# Utilizing Google Scholar as a Bibliographic Resource for Publications Search

ESIP 2023
July 18-21, 2023
Irina.Gerasimov@nasa.gov

https://disc.gsfc.nasa.gov/
NASA/Goddard Earth Sciences Data and Information Services Center (GES DISC)

**Irina Gerasimov[1,2], Jerome Alfred[1,2], Andrey Savtchenko[1,2], James Acker[1,2], Binita KC[1,2], Long Pham[1], Jennifer Wei[1]**
[1]Code 619, NASA Goddard Space Flight Center, Greenbelt, MD, USA [2]ADNET Systems Inc., Lanham, MD, USA

## Abstract

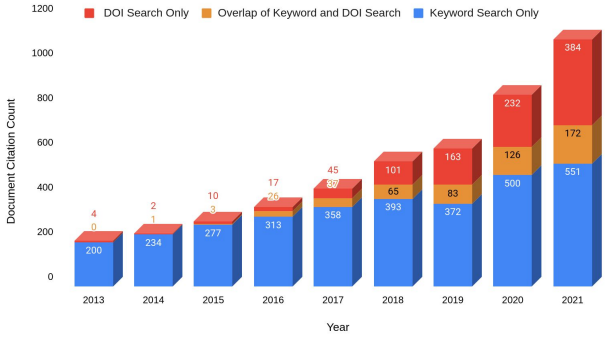
- The Earth Observing System Data and Information System (EOSDIS) began registering dataset Digital Object Identifiers (DOIs) in 2012.
- Automated search methods can be used to collect published works from a wide range of bibliometric databases.
- Presented here is a comparison of publication citation counts obtained from Google Scholar vs established bibliographic databases.
- Also presented is the methodology and open source tool to obtain publication citations from Google Scholar by dataset DOI and metadata keyword search.

## Google Scholar dataset DOI and Keyword Search

Google Scholar citation collection methodology:

- Use a subscription service, SerpAPI, to execute multiple search queries
- Process URLs using Zotero translation server to obtain full citation
- Retrieve document types from Crossref
- Use EOSDIS Common Metadata Repository to compose dataset keyword search queries; for example:
  *"/"OMBRO/" (/"Aura/" | /"OMI/" ) NASA"*, where "Aura" is the satellite platform carrying the "OMI" instrument that collected the data that was used to produce the "OMBRO" dataset (doi: 10.5067/Aura/OMI/DATA2006).

Overlap of Document Citations found by Dataset DOI and Keyword Search



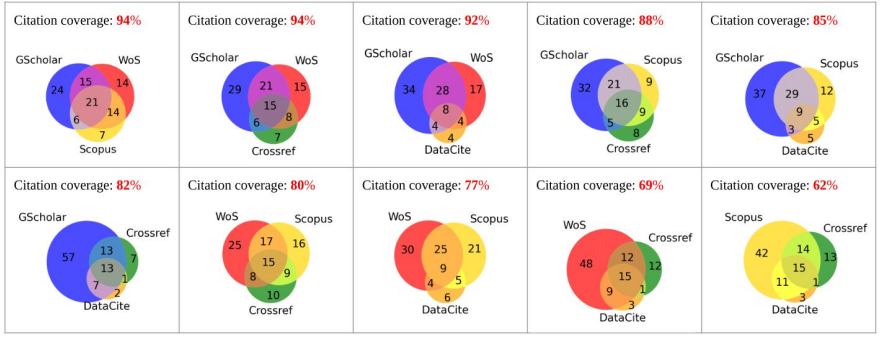

GES DISC results shown on the histogram above show that while authors are increasingly citing datasets by DOI, referencing datasets by their name is still widely practiced

## Comparison of Bibliographic Sources

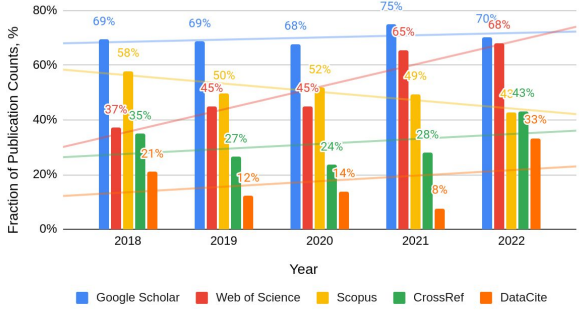
Methodology: Search Web of Science, Scopus, Crossref, Google Scholar, and DataCite for publications referencing each EOSDIS dataset DOI. Retain only citations of books, journal articles, conference papers, dissertations, and reports (exclude Web pages, preprints, discussions, works without DOI, etc.).

- Count of searched EOSDIS DOIs: **11,089**
- Total count of unique publications found: **17,093**
- Count of EOSDIS DOIs cited at least once: **3,012**
- Total count of dataset citations in publications: **24,925**
- Venn Diagrams below show overlap of citations obtained from different combinations of bibliographic sources for citations collected in 2012-2022.

- Google Scholar provides both the largest coverage (70%) and fraction of unique citations (19%) among major bibliographic sources: compar to Web of Science (53% coverage and 10% unique), Scopus (45% an 4%), and Crossref (33% and 3%) as well as EOSDIS dataset registry, DataCite (19% and 1%).
- The temporal trends show that Google Scholar steadily finds 70% of all citations and Web of Science's performance steadily improves, reaching 70% in 2022. Crossref and DataCite also show promising trends with significant performance improvement in 2022.




Yearly Publication Fraction Obtained From Bibliographic Sources




## EOSDIS Dataset Citation Library and Google Scholar Search Tool

✦ EOSDIS dataset citation library is accessible online from an open source Zotero citation manager: https://www.zotero.org/groups/4567966/eosdis_dci/library

✦ The Google Scholar Dataset Citing Documents Search tool is available at: https://github.com/iragerasimov/GSDCDS