



Harvesting Publications for Air Quality and Health Data



1,2,3 Binita KC, 1,2 Irina Gerasimov, 1,2 Armin Mehrabian, 1Mahabal Hegde, 1Jennifer Wei, 1,2Mo Khayat
1NASA GES DISC, 2Adnet Systems Inc., 3 <https://orcid.org/0000-0001-6126-5369>

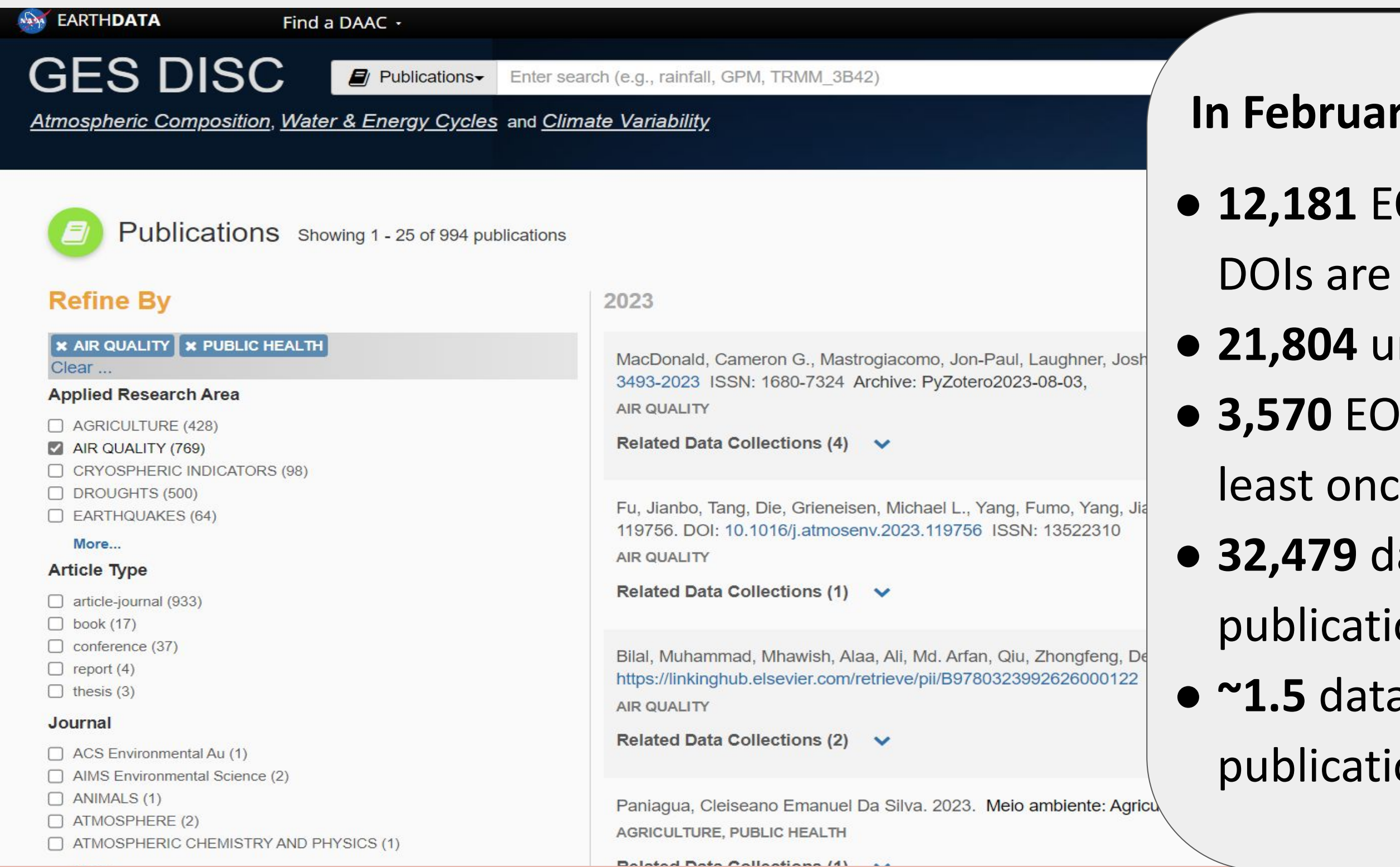
Harvesting Publications using EOSDIS data

We are the [Goddard Earth Sciences Data and Information Services Center \(GES DISC\)](#), one of the 12 **NASA Distributed Active Archive Centers (DAACs)** that manage, archive, and distribute data as part of NASA's Earth Observation Systems Data and Information System (EOSDIS). We provide support for the **archive and distribution** of the data from a range of **satellite observations, models, ground measurements, and field campaigns**.

Multi-disciplinary data holdings include observations and model data of:

- **Atmospheric composition**
- **Water/Energy cycles**
- **Climate variability**

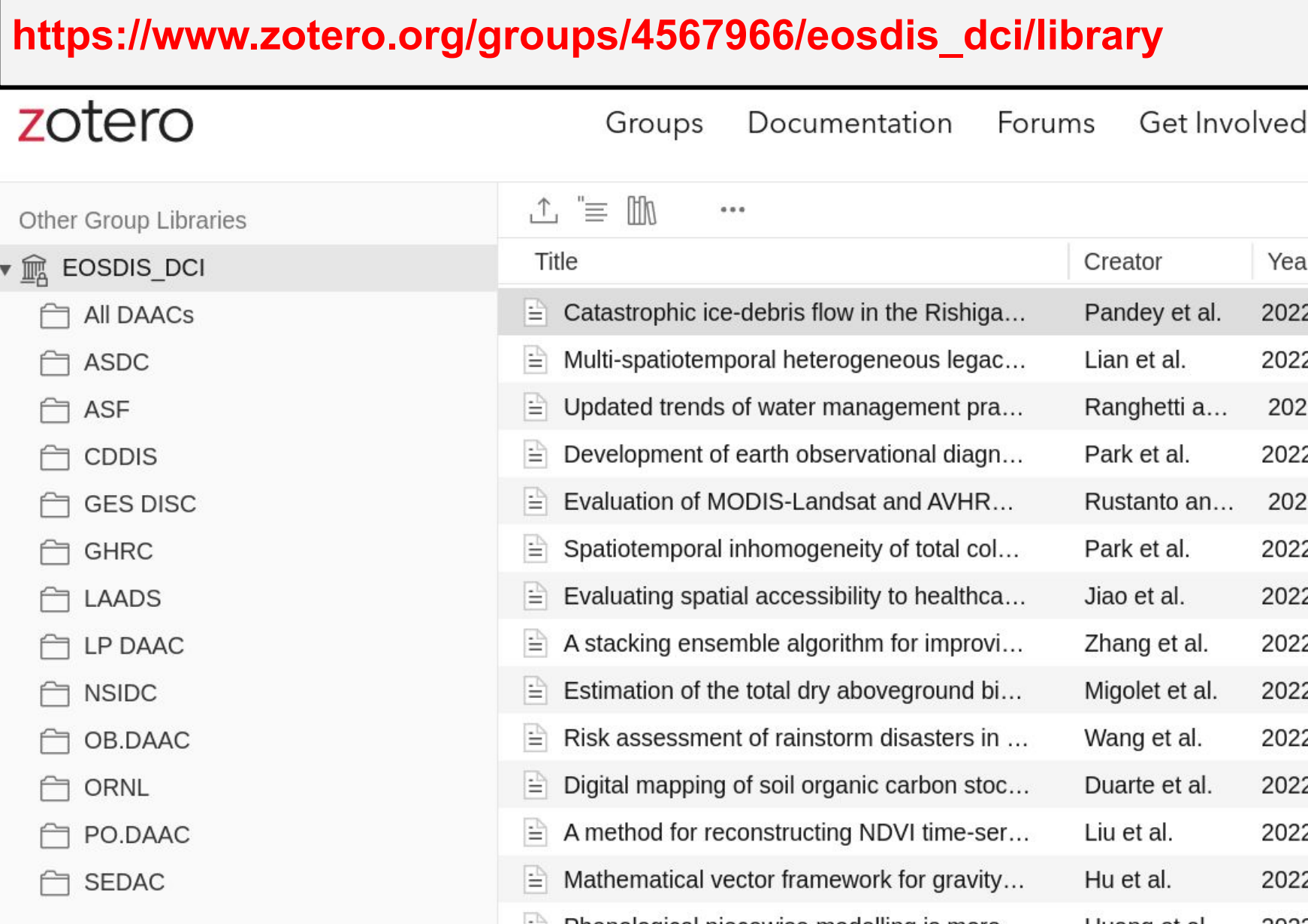
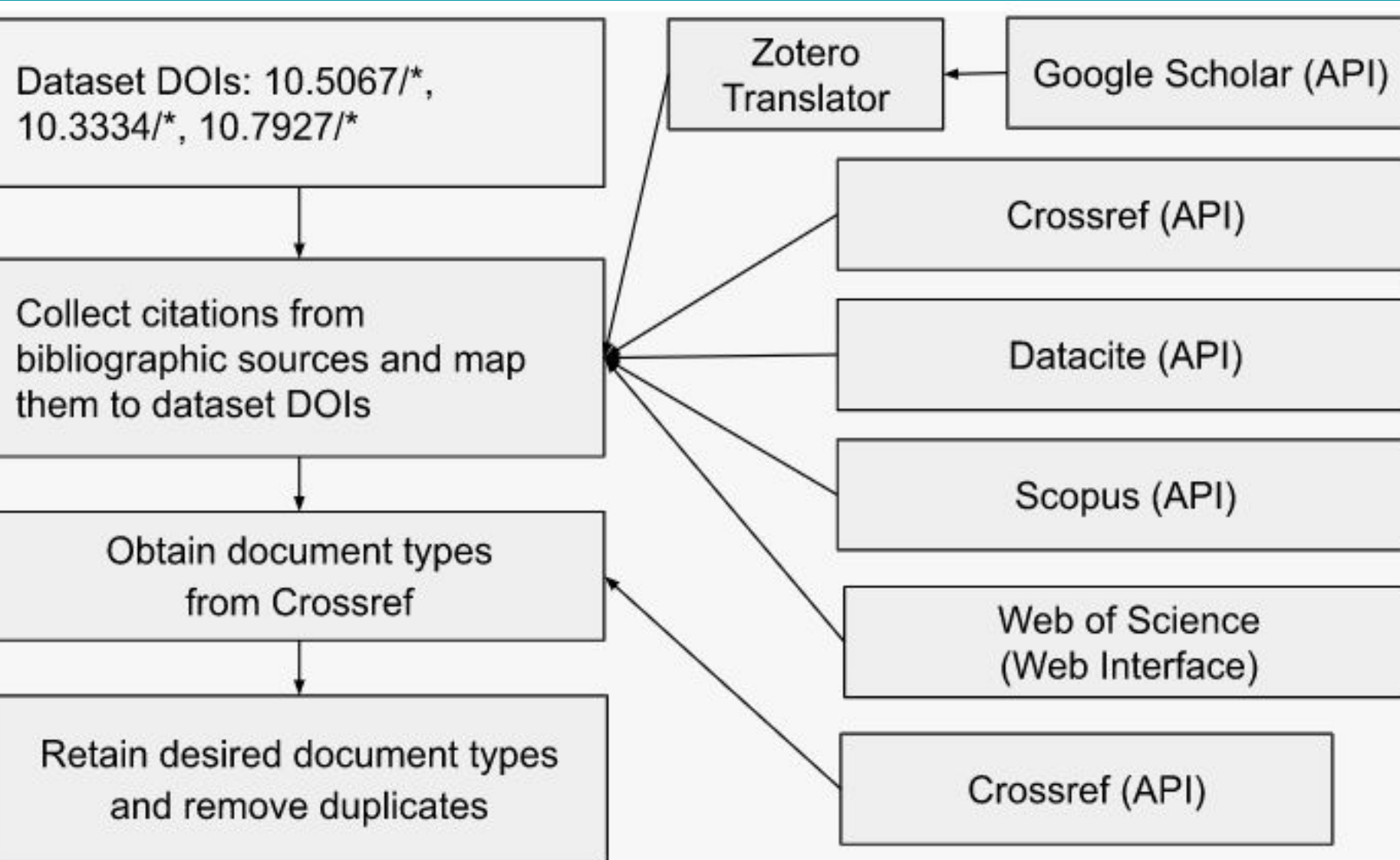
❖ **Data are free and open to the public**



In February 2024:

- **12,181** EOSDIS dataset DOIs are registered
- **21,804** unique publications
- **3,570** EOSDIS DOIs cited at least once
- **32,479** dataset citations in publications
- **~1.5** dataset citations per publication

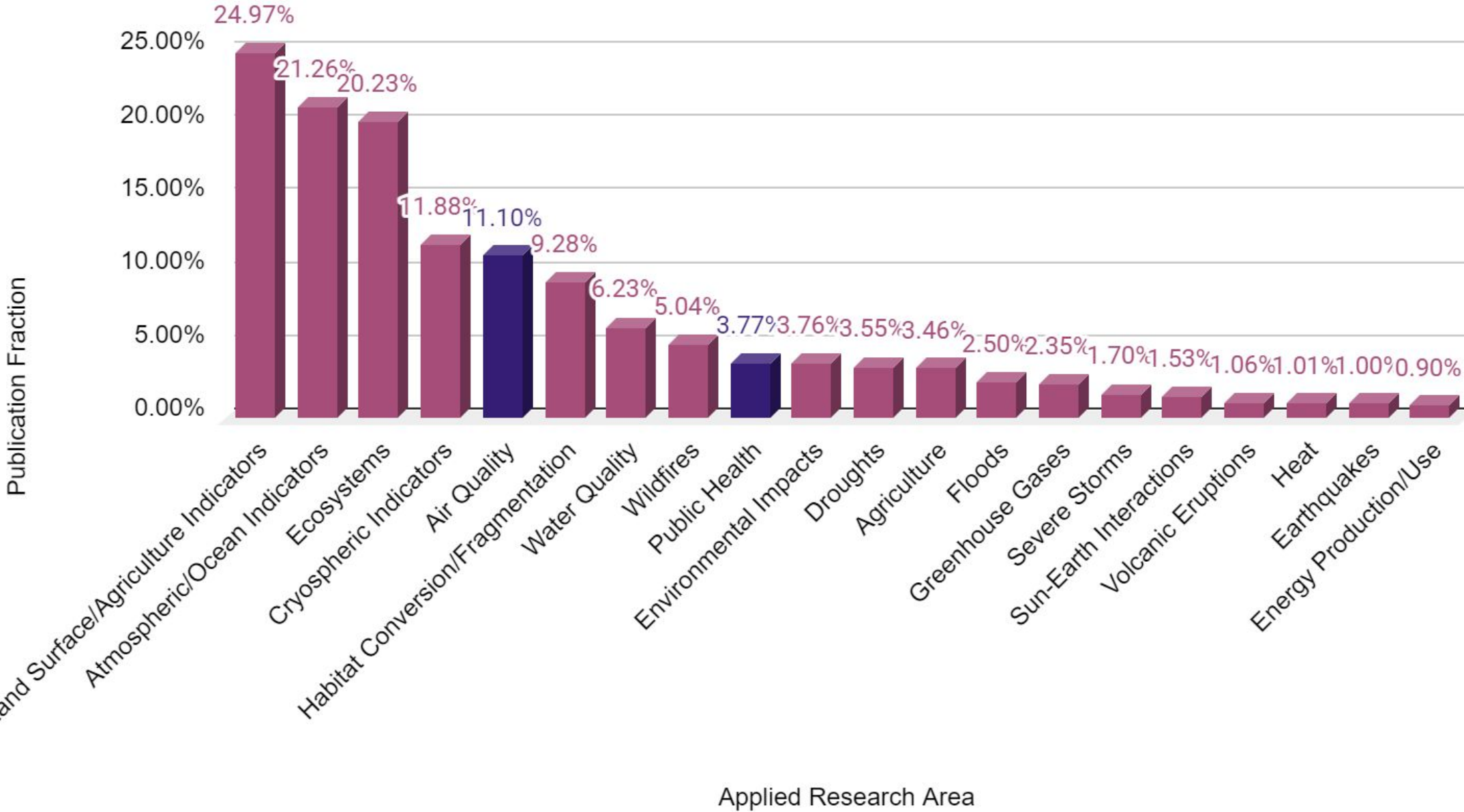
Applied Research Areas of EOSDIS data



Publication Citation Collection Flowchart

Air Quality and Public Health topics comprise 9% and 4% of 18,247 EOSDIS publications. For GES DISC, publications are gathered using Data DOI and short name.

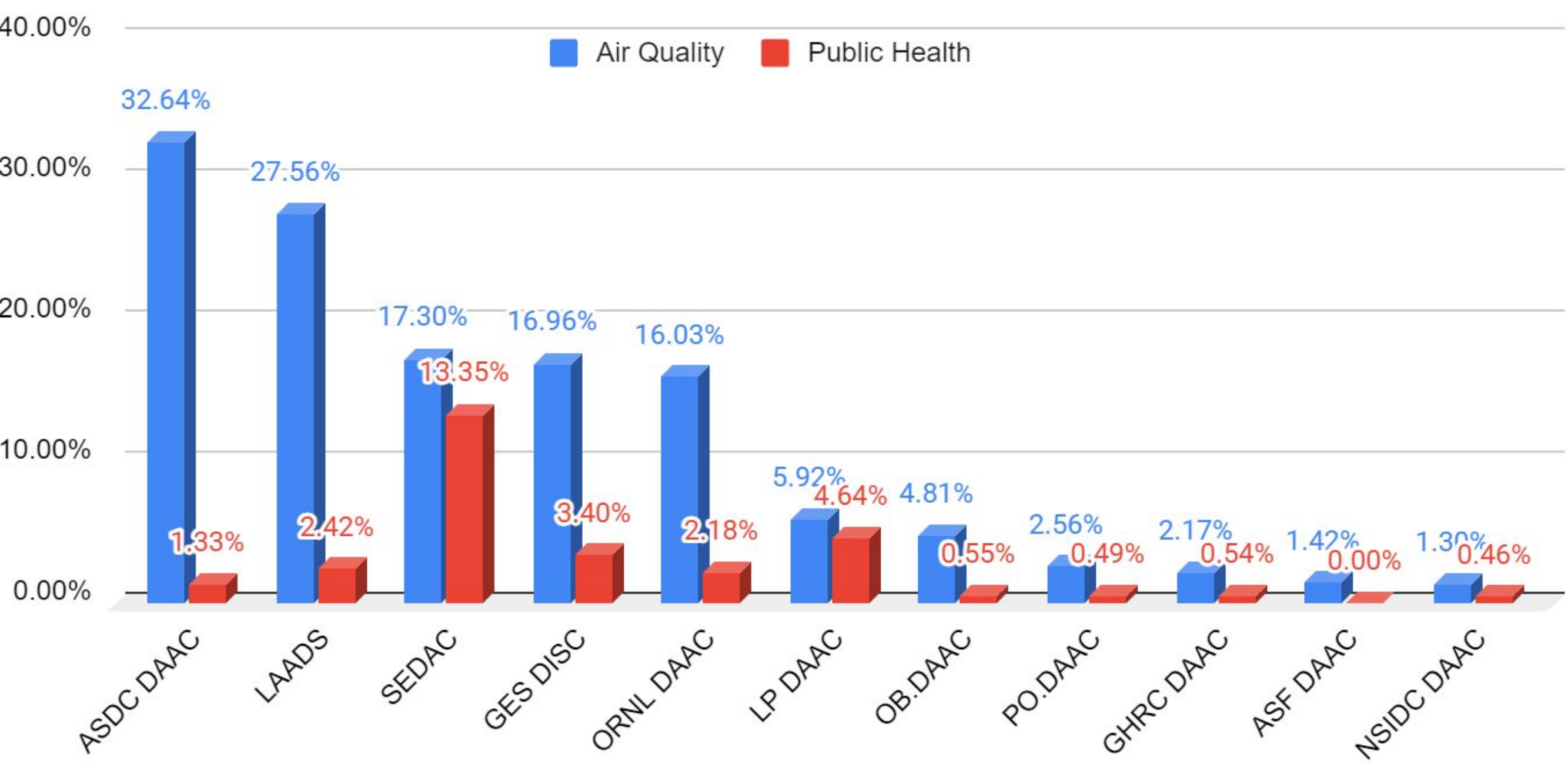
Applied Research Areas in 18,247 EOSDIS Dataset-Citing Publications



Publications by Applied Research Area in DAACs

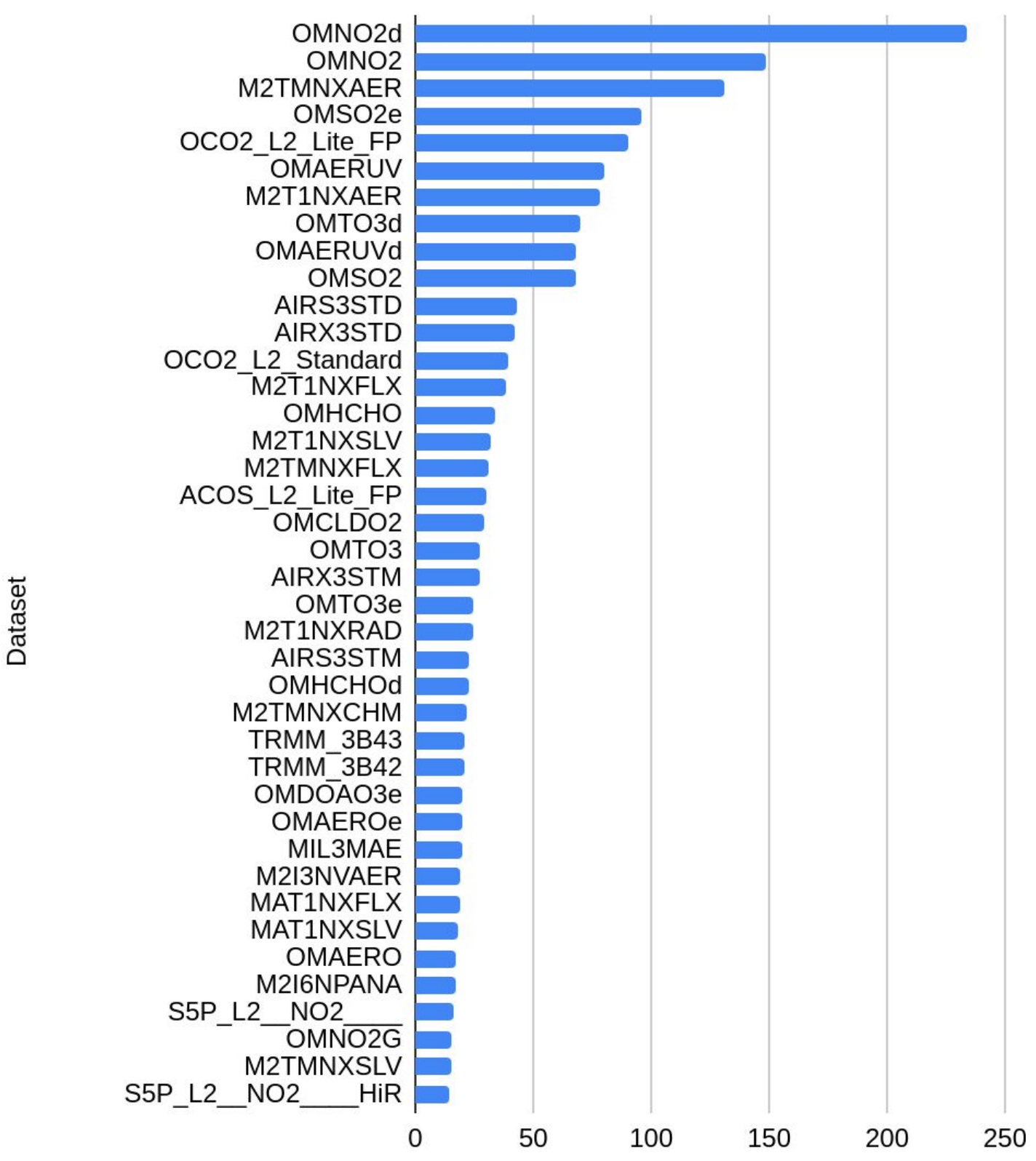
Percentage of datasets grouped by air quality and health distributed across the DAACs (figure below) shows that ASDC DAAC has ~33% of air quality data, with GES DISC having 17%. SEDAC has the majority of public health datasets.

EOSDIS Dataset DOI-Citing Publications Fractions for 'Air Quality' and 'Public Health' Applied Research Areas by DAAC

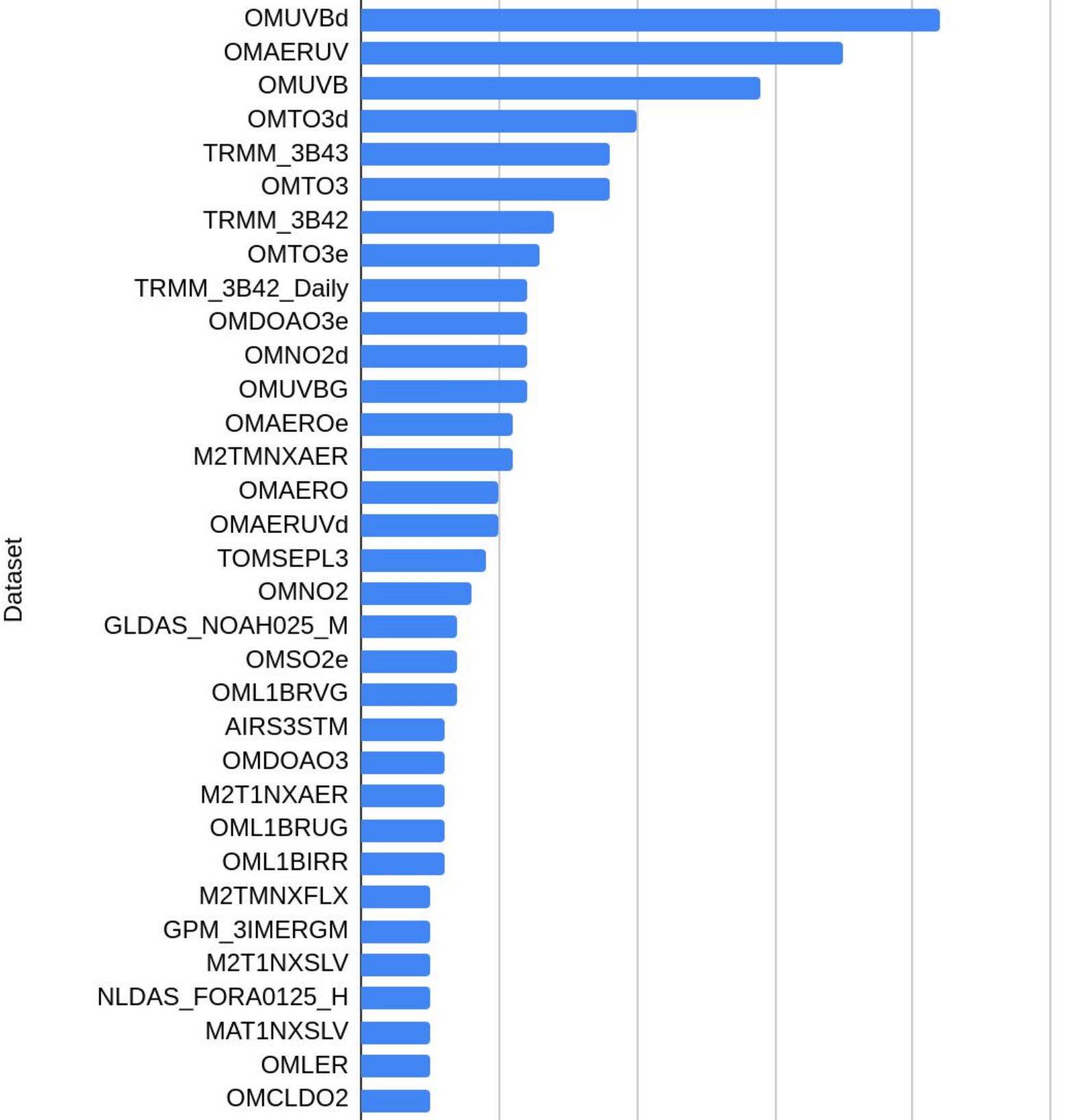


GES DISC total count of publications is 11,064. Of those, 1,394 are Air Quality and 272 are Public Health. Shown below are counts of the citing of GES DISC datasets in these publications.

Air Quality Publication Counts for GES DISC Datasets



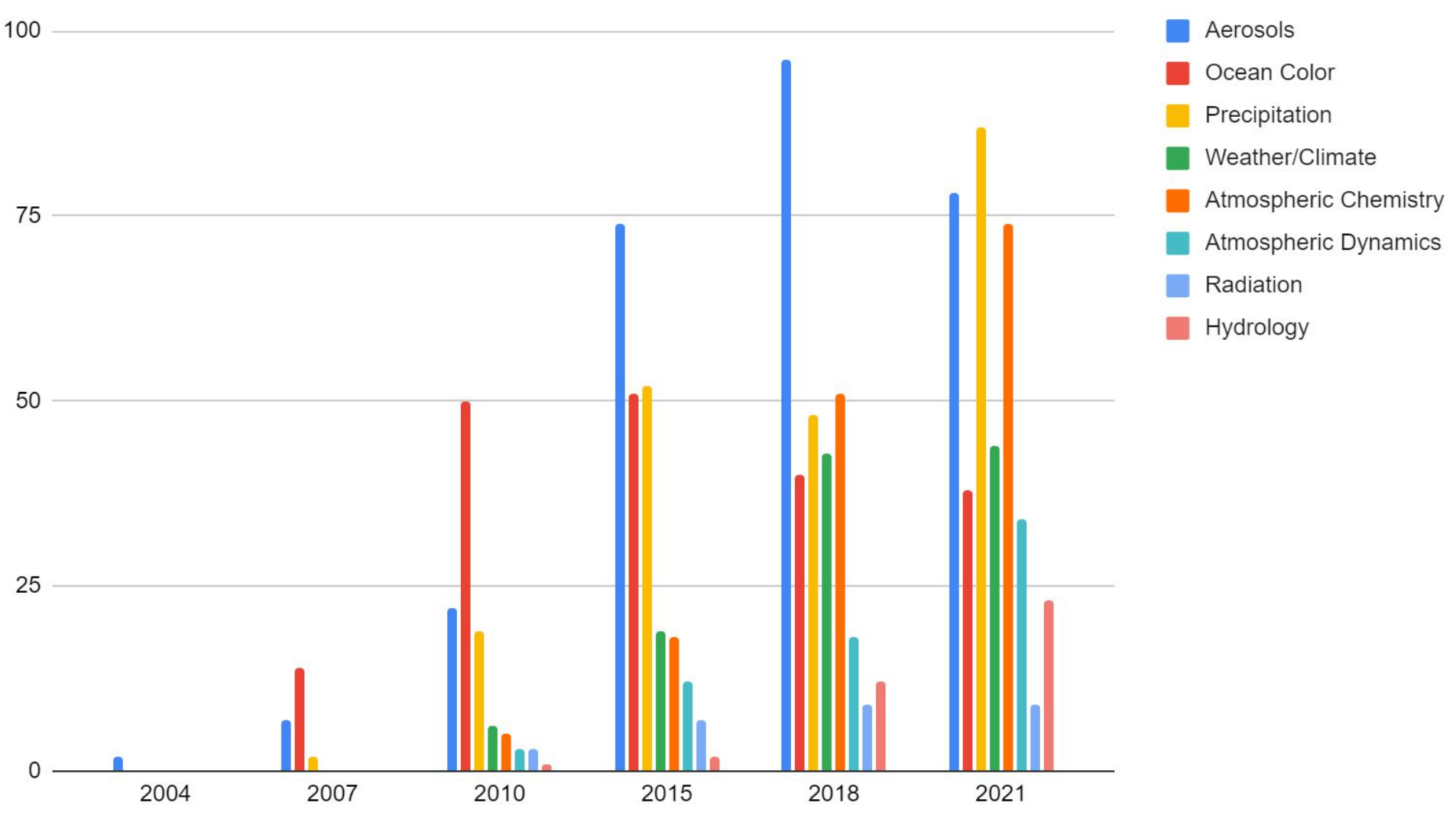
Public Health Publication Count for GES DISC Datasets



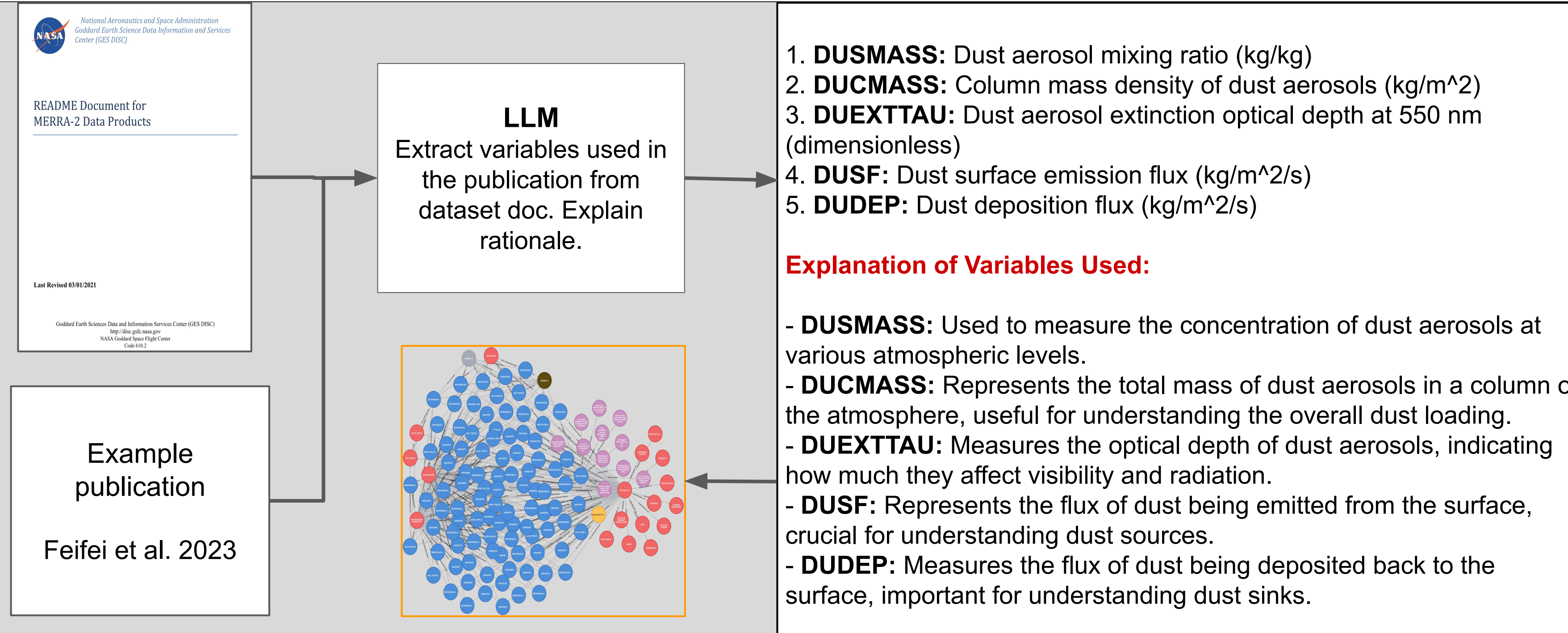
Air Quality Publications in Giovanni

The Giovanni tool has a great impact in the scientific community as it has enabled more than 3000 peer-reviewed publications to date. The breakdown of variables used in Giovanni by publication is shown in the figure at right. *Aerosol* is the leading discipline used in these papers.

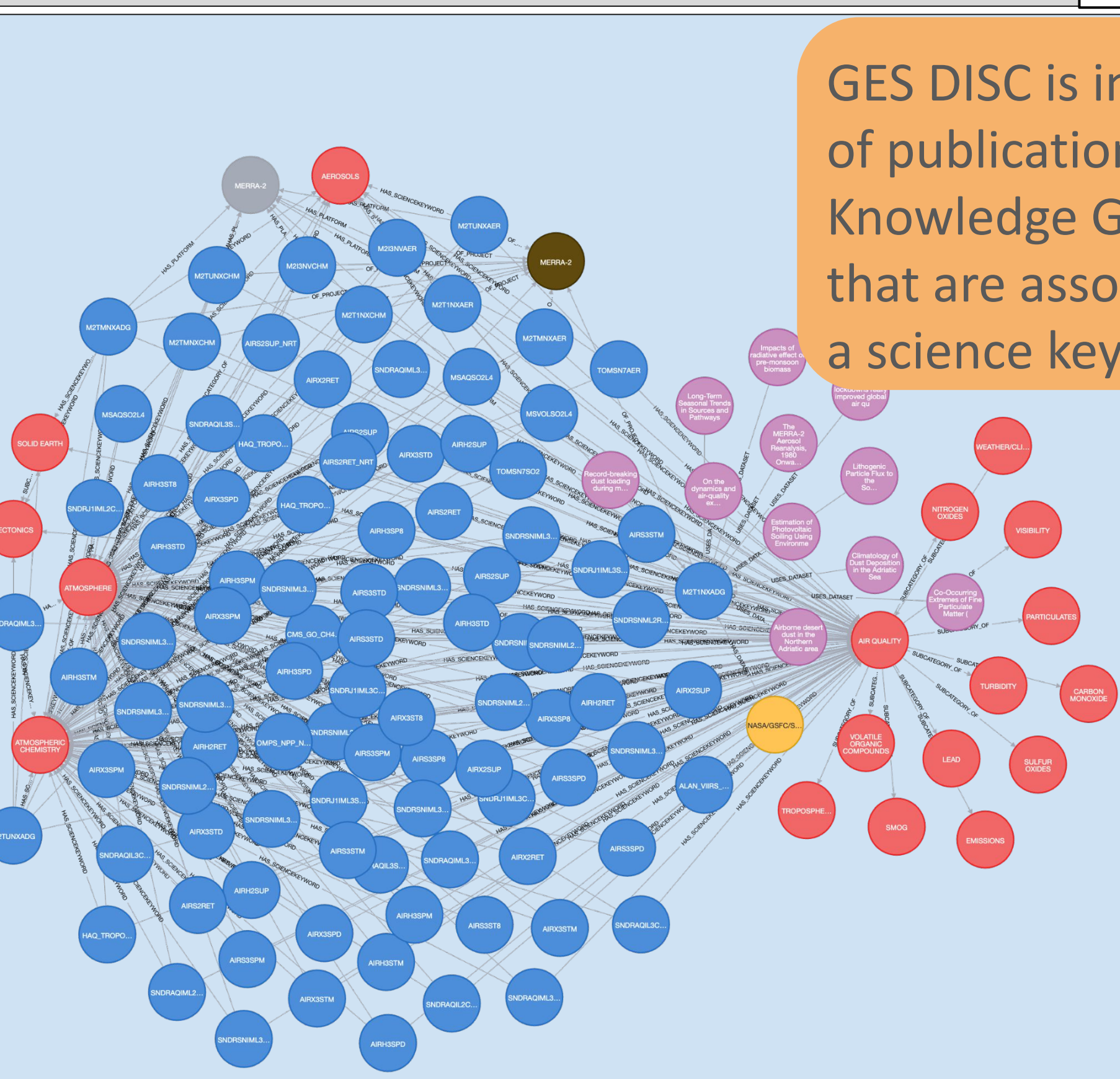
Giovanni Variable Disciplines Used in Publications, count



Coming Soon!

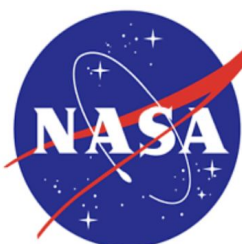


GES DISC is in the process of publishing Knowledge Graphs of publications using EOSDIS data in Kaggle Platform. The Knowledge Graph will help users to identify publications that are associated with a particular dataset that align with a science keyword.



Example: use of GES-DISC Knowledge Graph related to “**Air Quality**” keyword

- Blue nodes represent datasets
- Purple nodes represent publications that cite an example MERRA 2 dataset
- Red nodes represent Science Keywords



GES DISC Help Desk
gsfc-dl-help-disc@mail.nasa.gov

Follow us on X
[@NASAEarthData](https://twitter.com/NASAEarthData)

Earthdata Forum
<https://forum.earthdata.nasa.gov>

Earthdata YouTube
<https://www.youtube.com/c/NASAEarthData>