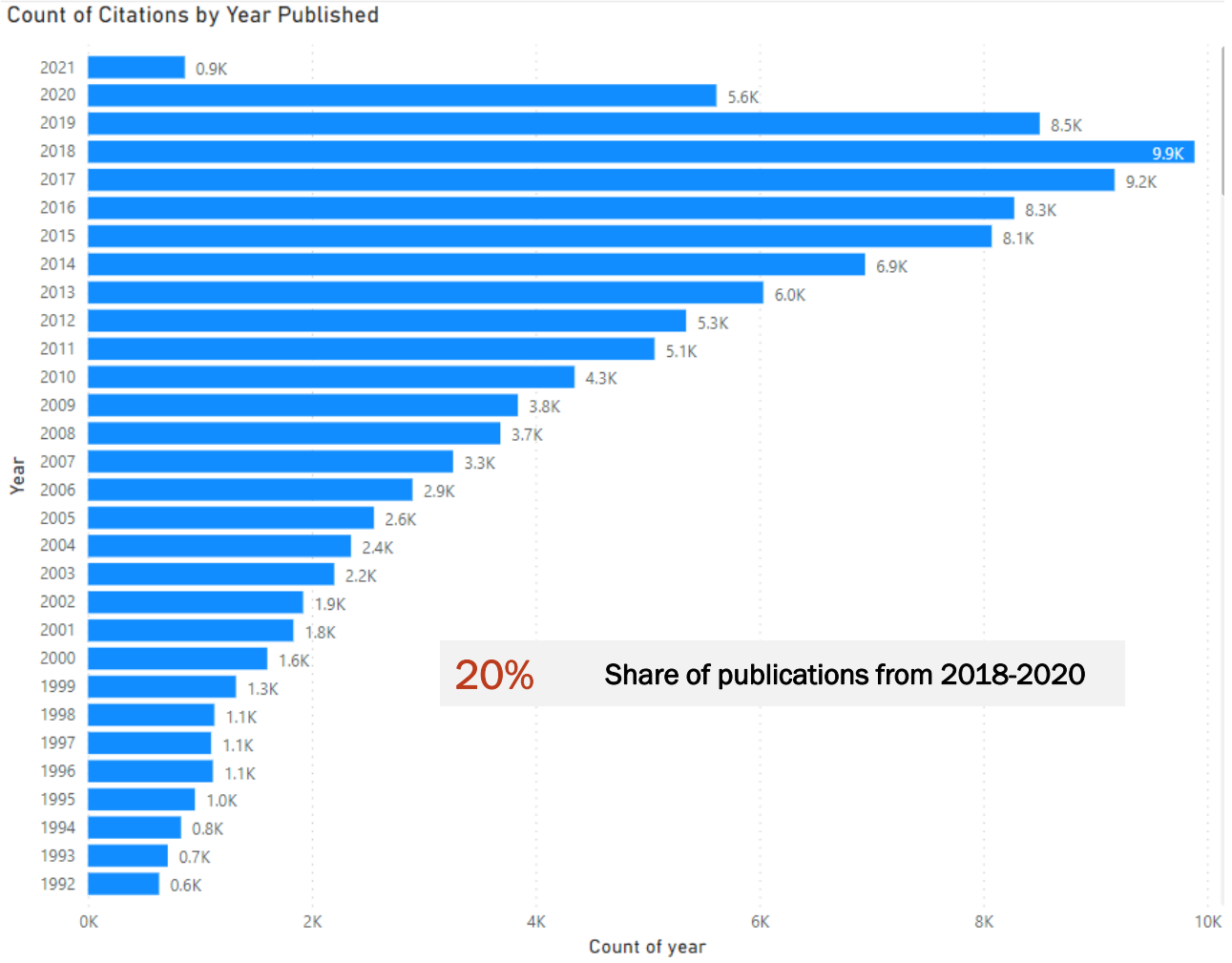
WHAT JOURNAL PACKAGES ARE WE USING?

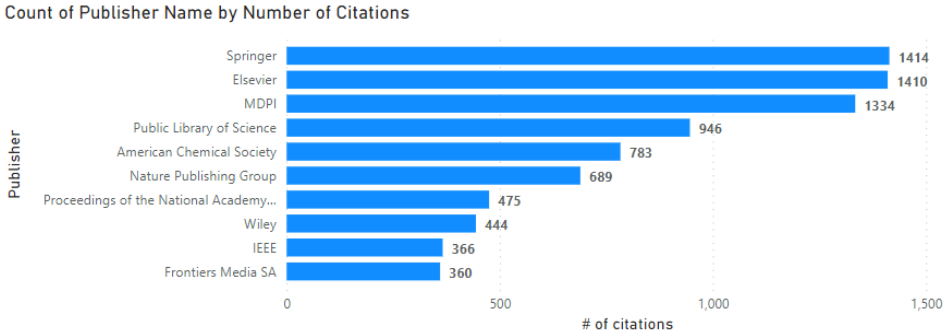


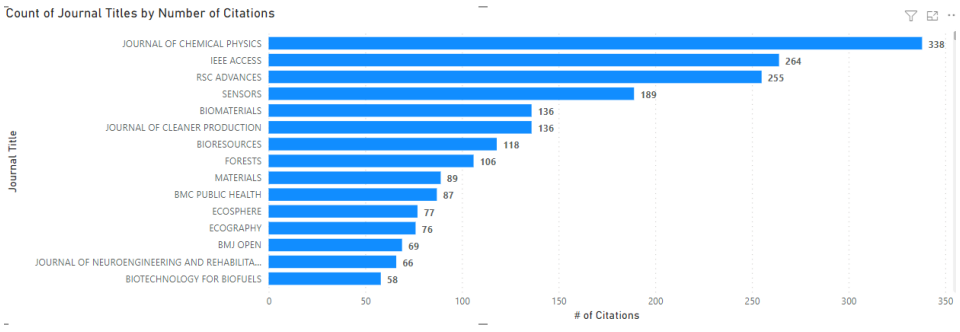


WHAT JOURNALS ARE WE USING?

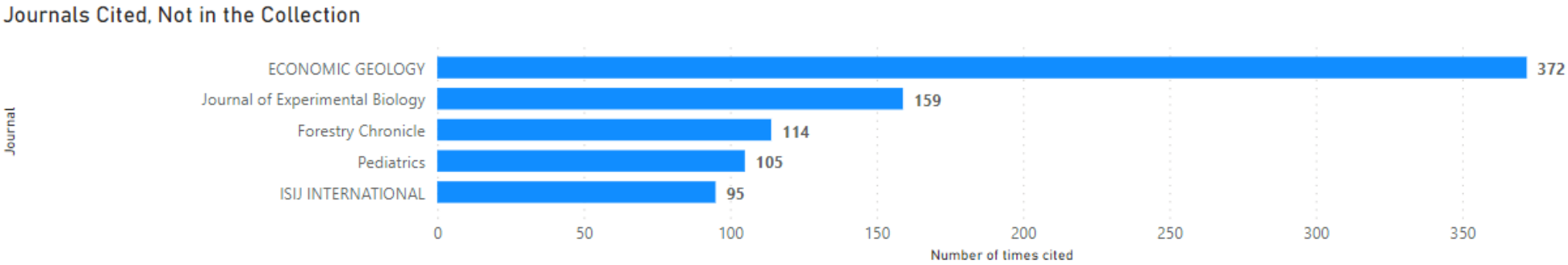
WHAT YEAR WERE MOST CITED JOURNAL ARTICLES PUBLISHED?





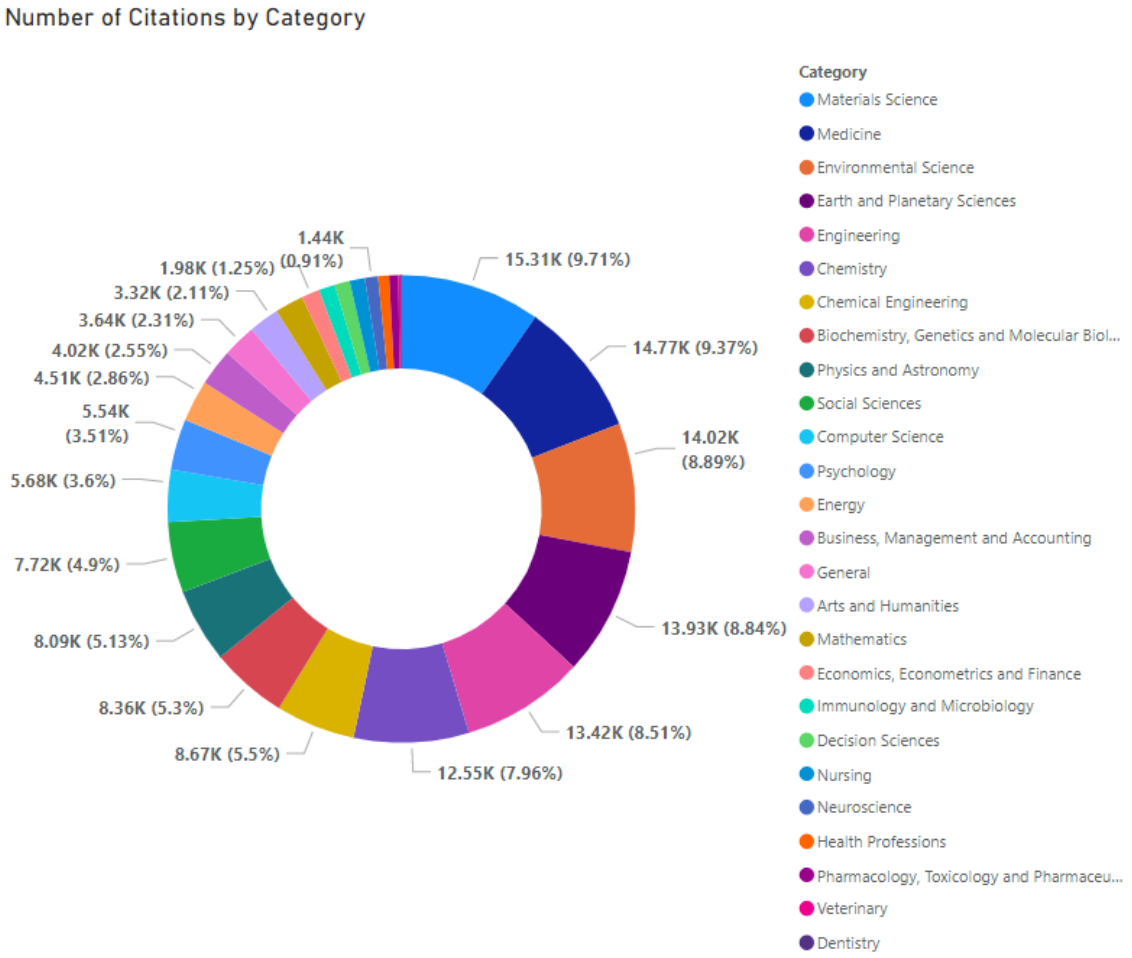


WHICH OA JOURNALS ARE WE USING MOST FREQUENTLY?



WHICH JOURNALS ARE BEING CITED THAT WE DON'T HAVE IN THE COLLECTION?

WHAT FIELDS ARE WE CITING MOST FREQUENTLY?

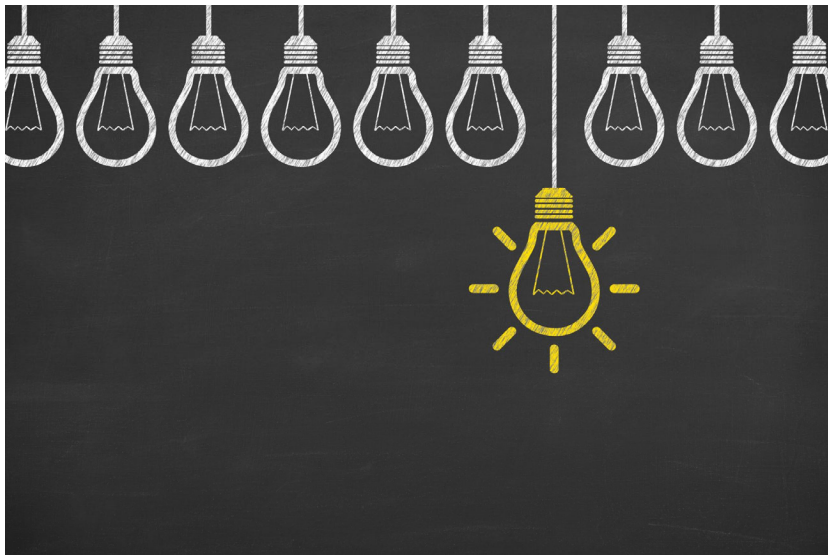


FEEDBACK

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

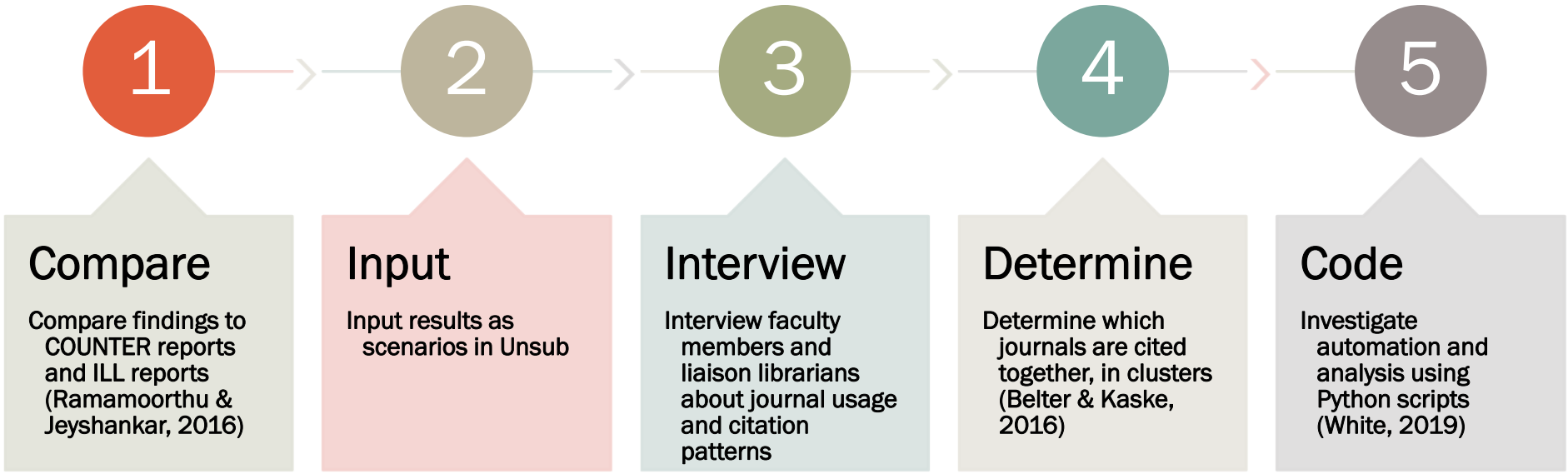
IF I COULD DO IT OVER AGAIN

- Start with more data fields
- Trial run with a more limited scope
- Find help for data cleanup process

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
2
3
4
5
Compare
Compare findings to COUNTER reports and ILL reports (Ramamoorthu & Jeyshankar, 2016)
Input
Input results as scenarios in Unsub
Interview
Interview faculty members and liaison librarians about journal usage and citation patterns
Determine
Determine which journals are cited together, in clusters (Belter & Kaske, 2016)
Code
Investigate automation and analysis using Python scripts (White, 2019)

QUESTIONS, COMMENTS & FEEDBACK

JULIE.MORRIS@UNB.CA

REFERENCES

Belter, C. W., & Kaske, N. K. (2016). Using bibliometrics to demonstrate the value of library journal collections. *College and Research Libraries*, *77*(4), 410–422. https://doi.org/10.5860/crl.77.4.410

deVries, S., Kelly, R., & Storm, P. M. (2010). Moving Beyond Citation Analysis: How Surveys and Interviews Enhance, Enrich, and Expand Your Research Findings. *College & Research Libraries*, *71*(5), 456–466. https://doi.org/10.5860/crl-45r1

Hoffman, K., & Doucette, L. (2012). A Review of Citation Analysis Methodologies for Collection Management. *College & Research Libraries*, *73*(4). https://doi.org/10.5860/crl-254

Pastva, J., Shank, J., Gutzman, K. E., Kaul, M., & Kubilius, R. K. (2018). Capturing and Analyzing Publication, Citation, and Usage Data for Contextual Collection Development. *Serials Librarian*, *74*(1–4), 102–110.

Ramamoorthu, P., & Jeyshankar, R. (2016). Is there a correlation between access to journals and their productive use? *Annals of Library and Information Science*, *63*, 24–30.

Tucker, C. (2013). Analyzing faculty citations for effective collection management decisions. *Library Collections, Acquisitions and Technical Services*, *37*(1–2), 19–33. https://doi.org/10.1016/j.lcats.2013.06.001

White, P. P. (2019). Using Data Mining for Citation Analysis. *College & Research Libraries*, *80*(1). https://doi.org/10.5860/crl.80.1.76