JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

Appendix 1: Citation Search & Download Protocol

1. Navigate to Web of Science's Advanced Search function. We're searching in the Web of Science Core Collection.

   Fill the search bar with our parameters:

   (SU= (Crystallography OR Geochemistry & Geophysics OR Geology OR Meteorology & Atmospheric Sciences OR Mineralogy OR Mining & Mineral Processing OR Oceanography OR Physical Geography OR Water Resources OR Remote Sensing OR Paleontology)) OR (WC=(Crystallography OR Geochemistry & Geophysics OR Geology OR Geosciences, Multidisciplinary OR Meteorology & Atmospheric Sciences OR Mineralogy OR Oceanography OR Paleontology OR Remote Sensing OR Soil Science OR Water Resources))

   For our purposes, we are restricting the language to English.

   In WoS, "Timespan" and "Publication Range" yield the same results. We're looking for results between 2010-2019.

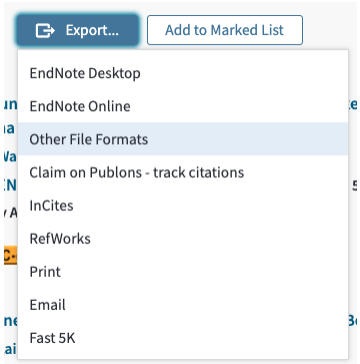
   Run the search.

2. Using the column on the left-hand side of the interface, refine by your institution.

   Limit to Document Type: Article or Review

JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

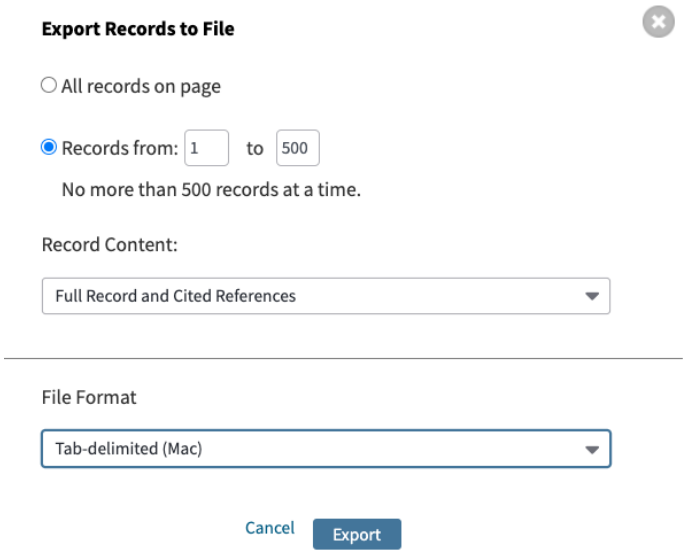
3. Next, we're going to download our results. Users may only download 500 results at a time, so this is a multi-step process.
    a. Click on "Export" above the first result on the page. Indicate "Other File Formats"



    b. For the first download, select records from 1-500.
        i. 1-500
        ii. 501-1000
        iii. 1001-1500
        iv. 1501-2000
        v. 2001-2500
    c. Record Content: Full Record and Cited References
    d. File Format: Tab-delimited UTF-8

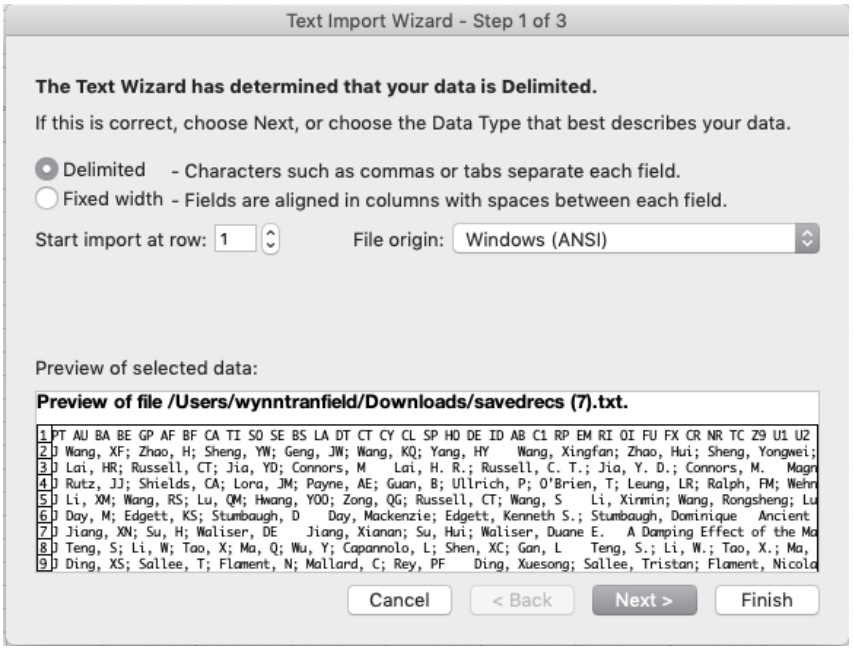
JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

e.

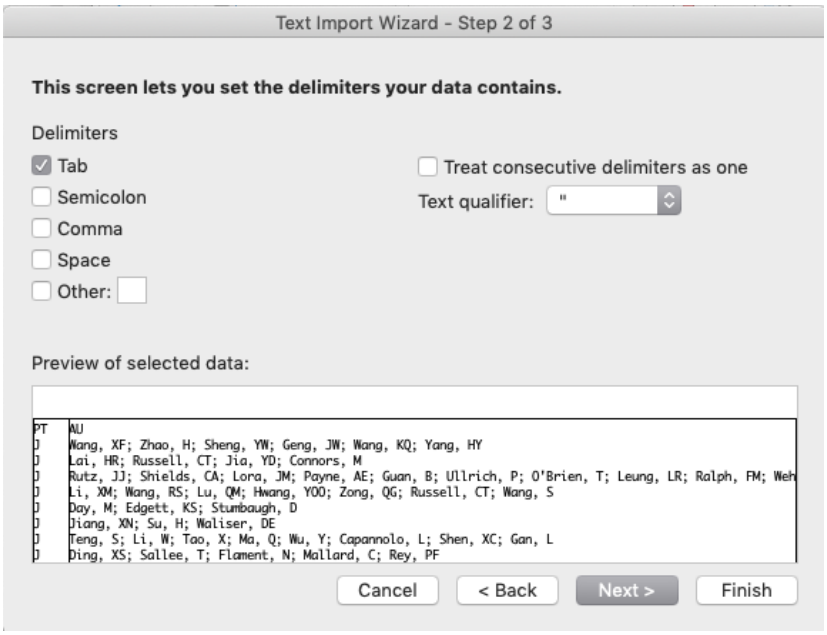


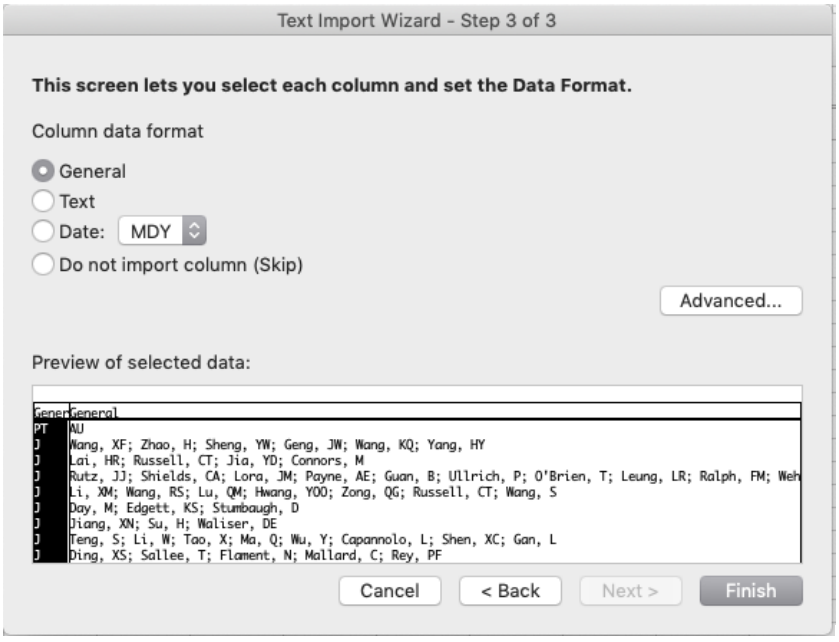
4. Once each pertinent export file is downloaded, you may combine them one at a time.
    a. Open downloaded file in Excel. A Text Import Wizard will appear to guide you.

b.



JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194





5. Add a column to identify the institution
    a. Example: Institution 1, Institution 2, Institution 3, Institution 4

JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

Appendix 2: Open Access Representation

	Frequency by institution			
OA Type	**Inst 1**	**Inst 2**	**Inst 3**	**Inst 4**
DOAJ Gold	1164	449	363	56
Bronze	985	529	477	34
Green Published	440	346	263	10
DOAJ Gold, Green Published	230	109	79	8
DOAJ Gold, Green Accepted	174	67	44	7
Other Gold	169	129	101	9
Green Accepted	162	126	135	10
Green Published, Other Gold	118	119	57	4
DOAJ Gold, Green Accepted, Green Published	93	32	23	2
Green Published, Bronze	131	93	65	4
Green Published, Green Accepted	56	29	35	0
Bronze, Green Accepted	33	26	16	0
Other Gold, Green Published, Green Accepted	33	21	12	1
Green Accepted, Other Gold	18	4	11	0
Other Gold, Bronze	1	1	0	0
Green Accepted, Green Published, Bronze	19	10	10	0
Total OA	3826	2090	1704	145
Total Articles	7439	5002	3862	1344
Percent, OA articles	51.43%	41.78%	44.12%	10.79%

JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

Appendix 3: Data Cleaning

Following our initial calculations, we were curious to understand the remaining 11,280 citations that either lacked a DOI or could not be matched with DOIs through CrossRef. We placed these citations in a separate .CSV file to examine and performed some limited clean up and consolidation in Microsoft Excel. The spreadsheet column "CR" or cited reference was split by comma. Titles were sorted and grouped by frequency. Titles appearing 20 or more times were standardized to match the original data set and were then added back into the main file and analyzed together. Following these steps, were left with 55,563 citations (from 55,580). Considering all 55,563 citations together did not change the list of top 5 cited titles listed in Table 4.

In a specific example, the Bulletin of the Seismological Society of America appeared in the DOI-matched dataset 168 times. After cleaning the variants and abbreviations of the title, it appeared 294 times in the full data set. While this may seem like a large discrepancy, the title was still within the top 50 before and after (moving from #50 to #30).

This process of data cleaning is the least replicable part of our method relying on some individual judgment and the vagaries of journal abbreviations, but in order to be thorough we wanted to explore these results, not just the portion whose DOIs matched through CrossRef. Future studies might skip this step depending on the presence of DOIs in the source cited references data.

Reintroducing the citations without DOIs changed the representation of the 80/20 rule for most of our institutions (Table 3-1 compare to Table 5). Boulder's data set is

JeSLIB 2021; 10(2): e1194
https://doi.org/10.7191/jeslib.2021.1194

larger than the other institutions and saw little change. Berkeley, UCLA, and Houston all shift with the reintroduction of the non-DOI matched data. Further cleaning of the long tail of works cited once would be required to confirm the accuracy and benefit of these additional steps.

Table 3-1: Citation counts and 80/20 rule representation, post-cleaning

	Aggregate	Inst 1	Inst 2	Inst 3	Inst 4
Total Citations	55,563	24,035	16,847	13,079	1,602
Total number of journal titles	9,671	4,231	4,322	2,931	603
% of journals responsible for 80% of citations	9.34%	9.71%	24.32%	19.38%	46.93%