



# Data Democratization: Challenges and Opportunities

Zhong Liu<sup>1,2</sup>, Binita KC<sup>1,3</sup>, James Acker<sup>1,3</sup>, Keith Bryant<sup>1,3</sup>, and  
Mahabaleshwa Hegde<sup>1</sup>

<sup>1</sup>NASA GES DISC; <sup>2</sup>CSISS, George Mason University; <sup>3</sup>ADNET Systems, Inc.

- Introduction
- Challenges and opportunities
- Example: NASA Giovanni
- Summary



# Introduction

- An overview of challenges and opportunities of data democratization
  - Key challenges: data service and data discovery, data services, user experiences, data analysis and visualization, data quality, trustworthiness, trainings, and more.
  - Key opportunities: enabling wider access, enhanced collaboration, user centricity, data literacy training, user-friendly visualization tools, data cataloging.
- In this talk, we focus on data quality and product development, data silos, information overload, data access, visualization, and data literacy; in particular, data product development, data access, and visualization.





# Data Product Development

A recent article discussed several key areas of developing Earth science data products: *Liu, Z., Yao, T. Strategizing Earth Science Data Development. Sci Data 11, 693 (2024). <https://doi.org/10.1038/s41597-024-03531-6>* including user needs, data collection, product development, product dissemination, metrics, and implementation.





# Data Product Development (Cont.)

## User needs:

- User needs vary significantly (professionals, students, citizens).
- Collection, integration, and holistic analysis of user needs.

## Data product development:

- Ideal products: long-term, global, well-calibrated, consistent, bias-free, low uncertainty, and low-latency.
- NASA's Making Earth System Data Records for Use in Research Environments (MEaSUREs) program (proposal-based).
- Identify key variables (e.g., climate change, disasters, Essential Climate Variables (WMO)).
- Long-term commitment (e.g., funding).
- Data services to generate on-demand products.



# Data Silos – NASA Earthdata Cloud Evolution

- NASA's 12 Distributed Active Archive Centers ( [DAACs](#)) are moving the data archives they manage into the cloud.
- Benefits:
  - Easy access to data (e.g., direct access to data in the cloud)
  - Enabling science (e.g., interdisciplinary science, AI/ML)
  - Rapid deployment (e.g., computing close to data)
  - Scalability (e.g., data archive)
  - Flexibility (e.g., software and operating systems)
  - Reduced redundancy (e.g., reduction of redundant tools and services)
  - Cost-effectiveness (e.g., Users cover their storage cost in the cloud)

Source: <https://www.earthdata.nasa.gov/about/earthdata-cloud-evolution>



# Information Overload

- Too many similar datasets or variables are available in the search results

The screenshot shows the GES DISC search results page. The search term is 'precipitation'. The page displays a list of 75 datasets associated with this term. The results are organized into a table with columns for Dataset, Source, Version, Time Res., Spatial Res., Process Level, Begin Date, and End Date. The first three rows show 'GPM IMERG Late Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V06 (GPM\_3IMERGHL\_06)'. The left sidebar contains various filters such as 'Refine By', 'Features', 'Subject', 'Measurement', 'Source', and 'Processing Level'.

Search results from the GES DISC landing page: 75 datasets are associated with “precipitation”.

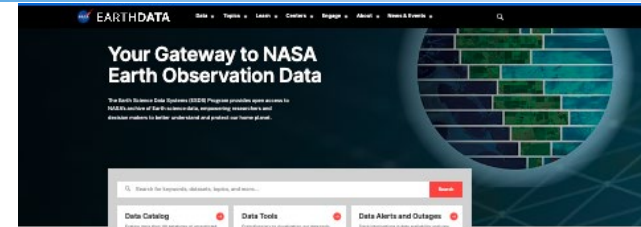
The screenshot shows the GIOVANNI search results page. The search term is 'precipitation'. The page displays a list of 156 data variables associated with this term. The results are organized into a table with columns for Variable, Units, Source, Temp. Res., Spat. Res., Begin Date, End Date, and Vert. The first row shows 'Fraction of total precipitation that is convective (NLDAS\_FORA0125\_H v2.0)'. The left sidebar contains various filters such as 'Select Plot', 'Select Date Range (UTC)', 'Select Region (Bounding Box or Shape)', 'Select Variables', 'Disciplines', 'Measurements', 'Platform / Instrument', 'Spatial Resolutions', 'Temporal Resolutions', and 'Special Features'.

Search results from GES DISC Giovanni: 156 data variables are associated with “precipitation”.



# Data Access:

- New NASA Earthdata access: the gateway to NASA Earth observation data.
- NASA Giovanni: a web application for accessing, visualizing, and analyzing Earth science remote sensing data without downloading the data.

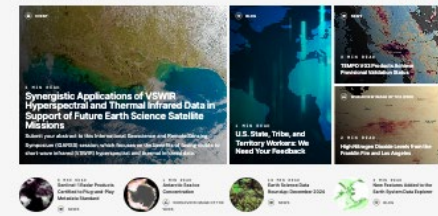


### Browse Data by Topic

Our data portfolio guides you by topic to the data, tools, and links associated for your research.

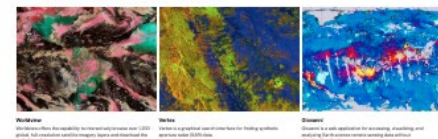


### Featured News and Events



### Featured Data Tools

From discovery to visualization, our data tools identify and streamline your research.





# NASA Giovanni

- 2000+ Earth science multidisciplinary variables.
- 22 plot types (time - averaged maps, comparison plots, vertical plots, climatology, and animations).
- Shapefiles (countries, US states, watersheds, world regions).
- Data and plots can be downloaded for further analysis.
- Results (e.g., plots) can be reproducible and shareable.

**GIOVANNI** The Bridge Between Data and Science v 4.40 Feedback Help Log out (ziuu)

High values in IMERG Early and Late in October 2024 ... [1 of 1 messages] [Read More](#)

Select Plot: Time Averaged Map Select Date Range (UTC): YYYY - MM - dd 00 : 00 to YYYY - MM - dd 23 : 59 Select Region (Bounding Box or Shape): -180, -90, 180, 90

Valid Range: 1948-01-01 to 2024-12-19  
Please specify a start date.

Select Variables

- Observations
  - Model (53)
  - Observation (53)
  - Reanalysis (50)
- Disciplines
  - Atmospheric Dynamics (44)
  - Cryosphere (1)
  - Hydrology (113)
  - Oceanography (2)
  - Water and Energy Cycle (71)
- Measurements
- Platform / Instrument
- Spatial Resolutions
- Temporal Resolutions
- Special Features
- Portal

Number of matching Variables: 156 of 2064 Total Variable(s) included in Plot: 0  
Please select at least 1 variable

Keyword: precipitation Search Clear

Variable	Units	Source	Temp. Res.	Spat. Res.	Begin Date	End Date	Vert. :
<input type="checkbox"/> Fraction of total precipitation that is convective (NLDAS_FORA0125_H v2.0)	fraction	NLDAS Model	Hourly	0.125 °	1979-01-01	2024-12-15	
<input type="checkbox"/> Convective Precipitation (NLDAS_FORA0125_M v2.0)	kg m-2	NLDAS Model	Monthly	0.125 °	1979-01-01	2024-11-30	
<input type="checkbox"/> Precipitation Rate (TRMM_3B42_Daily v7)	mm/day	TRMM	Daily	0.25 °	1998-01-01	2019-12-31	
<input type="checkbox"/> Near-Real-Time Precipitation Rate (TRMM_3B42RT v7)	mm/hr	TRMM	3-Hourly	0.25 °	2003-03-01	2019-12-31	
<input type="checkbox"/> Near-Real-Time Precipitation Rate (TRMM_3B42RT_Daily v7)	mm/day	TRMM	Daily	0.25 °	2000-03-01	2019-12-31	
<input type="checkbox"/> Precipitation (TRMM_3B42 v7)	mm/hr	TRMM	3-Hourly	0.25 °	1997-12-31	2019-12-31	
<input type="checkbox"/> Total precipitation (NLDAS_FORA0125_H v2.0)	kg m-2	NLDAS Model	Hourly	0.125 °	1979-01-01	2024-12-15	
<input type="checkbox"/> Total precipitation (NLDAS_FORA0125_M v2.0)	kg m-2	NLDAS Model	Monthly	0.125 °	1979-01-01	2024-11-30	
<input type="checkbox"/> Total precipitation rate (GLDAS_CLSM025_D v2.0)	kg m-2 s-1	GLDAS Model	Daily	0.25 °	1948-01-01	2014-12-30	
<input type="checkbox"/> Multisatellite precipitation (GPCPMON v3.1)	mm/day	AIRS,SSMI	Monthly	0.5 °	1983-01-01	2019-12-31	
<input type="checkbox"/> Wind-loss adjusted gauge precipitation (GPCPMON v3.1)	mm/day	AIRS,SSMI	Monthly	0.5 °	1983-01-01	2019-12-31	

Reset Plot Data Go to Results



# NASA Giovanni (Cont.)

**GIOVANNI** The Bridge Between Data and Science

High values in IMERG Early and Late in

Select Plot

Time Averaged Map

Select Variables

Observations

- Model (560)
- Observation (774)
- Reanalysis (730)

Disciplines

- Aerosols (274)
- Atmospheric Chemistry (236)
- Atmospheric Dynamics (769)
- Cryosphere (18)
- Hydrology (677)
- Ocean Biology (57)
- Oceanography (87)
- Water and Energy Cycle (816)

Measurements

Platform / Instrument

Spatial Resolutions

Temporal Resolutions

Wavelengths

Depths

Special Features

**GIOVANNI** The Bridge Between Data and Science v 4.40

High values in IMERG Early and Late in October 2024 ... [1 of 1 messages] [Read More](#)

Select Plot: Time Averaged Map

Select Date Range (UTC): YYYY - MM - dd 00 : 00 to YYYY - MM - dd 23 : 59 -180, -90, 180, 90

<b>Maps</b>	Scatter, Area Averaged (Static)	Time Series, Recurring Averages
<b>Time Averaged Map</b>	Scatter (Interactive) Limited to: 30000 points	
Map, Recurring Averages	Scatter (Static)	<b>Miscellaneous</b>
Time Averaged Overlay Map	Scatter, Time-Averaged (Interactive) Limited to: 30000 points	Histogram
Map, Accumulated		Zonal Mean
<b>Animation</b> Limited to: 365 time steps	<b>Time Series</b>	<b>Vertical</b>
Map, Difference of Time Averaged	Time Series, Area-Averaged Differences	Cross Section, Latitude-Pressure
	Time Series, Area-Averaged	Cross Section, Longitude-Pressure
<b>Comparisons</b>	Hovmoller, Longitude-Averaged	Cross Section, Time-Pressure
Map, Correlation	Hovmoller, Latitude-Averaged	Vertical Profile

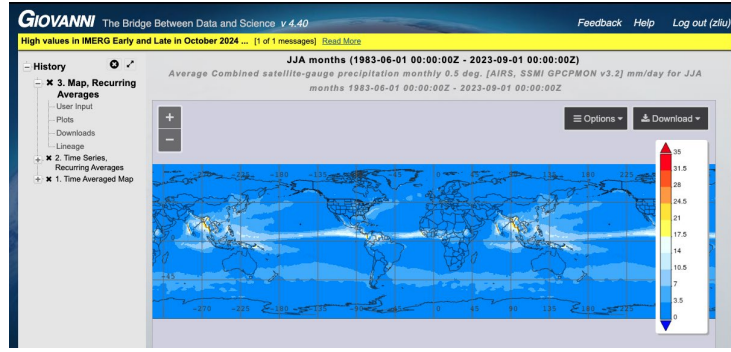
Select Region (Bounding Box or Shape)

0 : 00 to YYYY - MM - dd 23 : 59 -180, -90, 180, 90

Select a Shape...

- Countries and Areas (source: [US State Department](#))
- Lakes and Reservoirs (source: [World Wildlife Fund](#))
- Land Only file (source: [GES DISC](#))
- Sea Only file (source: [GES DISC](#))
- US States (source: [TIGER/Line\\_US Census Bureau](#))
- Watersheds (source: [Major Hydrological Basins, FAQ \(United Nations\)](#))
- World\_Regions (source: [ESRI](#))

135°00'W 90°00'W 45°00'W 00°00'E 45°00'E 90°00'E 135°00'E



JJA months (1983-06-01 00:00:00Z - 2023-09-01 00:00:00Z)

Combined satellite-gauge precipitation monthly 0.5 deg. [AIRS, SSMI GPCPMON v3.2] mm/day for JJA months 1983-06-01 00:00:00Z - 2023-09-01 00:00:00Z

Options

Layer Titles

Download

- GeoTIFF
- KMZ
- PNG
- NetCDF

Decorations

- Title, Sub-title
- Caption
- Legend

Supporting Overlays

- Coastlines
- Countries
- US States
- Grid





# Data Information and Training Resources

- Data information and training resources are available.
- Giovanni User Guide, definition for each variable, and Giovanni-related publications.
- FAQs, How-to's, dataset landing pages, and Help Desk.
- NASA Applied Remote Sensing Training Program (ARSET, <https://appliedsciences.nasa.gov/what-we-do/capacity-building/arset>)

## Giovanni Learning Resources

QUICK FILTERS

All Webinar (3)



### Create Difference Maps for NASA Data with Giovanni, Panoply, and Excel

Two fundamental ways to use Earth remote sensing data are to examine anomalies and to monitor change. During this webinar we will show you how to generate difference maps with NASA Giovanni, Panoply and Excel. Join us to learn more!

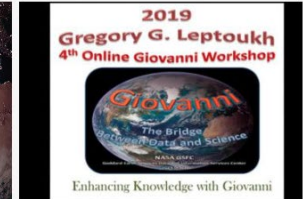
▶ WEBINAR APR 18, 2022



### Redesigning Giovanni: More Data, More Plots, Faster Results

Join us on Wednesday, 6 May, 2015, 2-3PM ET to learn how to easily access, visualize and download many of NASA's Earth science data sets using the redesigned NASA Giovanni-4 Tool.

▶ WEBINAR APR 15, 2022



### 2019 Gregory G. Leptoukh 4th Online Giovanni Workshop

Join us for the 4th Gregory G. Leptoukh Online Giovanni Workshop, hosted by NASA's Goddard Earth Sciences Data and Information Services Center (GES DISC). The theme for this workshop is "Enhancing Knowledge with Giovanni" and will showcase different ways Giovanni is used to enhance scientific knowledge.

▶ WEBINAR FEB 18, 2020

*Giovanni learning resources:*  
<https://www.earthdata.nasa.gov/data/tools/giovanni>



# Summary

- There are many barriers to democratizing Earth data.
- In this talk, we focus on:
  - Data product development
  - Data access, analysis and visualization (NASA Giovanni)
- More efforts are needed to lower the barriers.

## Information:

NASA Earthdata: <https://www.earthdata.nasa.gov/>

NASA Giovanni: <https://giovanni.gsfc.nasa.gov/giovanni/>

GES DISC: <https://disc.gsfc.nasa.gov/>

NASA ARSET: <https://appliedsciences.nasa.gov/what-we-do/capacity-building/arset>

*Reference: Liu, Z., Yao, T. Strategizing Earth Science Data Development. Sci Data 11, 693 (2024). <https://doi.org/10.1038/s41597-024-03531-6>*

