

Unveiling Knowledge: Enhancing Data Discovery through Integrative Knowledge Graphs at NASA's GES-DISC AGU Fall Meeting 2023



Armin Mehrabian^{1,2}, Irina Gerasimov^{1,2}, Mohammad Khayat^{1,2}, Brianna Pagán^{1,2}
Ben Forshey^{1,2}, Binita KC^{1,2}, Mahabal Hegde¹, and David J Meyer¹

¹Code 619, NASA Goddard Space Flight Center

²ADNET Systems Inc

Funny Meme about:
How AI is moving fast but people are
becoming AI experts even faster

Who We Are



- ▶ We are the NASA Goddard Earth Sciences Data and Information Services Center (GES-DISC) <https://disc.gsfc.nasa.gov/>
- ▶ Located at the Goddard Space Flight Center (GSFC) in Greenbelt, Maryland, USA
- ▶ One of 12 NASA Science Mission Directorate Data Centers that provide Earth science data, information, and services
- ▶ Our area of focus is *Atmospheric Composition, Water & Energy Cycles* and *Climate Variability*
- ▶ Our number one goal is to serve your Earth science data and information needs
- ▶ Currently we host more than **1500** datasets

Where the knowledge exist

1. Dataset metadata

```

}, {
  "Publisher": "Goddard Earth Sciences Data and Information Services Center (GES DISC)",
  "Title": "GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06",
  "SeriesName": "GPM_3IMERGHH",
  "ReleaseDate": "2019-03-20T00:00:00.000Z",
  "Version": "06",
  "ReleasePlace": "Greenbelt, MD"
}, {
  "SpatialExtent": {
    "GranuleSpatialRepresentation": "CARTESIAN",
    "HorizontalSpatialDomain": {
      "Geometry": {
        "CoordinateSystem": "CARTESIAN",
        "BoundingRectangles": [ {
          "WestBoundingCoordinate": -180,
          "NorthBoundingCoordinate": 90,
          "EastBoundingCoordinate": 180,
          "SouthBoundingCoordinate": -90
        } ]
      }
    }
  }, {
    "Type": "GET RELATED VISUALIZATION",
    "Subtype": "GIOVANNI",
    "URL": "https://giovanni.gsfc.nasa.gov/#dataKeyword",
    "Description": "Online visualization and analysis",
    "URLContentType": "VisualizationURL"
  }, {
    "Type": "USE SERVICE API",
    "Subtype": "OPENDAP DATA",
    "URL": "https://gpm1.gesdisc.eosdis.nasa.gov/opendap",
    "Description": "Access the data via the OPENDAP protocol",
    "URLContentType": "DistributionURL"
  }, {
    "Type": "USE SERVICE API",
    "Subtype": "GRADS DATA SERVER (GDS)",
    "URL": "https://gpm1.gesdisc.eosdis.nasa.gov/dods/G",
    "Description": "The GrADS Data Server (GDS) is another way to access the data",
    "URLContentType": "DistributionURL"
  }, {
    "Category": "EARTH SCIENCE",
    "Topic": "ATMOSPHERE",
    "Term": "PRECIPITATION",
    "VariableLevel1": "PRECIPITATION RATE"
  }, {
    "Category": "EARTH SCIENCE",
    "Topic": "ATMOSPHERE",
    "Term": "PRECIPITATION",
    "VariableLevel1": "SOLID PRECIPITATION",
    "VariableLevel2": "SNOW"
  }, {
    "Category": "EARTH SCIENCE",
    "Topic": "ATMOSPHERE",
    "Term": "PRECIPITATION",
    "VariableLevel1": "LIQUID PRECIPITATION",
    "VariableLevel2": "RAIN"
  }, {
    "TemporalExtents": [ {
      "RangeDateTimes": [ {
        "BeginningDateTime": "2000-06-01T00:00:00.000Z",
        "EndsAtPresentFlag": false
      } ],
      "ProcessingLevel": {
        "Id": "3"
      }
    } ],
    "DOI": {
      "DOI": "10.5067/GPM/IMERG/3B-HH/06"
    }
  }
}

```

GES DISC Find a DAAC

Data Collections precipitation

Atmospheric Composition, Water & Energy Cycles and Climate Variability

Data Collections Showing 1 - 25 of 78 datasets associated with precipitation

Refine By

Features Sort

- ☐ Cloud Enabled (60)

Subject Sort

- ☐ Altitude (2)
- ☐ Atmospheric Radiation (3)
- ☐ Atmospheric Temperature (2)
- ☐ Atmospheric Water Vapor (39)
- ☐ Atmospheric Winds (2)

Measurement Sort



- ☐ Air Temperature (2)
- ☐ Brightness Temperature (10)
- ☐ Cloud Fraction (2)
- ☐ Cloud Frequency (2)
- ☐ Cloud Liquid Water/Ice (4)

Source Sort

- ☐ Aqua AIRS (3)
- ☐ DMSP 5D-3/F16 SSMIS (1)
- ☐ DMSP 5D-3/F17 SSMIS (1)

Want to focus your search?


- Add more keywords to your search (e.g., surface precipitation).
- Use the filters in 'Refine By'.

Dataset	Source	Version	Time Res.	Spatial Res.	Process Level	Begin Date	End Date
 GPM IMERG Final Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06 (GPM_3IMERGHH 06)	Multi-sensor Analysis	06	30 minutes	0.1 ° x 0.1 °	3	2000-06-01	2021-09-30
 GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06 (GPM_3IMERGDF 06)	Multi-sensor Analysis	06	1 day	0.1 ° x 0.1 °	3	2000-06-01	2021-09-30

Where the knowledge exist



1. Dataset metadata
2. Data documentations



National Aeronautics and Space Administration

Algorithm Theoretical Basis Document (ATBD)
Version 06

**NASA Global Precipitation Measurement (GPM)
Integrated Multi-satellite Retrievals for GPM
(IMERG)**

Prepared for:


**Global Precipitation Measurement (GPM)
National Aeronautics and Space Administration (NASA)**

Prepared by:

George J. Huffman
NASA/GSFC
NASA/GSFC Code 612
Greenbelt, MD 20771

and David T. Bolvin, Dan Braithwaite, Kuolin Hsu, Robert Joyce,
Christopher Kidd, Eric J. Nelkin, Soroosh Sorooshian, Jackson Tan,
Pingping Xie

13 March 2019



**National Aeronautics and Space
Administration
Goddard Earth Science Data
Information and**

**README Document for the
GPM Data**

Integrated Multi-satellite Retrievals for GPM (IMERG) Technical Documentation

George J. Huffman (1), David T. Bolvin (1,2), Eric J. Nelkin (1,2), Jackson Tan (1,3)

(1) Mesoscale Atmospheric Processes Laboratory, NASA Goddard Space Flight Center
(2) Science Systems and Applications, Inc.
(3) University Space Research Association

6 October 2020

News Archive

22 September 2020 As of 07:43:54 UTC on 27 September 2020 the Product/Version metadata for IMERG Early and Late was corrected from 06A to 06B, which presumably was incorrect since the shift to 06B processing at 13:33 EDT 22 May 2019. [Note: the actual product hours of the shifts will be later by the latency of the product.]

22 September 2020 A bug fix was installed in GPROF that eliminates negative POP values in southerly retrievals.

9 September 2020 The first Fall 2020 GOES-W noise season was less intrusive than previously, running up to 3 half-hours during 25-30 August 2020.

27 August 2020 GPM Core Observatory went into safe-hold -1100 UTC 10 August 2020 over the Southern Indian Ocean due to a single electron-magnetic event that affected the spacecraft processor. Science data were lost until operations resumed on 26 August 2020, starting -1400 UTC for GMI and -2300 UTC for DPR. IMERG continued production throughout the outage, but without the GMI data or accumulating calibration information from CORRA.

10 April 2020 As about 13:15 UTC on 30 April 2020 the Ka was returned to service, including driving the "full" DPR-based Combined, which then replaced the Ka-only Combined in the CORRA-GMI calibration accumulation.

27 April 2020 At about 07:50 UTC on 22 April 2020 the DPR went off-line (over Antarctica). The DPR Ka was returned to service at about 18:17 UTC, but the DPR Ka continued to be off-line. A Ka-only CORRA was initiated around 13:34 UTC on 26 April 2020, which started feeding the CORRA-GMI calibration accumulation this around 11:20 UTC on 27 April 2020, meaning the IMERG calibration will be about 6 days short of a full set, but very likely still stable.

26 February 2020 As in August, the GOES-W cooling problems resulted in noise that contaminated the CPC 4-km merged global 0.2 product, and so the Early and Late estimates. The date span was 15-26 February 2020. Furthermore, it appears that this problem is predictable for 4 times a year in February, April, August, and October. The IMERG team is experimenting with automated QC.

7 February 2020 The METOP-A channel 2 started degrading in early December 2019 and has been judged unusable for IMERG. Following the discovery of this problem in early February 2020, use was halted in the Early and Late Runs after the 3 February 2020 orbit ending 13:33:06 UTC. Users should expect to see some noise from METOP-A starting in mid-December 2019 up to this date/time. The Final Run stops using METOP-A after 3 December 2019 based on a conservative analysis that includes the fact that METOP-B and -C are in nearly redundant orbits. 17 December 2019 Australian SAPHIR lightning data were discovered starting in orbit 9964 and ending in 10001, which covers most of the period 12 September 2019 04:00 UTC to 19

Product Summary

Variables

Data Citation

Documentation

References

ALGORITHM THEORETICAL BASIS DOCUMENT (ATBD): [IMERG_ATBD_V06.pdf](#)

PI DOCUMENTATION: [Release notes; New Morphing algorithm](#)

PI DOCUMENTATION: [IMERG Technical Documentation](#)

PI DOCUMENTATION: [IMERG Quality Index](#)

PI DOCUMENTATION: [Caveats for IMERG extension into TRMM era](#)

READ-ME: [README Document](#)

IMPORTANT NOTICE: [IMERG Release Notes](#)

ANOMALIES: [GPM and partner sensors anomalous events](#)

PROJECT HOME PAGE: [GPM Project Home Page](#)

PRODUCT USAGE: [In-region Direct S3 Zarr Cache Access](#)

GIS: [ArcGIS REST API portal](#)

GIS: [ArcGIS Enterprise Web Map](#)

Where the knowledge exist



1. Dataset metadata
2. Data documentations
3. How tos and FAQs

How to Read IMERG Data Using Python

1. Download the data

Before accessing data at GES DISC, a user must first register with Earthdata Login, then be authorized to access data at GES DISC

- In a web browser, go to: <https://disc.gsfc.nasa.gov>
- In the Search field, enter GPM_3IMERGM and press enter. Figure 1 shows a screen shot of what the search results look like.
- Click on the latest version of the GPM_3IMERGM data, currently version 5.
- Click on the "Online Archive" button on the right.
 - click on the "2014/" folder
 - click on the link "3B-MQ-MS-MRG-3IMERG-20140101-S000000-E235959.01V06B.HDF5" to download the data file.
- Note: this recipe works for any IMERG data, not just monthly estimates.



2. Run the following cells to learn how to read and plot IMERG data with Python

The first step is to import the required Python libraries. If any of the following import commands fail, check the local Python environment and install any missing packages.

```
In [ ]: import numpy as np
import matplotlib.pyplot as plt
import cartopy.crs as ccrs
from cartopy.mpl.gridliner import LONGITUDE_FORMATTER, LATITUDE_FORMATTER
import matplotlib.ticker as mticker
```

Open the IMERG data for reading:

```
In [ ]: # If it may be necessary to add a path to the filename if it is not in the working directory
fn = "3B-MQ-MS-MRG-3IMERG-20140101-S000000-E235959.01V06B.HDF5"
f = h5py.File(fn, "r")
```

View the available groups in the file and the variables in the 'Grid' group:

```
In [ ]: groups = [ x for x in f.keys() ]
print(groups)
gridMembers = [ x for x in f['Grid'] ]
print(gridMembers)
```

Read the precipitation, latitude, and longitude data:

How-To's

Showing 1 - 25 of 94 how-to's

Sort by: Last Updated

Refine By

Data Tools Sort ▾

- ☐ API (2)
- ☐ ArcGIS (7)
- ☐ CMR (1)
- ☐ Direct S3 Access (2)
- ☐ GES DISC L3/4 Subsetter (3)
- ☐ GES DISC Level 2 Subsetting
- ☐ Service (1)
- ☐ Giovanni (2)

[More...](#)

Data Types Sort ▾

- ☐ Grid (25)
- ☐ Point (1)
- ☐ Shape (1)
- ☐ Sounder (1)
- ☐ Swath (5)

Data Formats Sort ▾

- ☐ ASCII (1)
- ☐ CSV (1)
- ☐ GRIB (8)
- ☐ HDF (10)

How to configure a Web map in the ArcGIS Enterprise Web Map Viewer

This How-to document will provide steps to configure several aspects of a Web map in the ArcGIS Web Map Viewer.

Related Data Collections (1) ▾

How to Access GES DISC Data Using Python

There are multiple ways to work with GES DISC data resources using Python. For example, the data can be accessed using techniques that rely on ...

[Download the companion Jupyter Notebook](#)

How to Access the Hydrology Data Rods Time Series API Using Python

This notebook describes accessing the Hydrology Data Rods Time Series API using Python. It queries a grid nearest Newton, IL, and queries ...

[Download the companion Jupyter Notebook](#)

Related Data Collections (1) ▾

How to use the line plot function in Panoply with Giovanni data maps

The software package Panoply provides a line plot function that creates a plot of data values along a latitude or longitude line in a data ...

Related Data Collections (1) ▾

FAQs

Showing 1 - 25 of 232 faqs

Sort by: Alphabetic

Any information about TRMM lifetime?

TRMM mission comes to an end after 17 years. Please see ...

Any other FAQs for TRMM, GPM, and precipitation?

The NASA PMM (Precipitation Measurement Missions) website has them,

...

Any quick overview about GPM products?

Yes, check this document: ...

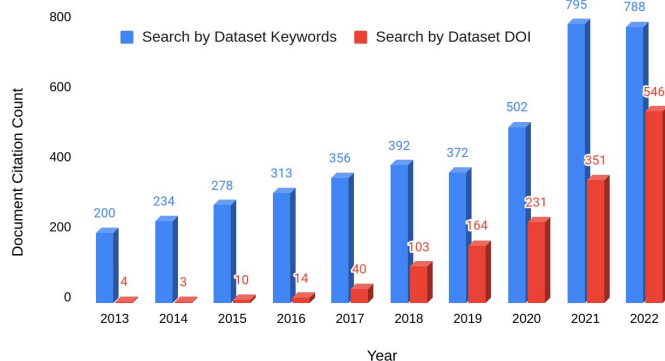
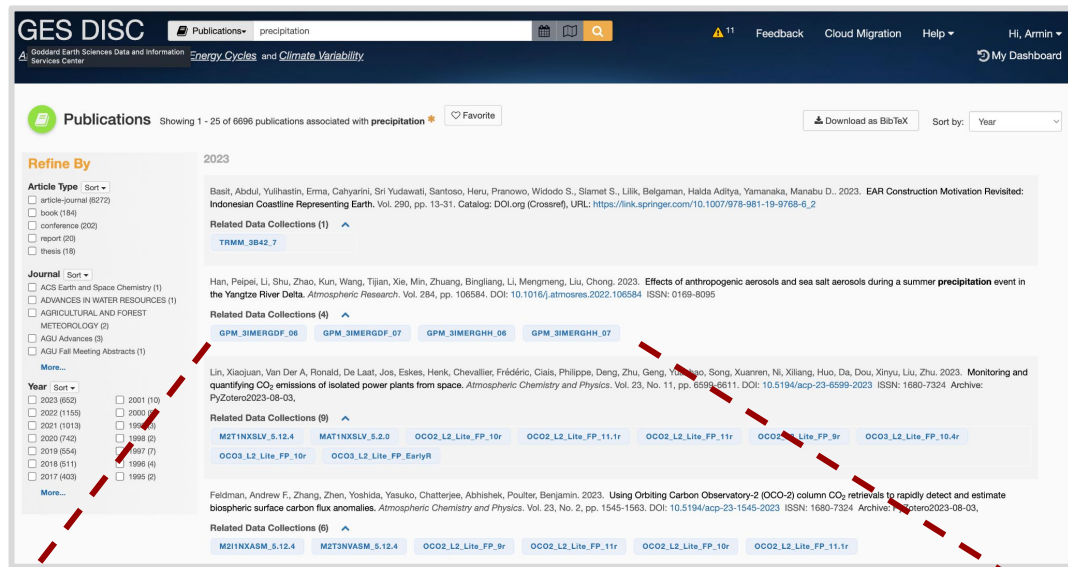
Any tools available to convert GPM HDF5 data to NetCDF?

The netCDF-4/HDF5 format was introduced in version 4.0; it is the HDF5 data format, with some restrictions (from Wiki). The HDF Group ...

Where the knowledge exist

1. Dataset metadata
2. Data documentations
3. How tos and FAQs
4. **Publications**

Document Citations found by Dataset DOI and Keyword Search

GES DISC Publications - precipitation

Showing 1 - 25 of 6696 publications associated with precipitation

Refine By

Article Type (Sort)

- ☐ article-journal (8772)
- ☐ book (134)
- ☐ conference (202)
- ☐ report (23)
- ☐ thesis (18)

Journal (Sort)

- ☐ ACS Earth and Space Chemistry (1)
- ☐ ADVANCES IN WATER RESOURCES (1)
- ☐ AGRICULTURAL AND FOREST METEOROLOGY (2)
- ☐ AGU Advances (3)
- ☐ AGU Fall Meeting Abstracts (1)

Year (Sort)

- ☐ 2023 (652)
- ☐ 2022 (1155)
- ☐ 2021 (1013)
- ☐ 2020 (142)
- ☐ 2019 (654)
- ☐ 2018 (611)
- ☐ 2017 (403)
- ☐ 2001 (10)
- ☐ 2000 (9)
- ☐ 1999 (9)
- ☐ 1998 (7)
- ☐ 1996 (4)
- ☐ 1995 (2)

Publications

Basil, Abdul, Yulhaslin, Erma, Cahyarin, Sri Yudawati, Santoso, Heru, Pranowo, Widodo S., Slamet S., Lilik, Belgaman, Haida Aditya, Yamanaka, Manabu D., 2023. EAR Construction Motivation Revisited: Indonesian Coastline Representing Earth. Vol. 290, pp. 13-31. Catalog: DOI.org (Crossref), URL: https://link.springer.com/10.1007/978-981-19-9768-6_2

Related Data Collections (1)

TRMM_3B42_7

Han, Peipei, Li, Shu, Zhao, Kun, Wang, Tijian, Xie, Min, Zhuang, Bingliang, Li, Mengmeng, Liu, Chong, 2023. Effects of anthropogenic aerosols and sea salt aerosols during a summer precipitation event in the Yangtze River Delta. *Atmospheric Research*. Vol. 284, pp. 106584. DOI: 10.1016/j.atmosres.2022.106584 ISSN: 0169-8095

Related Data Collections (4)

GPM_3IMERGDF_06 GPM_3IMERGDF_07 GPM_3IMERGHH_06 GPM_3IMERGHH_07

Lin, Xiaojuan, Van Der A, Ronald, De Laat, Jos, Eskens, Henk, Chevallier, Frédéric, Cials, Philippe, Deng, Zhu, Geng, Yuesha, Song, Xuanren, Ni, Xiliang, Huo, Da, Dou, Xinyu, Liu, Zhu, 2023. Monitoring and quantifying CO₂ emissions of isolated power plants from space. *Atmospheric Chemistry and Physics*. Vol. 23, No. 11, pp. 6599-6611. DOI: 10.5194/acp-23-6599-2023 ISSN: 1680-7324 Archive: <https://doi.org/10.5194/acp-23-6599-2023>

Related Data Collections (9)

M21NXSLV_5.12.4 MAT1NXSLV_5.2.9 OCO2_L2_Lite_FP_10r OCO2_L2_Lite_FP_11.1r OCO2_L2_Lite_FP_11r OCO2_L2_Lite_FP_9r OCO3_L2_Lite_FP_10.4r OCO3_L2_Lite_FP_10r OCO3_L2_Lite_FP_EarlyR

Feldman, Andrew F., Zhang, Zhen, Yoshida, Yasuko, Chatterjee, Abhishek, Poulter, Benjamin, 2023. Using Orbiting Carbon Observatory-2 (OCO-2) column CO₂ retrievals to rapidly detect and estimate biospheric surface carbon flux anomalies. *Atmospheric Chemistry and Physics*. Vol. 23, No. 2, pp. 1545-1563. DOI: 10.5194/acp-23-1545-2023 ISSN: 1680-7324 Archive: <https://doi.org/10.5194/acp-23-1545-2023>

Related Data Collections (6)

M21NXASM_5.12.4 M23NXASM_5.12.4 OCO2_L2_Lite_FP_9r OCO2_L2_Lite_FP_11r OCO2_L2_Lite_FP_10r OCO2_L2_Lite_FP_11.1r

Han, Peipei, Li, Shu, Zhao, Kun, Wang, Tijian, Xie, Min, Zhuang, Bingliang, Li, Mengmeng, Liu, Chong, 2023. Effects of anthropogenic the Yangtze River Delta. *Atmospheric Research*. Vol. 284, pp. 106584. DOI: 10.1016/j.atmosres.2022.106584 ISSN: 0169-8095

Related Data Collections (4)

GPM_3IMERGDF_06

GPM_3IMERGDF_07

GPM_3IMERGHH_06

GPM_3IMERGHH_07

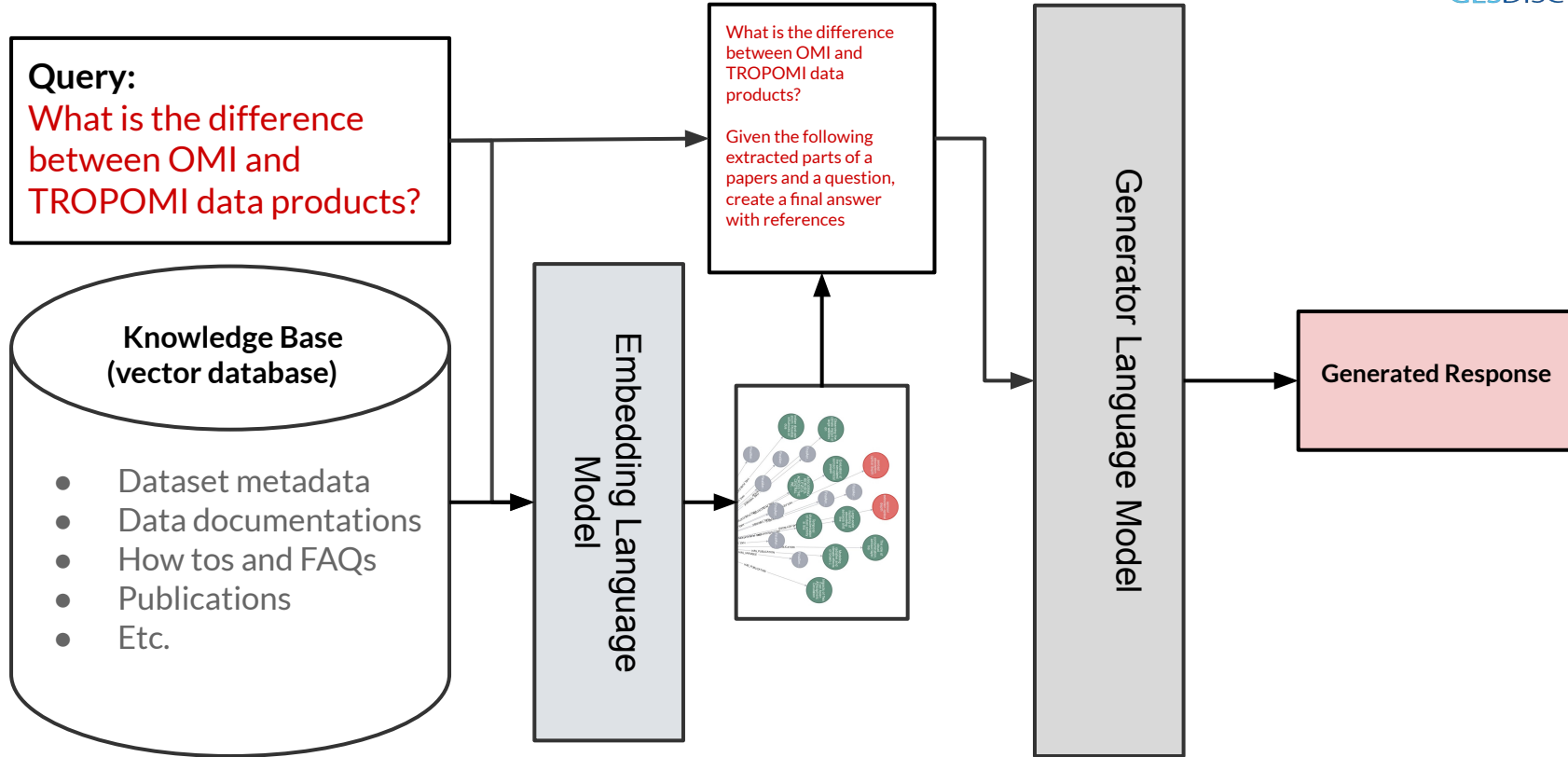
Another funny Meme about:
LLM this, LLM that, LLM ...

Traditional Search and Discovery

- ▷ Traditional search and data discovery primarily relies on curated metadata
- ▷ All of these sources are siloed
- ▷ We need to combine all of these into a knowledge base



Retrieval Augmented Generation (RAG)



Retrieval Augmented Generation (RAG)



A way of leveraging LLMs and your own data

73 TropOMAER_README_C0

74 TropOMAER_README_C0

66 OMBRO_README.pdf

73 OMBRO_README.pdf

RAGtriever Chat

Online v0.3.0

Powered by Weaviate

Copy

Cached (0.007461667)

The OMI (Ozone Monitoring Instrument) and TROPOMI (Tropospheric Monitoring Instrument) data products differ in several aspects, including their spatial resolution, spectral coverage, and the instruments' capabilities. Based on the provided context, here are some key differences:

What is a vector database?

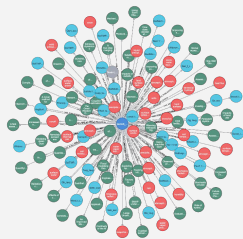
The OMI (Ozone Monitoring Instrument) and TROPOMI (Tropospheric Monitoring Instrument) data products differ in several aspects, including their spatial resolution, spectral coverage, and the instruments' capabilities. Based on the provided context, here are some key differences:

- 1. Spatial Resolution:** TROPOMI provides data with a typical pixel size of 5.5x3.5 km² for most spectral bands, which is a significant improvement over OMI's spatial resolution. This higher resolution allows for more detailed observations and better detection of local sources of pollutants.
- 2. Spectral Coverage:** TROPOMI covers eight spectral windows from the ultraviolet (UV) to the shortwave infrared (SWIR) regions, which allows for the detection and measurement of a wide range of atmospheric constituents. The provided context does not specify OMI's spectral coverage, but it is generally known that TROPOMI has an enhanced spectral resolution compared to OMI.
- 3. Instrument Configuration:** TROPOMI operates in a push-broom configuration with a swath width of about 2600 km, which enables it to capture a wide area in a single pass. This is an advancement over the OMI instrument, which has a different configuration and swath width.
- 4. Data Availability and Processing:** TROPOMI data products are generated by the Copernicus Sentinel project and distributed by ESA through NASA GES DISC. The context does not provide specific details on OMI data distribution, but it is known that OMI data are also available through NASA's data centers.
- 5. Algorithm and Product Development:** The TROPOMI aerosol inversion procedure (TROPOMAER) is a direct application of the OMI near UV aerosol algorithm (OMAERUV), indicating that the methodologies for aerosol data retrieval have been adapted and improved from OMI to TROPOMI.
- 6. Data Format:** TROPOMI data files are provided in the NetCDF4.0 format, which is a self-describing format with metadata included. The context does not specify OMI's data format, but traditionally, OMI data have been provided in HDF-EOS format.
- 7. Cloud Masking:** TROPOMI initially used an ESA-produced cloud mask based on NOAA VIIRS measurements, which was later replaced with the NOAA Enterprise Cloud Mask (ECM) product. OMI's cloud information handling is not detailed in the provided context, but it is known that cloud treatment is a critical aspect of satellite data retrieval.
- 8. Validation and Comparison:** TROPOMI data products are compared with AERONET observations for validation purposes, and the context mentions reasonable agreement between TROPOMI SSA retrievals and AERONET results. OMI data products also undergo validation, but the specific methods and results are not detailed in the provided context.

In summary, TROPOMI represents an advancement over OMI in terms of spatial resolution, spectral coverage, and data processing capabilities, which leads to improved atmospheric measurements and data products.

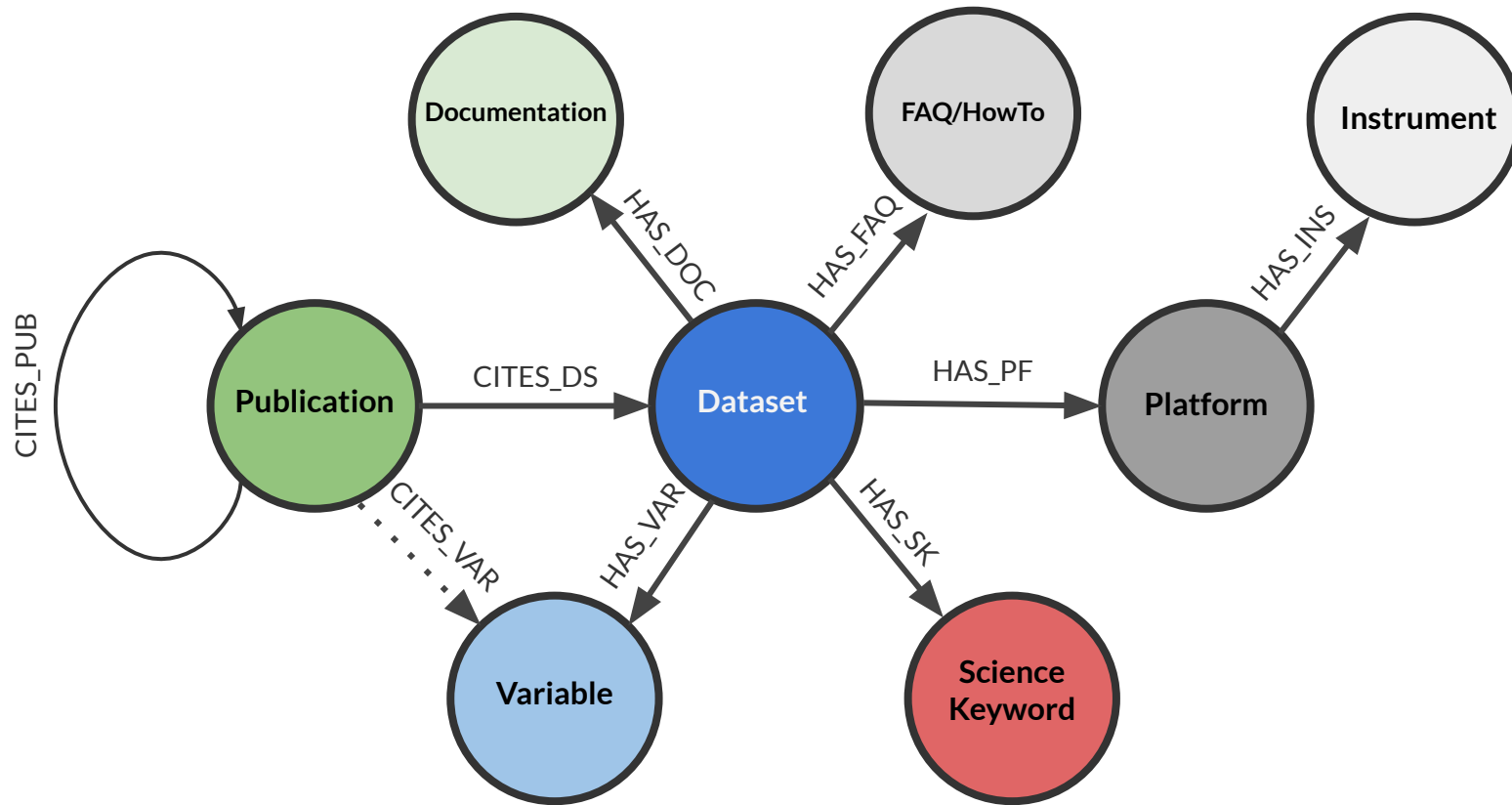
Knowledge Graph + (L)LM

Knowledge Base

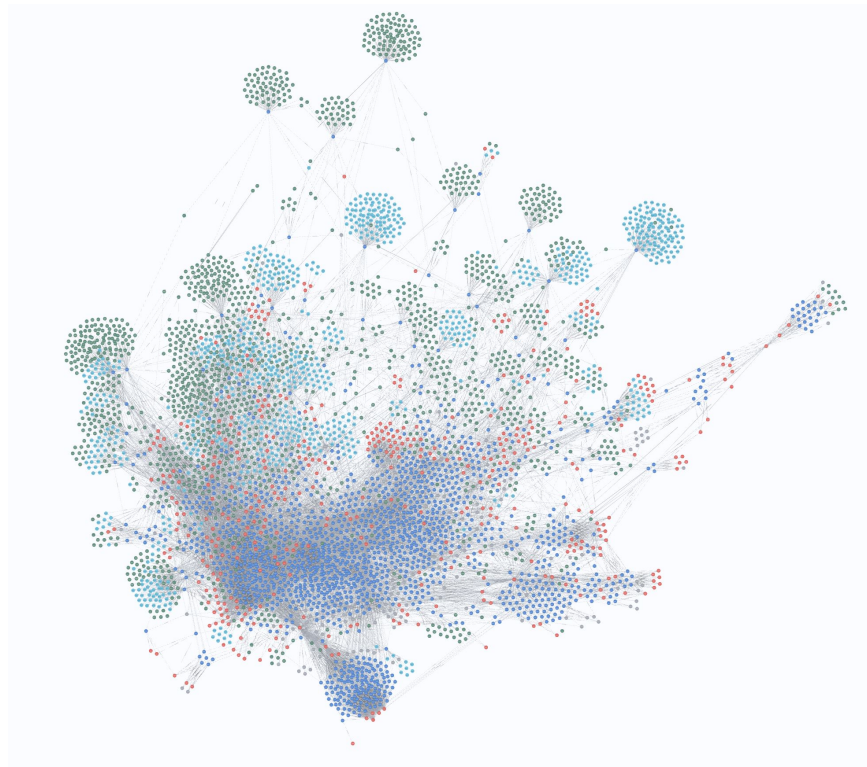


Large Language Model
(LLM)

High-level Schema



GES-DISC Knowledge Graph



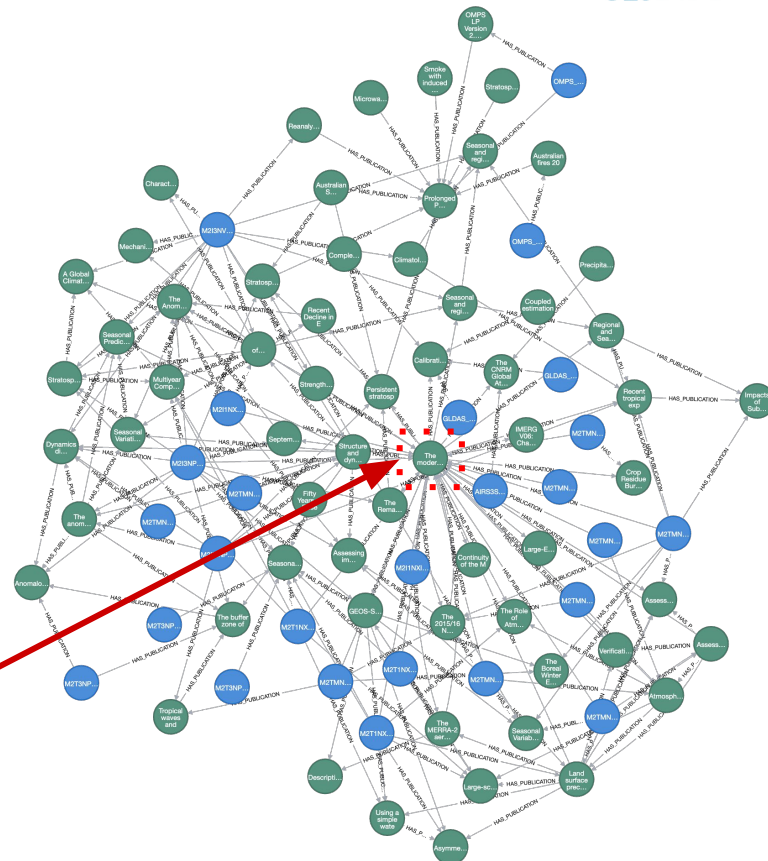
- Dataset: **1571**
- Publication: **2045**
- Science Keyword: **431**
- Variables: **951**
- Edges: **208K**

Graph Analytics and Graph ML



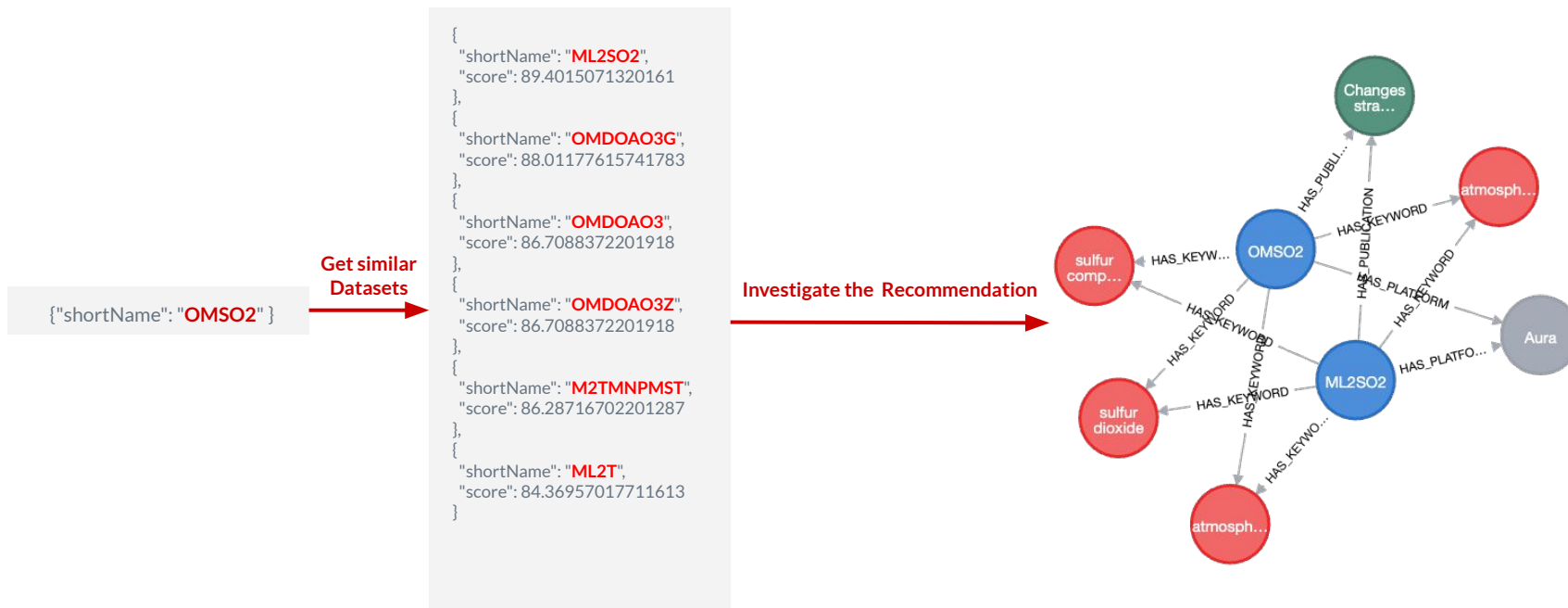
Centrality measures (I.E. pagerank)

- Rank influence of nodes in your network
- Identify influential publications within out citation network
- This publication has high PR score not just many publications cite it, but the citing publications have also been cited frequently



Graph Analytics and Graph ML

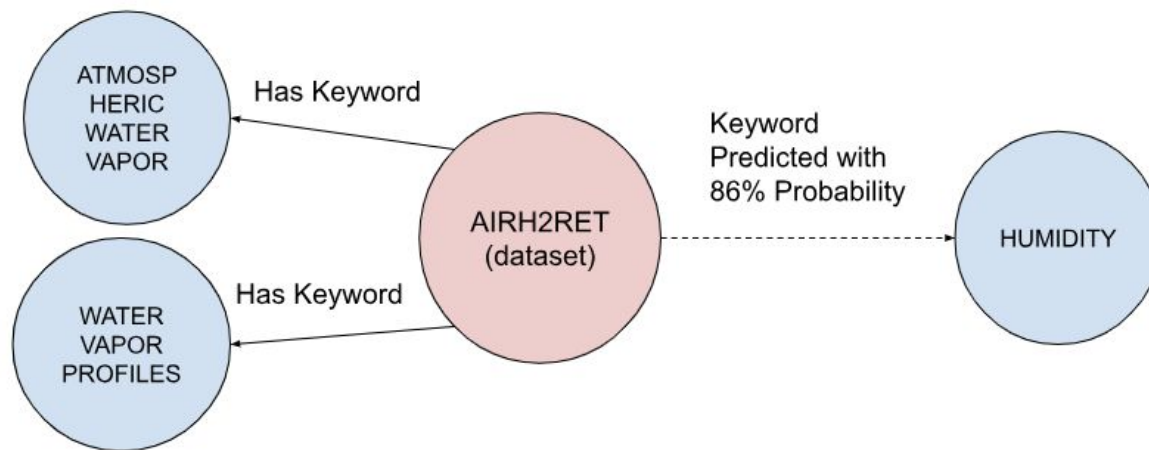
- ▷ Graph node embeddings (I.E. FastRP)
 - Used for dataset recommendation



Graph Analytics and Graph ML

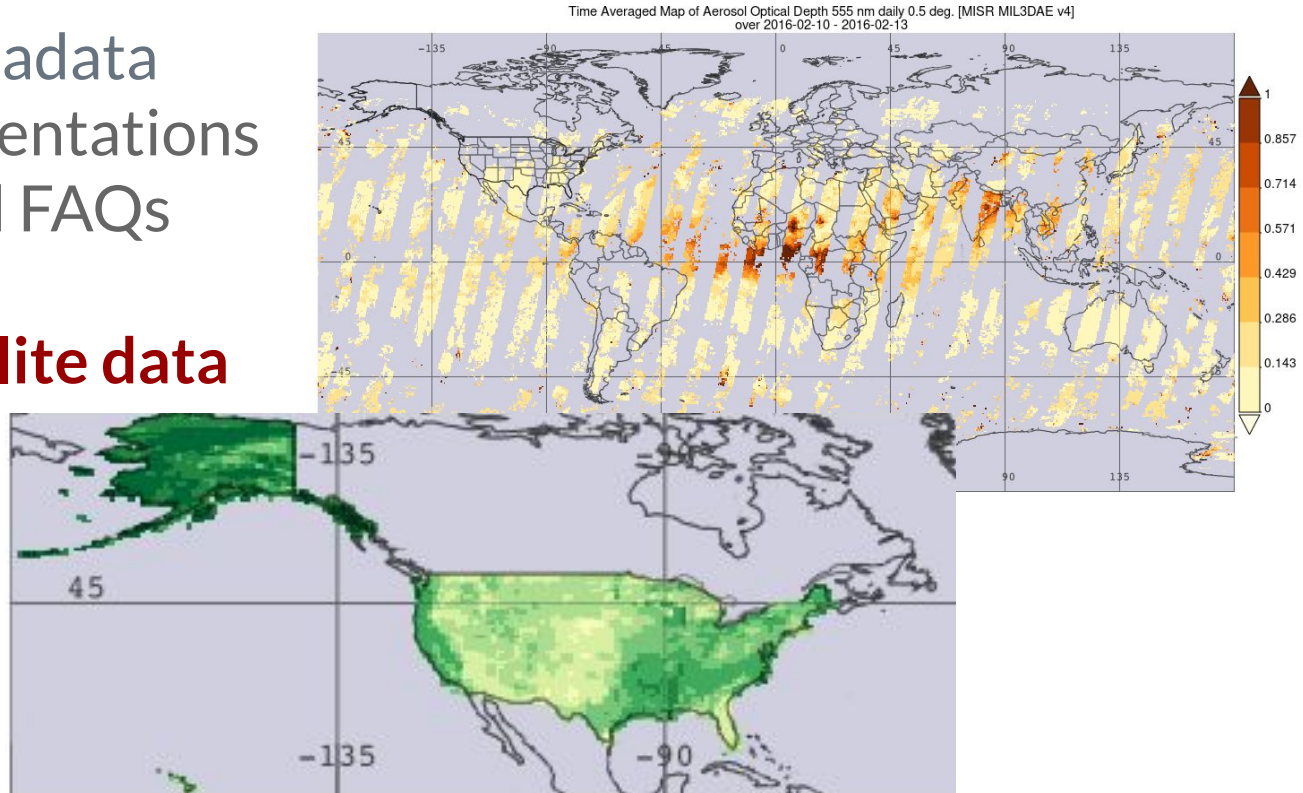
Missing link prediction

- Completing missing links between graph nodes (I.E. Datasets → Science Keywords)



Where the knowledge exist (Prediction)

1. Dataset metadata
2. Data documentations
3. How tos and FAQs
4. Publications
5. **Actual satellite data
(pixels)**



Multi-modal AI Models

- ▷ Humans experience the world multi-modal
- ▷ Data and applications we have are multi-modal
- ▷ Multi-modal data is rich and high-bandwidth - compared to language which is a low-bandwidth serialization
- ▷ We are running out of text data
- ▷ Vision-Text models
 - Image Text (Retrieval)
 - Image → Text (Captioning)
 - Text → Image (Generation)
- ▷ I.E. CLIP, Align (Late fusion of vision and text)
 - Train vision model and language model separately
 - Fuse them late using contrastive learning
 - Perhaps **NASA Geospatial AI Foundation + NASA LLM???**

THANK YOU

Please feel free to contact me if you
have any questions.

armin.mehrabian@nasa.gov

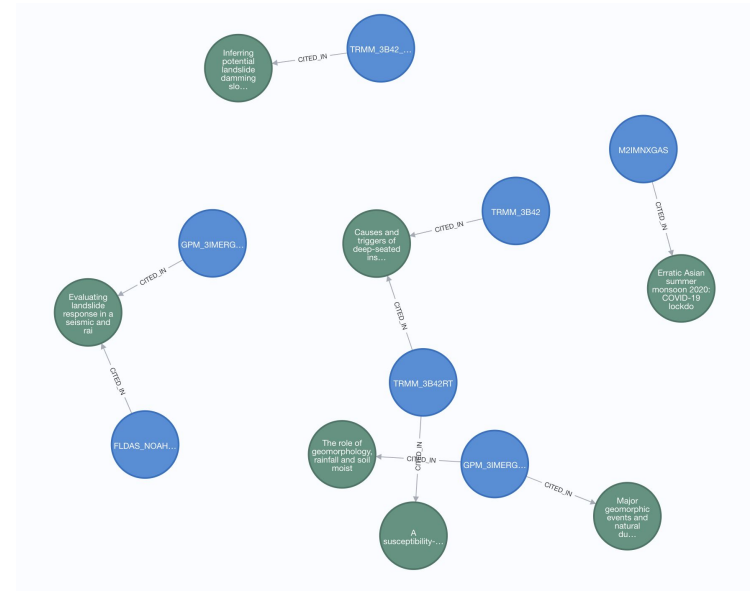


Backup Slides

Search Query “LANDSLIDE”

MATCH (d:Dataset)-[i:CITED_IN]->(p:Publication) **WHERE** p.abstract **CONTAINS** 'landslide' **RETURN** d.shortName, p.title

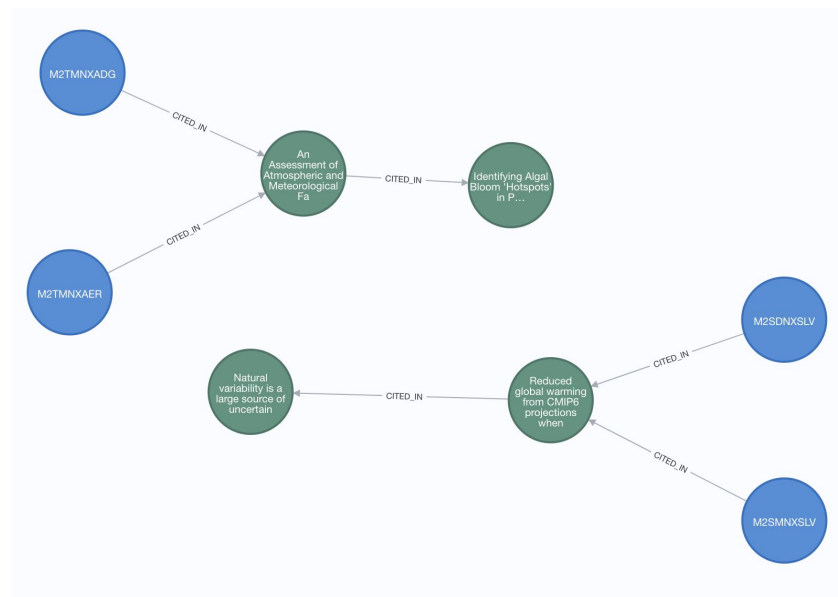
	d.shortName	p.title
1	"GPM_3IMERGHH"	"The role of geomorphology, rainfall and soil moisture in the occurrence of landslides triggered by 2018 Typhoon Mangkhut in the Philippines"
2	"GPM_3IMERGHH"	"Major geomorphic events and natural hazards during monsoonal precipitation 2018 in the Kali Gandaki Valley, Nepal Himalaya"
3	"M2IMNXGAS"	"Erratic Asian summer monsoon 2020: COVID-19 lockdown initiatives possible cause for these episodes?"
4	"TRMM_3B42"	"Causes and triggers of deep-seated hillslope instability in the tropics Insights from a 60-year record of Ikoma landslide (DR Congo)"
5	"TRMM_3B42_Daily"	"Inferring potential landslide damming using slope stability, geomorphic constraints, and run-out analysis: a case study from the NW Himalaya"
6	"FLDAS_NOAH01_C_GL_M"	"Evaluating landslide response in a seismic and rainfall regime: a case study from the SE Carpathians, Romania"
7	"GPM_3IMERGDF"	"Evaluating landslide response in a seismic and rainfall regime: a case study from the SE Carpathians, Romania"
8	"TRMM_3B42RT"	"Causes and triggers of deep-seated hillslope instability in the tropics Insights from a 60-year record of Ikoma landslide (DR Congo)"
9	"TRMM_3B42RT"	"A susceptibility-based rainfall threshold approach for landslide occurrence"



Search Query "ALGAL BLOOM"

MATCH (d:Dataset)-[i:CITED_IN*1..2]->(p:Publication) **WHERE** p.abstract **CONTAINS** 'algal bloom' **RETURN** d.shortName, p.title

	d.shortName	p.title
1	"M2SDNXSLV"	["Natural variability is a large source of uncertainty in future projections of hypoxia in the Baltic Sea"]
2	"M2TMNXAER"	["Identifying Algal Bloom 'Hotspots' in Marginal Productive Seas: A Review and Geospatial Analysis"]
3	"M2SMNXSLV"	["Natural variability is a large source of uncertainty in future projections of hypoxia in the Baltic Sea"]
4	"M2TMNXADG"	["Identifying Algal Bloom 'Hotspots' in Marginal Productive Seas: A Review and Geospatial Analysis"]



Search Query “PRECIPITATION”

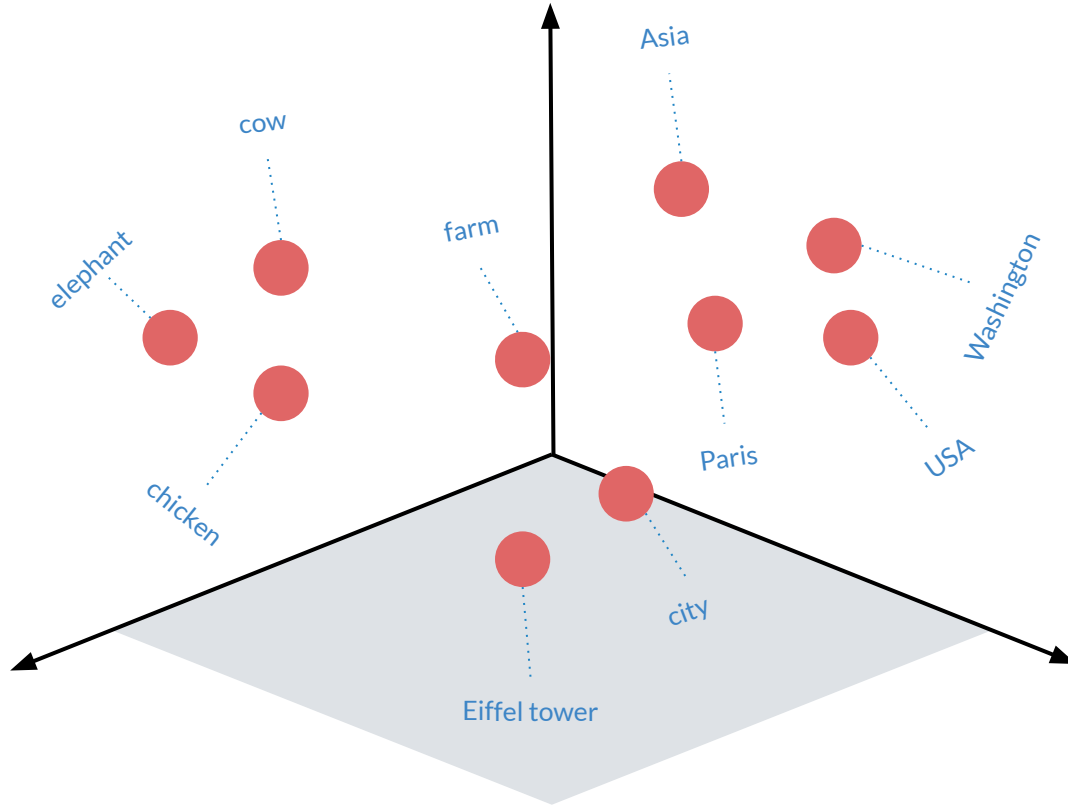


- ▶ We calculate the PageRank score for publications in our graph

```
MATCH (d:Dataset)-[i:CITED_IN]->(p:Publication) WHERE p.abstract CONTAINS 'precipitation' RETURN d.shortName, p.pagerank, p.title ORDER by p.pagerank DESC
```

	d.shortName	p.pagerank	p.title
1	"GPM_3IMERGDF"	0.48290645178018216	["Precipitation-Moisture Coupling Over Tropical Oceans: Sequential Roles of Shallow,
2	"MODIS_CR_Equal_Angle_3h"	0.47562706787251907	["Evaluation of GPROF V05 Precipitation Retrievals under Different Cloud Regimes"]
3	"MODIS_CR_Equal_Area_3h"	0.47562706787251907	["Evaluation of GPROF V05 Precipitation Retrievals under Different Cloud Regimes"]
4	"MODIS_CR_Equal_Angle_Daily"	0.26466569052647637	["Classifying Planetary Cloudiness with an Updated Set of MODIS Cloud Regimes"]
5	"MODIS_CR_Equal_Area_3h"	0.26466569052647637	["Classifying Planetary Cloudiness with an Updated Set of MODIS Cloud Regimes"]

Language embeddings



We can represent every word, sentence, phrase, ... document with a meaningful vector

elephant = [0.31, 0.62, ..., 0.87]

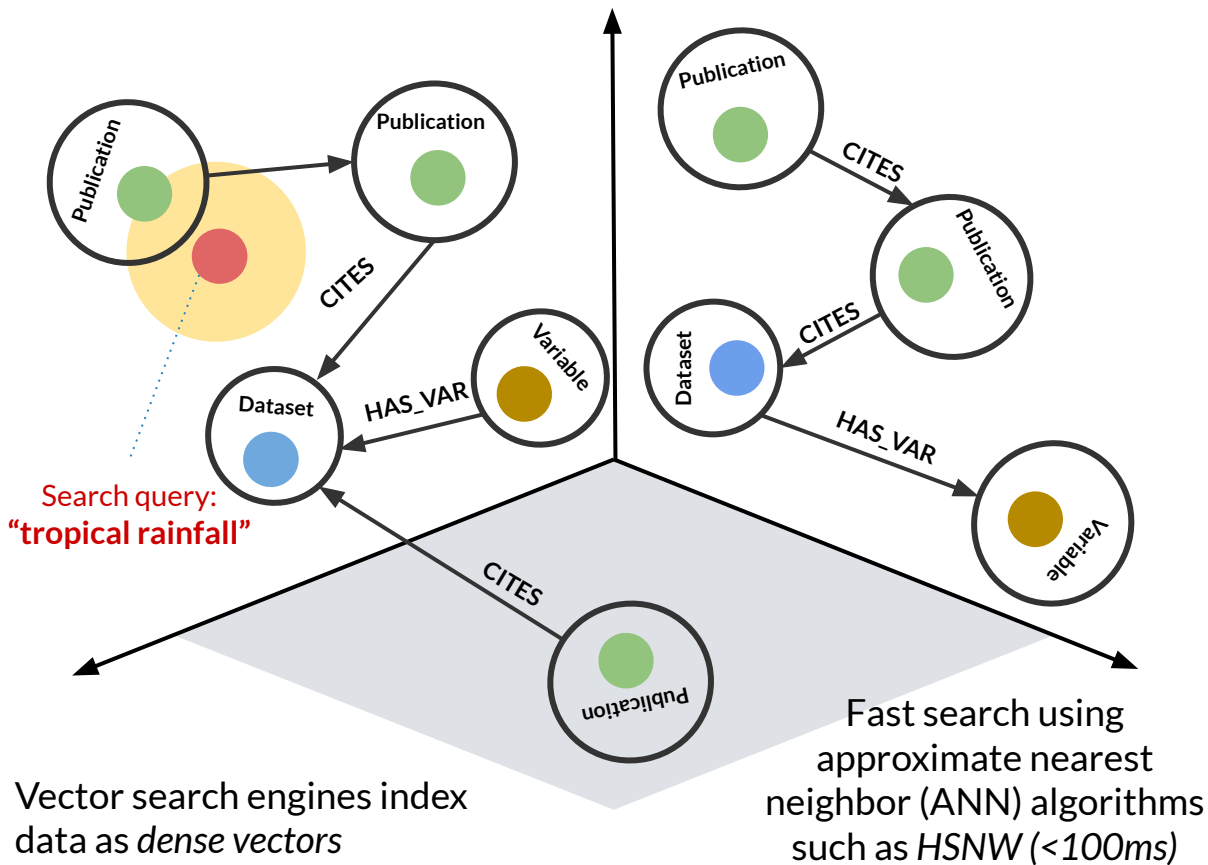
cheetah = [0.94, 0.54, ..., 0.88]

Eiffel tower = [0.45, 0.67, ..., 0.87]

France = [0.56, 0.83, ..., 0.22]

farm = [0.32, 0.68, ..., 0.76]

Graph-enabled Vector Search



Graph-enables vector search tools such as [Weaviate](#) create a graph layer on top of the vector search.

For a given user query i.e.

"tropical rainfall"

1. The query is placed in the vector space
2. All objects including publications and datasets within a radius of the query are identified.
3. If identified object is of type "publication", we can traverse graph to find the closest dataset (minimum hops)

What is the difference between OMI and TROPOMI data products?



The Ozone Monitoring Instrument (OMI) and TROPOspheric Monitoring Instrument (TROPOMI) are both used to measure total column ozone (TCO3). OMI data products are used to estimate surface-level concentrations of NO₂, SO₂, and other trace gases, while TROPOMI data products are used to monitor atmospheric composition data and are included in the data assimilation system of the Copernicus Atmosphere Monitoring Service (CAMS). OMSO2 V2, a new anthropogenic SO₂ vertical column density dataset, is generated with an algorithm based on principal component analysis of OMI radiances and features several updates, such as the use of expanded lookup tables and model a priori profiles to estimate SO₂ Jacobians for individual OMI pixels. The OMI OMCLDO2 cloud product is used to derive the effective cloud fraction and effective cloud pressure using a DOAS fit of the O₂-O₂ absorption feature around 477 nm. A new version of the OMI OMCLDO2 cloud product has been released that contains several improvements, such as the introduction of a temperature correction on the O₂-O₂ slant columns and the updated look-up tables.

SOURCES:

- [Monitoring and assimilation tests with TROPOMI data in the CAMS system: near-real-time total column ozone](#)
- [Ozone Monitoring Instrument \(OMI\) collection 4: establishing a 17-year-long series of detrended level-1b data](#)
- [A high-resolution and observationally constrained OMI NO₂ satellite retrieval](#)
- [Version 2 Ozone Monitoring Instrument SO₂ product \(OMSO2 V2\): new anthropogenic SO₂ vertical column density dataset](#)
- [Improvements to the OMI O₂-O₂ operational cloud algorithm and comparisons with ground-based radar-lidar observations](#)

S5P_L2_O3_TOT 1

S5P_L2_O3_TCL 1

M2TMNXSLV 5.12.4

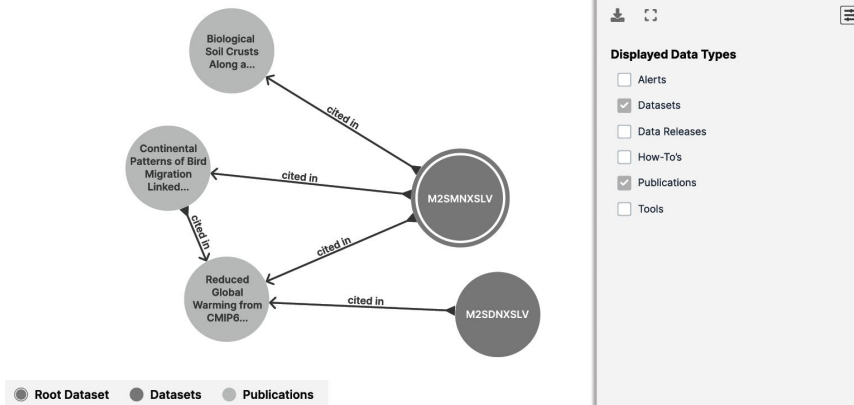
AIRS3STM 006

surface_air_pressure

ozone averaging kernel

surface albedo

Data Related to M2SMNXSLV.5.12.4



click and drag to pan scroll to zoom

Close

Query: what is the correlation between human population and wildfire?



Returns the starting and ending position of text from the documents that contain the answers

```
1 {  
2   Get {  
3     Publication(  
4       ask: {  
5         question: "what is the correlation between human population and wildfire?",  
6         properties: ["abstract"],  
7         rerank: true # supported from v1.10.0 on  
8       },  
9       limit: 5  
10    ) {  
11      title  
12      _additional {  
13        answer {  
14          hasAnswer  
15          certainty  
16          property  
17          result  
18          startPosition  
19          endPosition  
20        }  
21      }  
22    }  
23  }  
24 }
```

```
"data": {  
  "Get": {  
    "Publication": [  
      {  
        "_additional": {  
          "answer": {  
            "certainty": 0.4658797264099121,  
            "endPosition": 1973,  
            "hasAnswer": true,  
            "property": "abstract",  
            "result": "positively with burnt area only in densely  
forested regions",  
            "startPosition": 1914  
          }  
        },  
        "title": "Understanding and modelling wildfire regimes: an  
ecological perspective"  
      }  
    ],  
  }  
}
```