Instructions and advice for downloading data from the Web of Science to reproduce our study

This Supplementary Material accompanies the article:

Lerner, G. A., Williams, G. T., Meredith, E. S., Jenkins, S. F. and Barclay, J. (2023) “How inclusive is volcanology? Insights from global bibliometric analyses”, Volcanica, 6(1). pp. 77-94 doi: 10.30909/vol.06.01.7794.

**S8. Instructions to download datasets from Web of Science**

Web of Science denied a request to make our dataset publicly available. However, the reader can download the same dataset by following these steps or altering the steps to download a dataset of their choosing. Note that this procedure requires institutional access to Web of Science. If you do not have this access, please contact us for assistance.

1. Login to the Web of Science (<https://www.webofscience.com>).
2. Type “volcan\*” into the search bar, searching across ‘All Fields’ within the Web of Science Core Collection. Hit the purple ‘Search’ button when ready. Alternatively you can use this link to perform the same database query <https://www.webofscience.com/wos/woscc/summary/0f203d5f-45d2-4799-8921-ed19ecbf32dd-44dc89fd/relevance/1>



1. After the query finishes running, it will have generated many results to download. At the time of writing, this query generated slightly over 150,000 results (i.e. articles/books/conference proceedings), but only 1000-5000 can be downloaded at once (depending on your selected export option).



1. To start downloading results simply hit the ‘Export’ button and select a format/option for your exports. For this study we exported the ‘Full Record’ for all search results to Excel files, 1000 results at a time. As you can guess this required us to perform the download process 154 times to get all results. With a good internet connection and a focused mind, this tedious task can be completed in under 2 hours.



\*\*\*IMPORTANT TIP\*\*\* before you start downloading results, we highly recommend sorting your results by either date, citations, first author name or publication title. This is important because we have > 153,000 results and the Web of Science will fail to download results that are indexed beyond the 100,000 results point. In other words, if you have already downloaded the first 100,000 results of the entire search, when you go to download the next 1000 results (from 100,001 to 101,000) you will be unable to do so.



One solution to this problem is to order your results in such a way that when you cross the halfway point in downloading (roughly 77,000 results in – or 77 downloads at 1000 results per download) you can then reverse the order of your results to get the other half that remain. Note that this strategy only works whilst total results remain below 200,000, beyond which point you could limit your search to specific ranges of years in order to download all desired data.

1. If you wish to analyze your data using the R scripts provided alongside this study (available at <https://github.com/vharg/Bibli_Volc>), you should be able to use the scripts with minimal editing by saving your oldest and newest .xls file downloads in the ‘oldest\_WoS’ and ‘newest\_WoS’ folders in the ‘raw\_data’ folder.



1. Now you have dozens of .xls files containing 1000 articles each. To analyze all of these efficiently the first R script in the ‘processing\_scripts’ folder ‘Step0.1-binding\_raw\_data’ will read in each of these files and combine their articles into one large .csv file. This file can then be interrogated for whatever analysis you would like to perform. Enjoy.