Data Democratization: Challenges and Opportunities

Zhong Liu[1,2], Binita KC[1,3], James Acker[1,3], Keith Bryant[1,3], and Mahabaleshwa Hegde[1]

[1]NASA GES DISC; [2]CSISS, George Mason University; [3]ADNET Systems, Inc.

GESDISC

Outline

- 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.

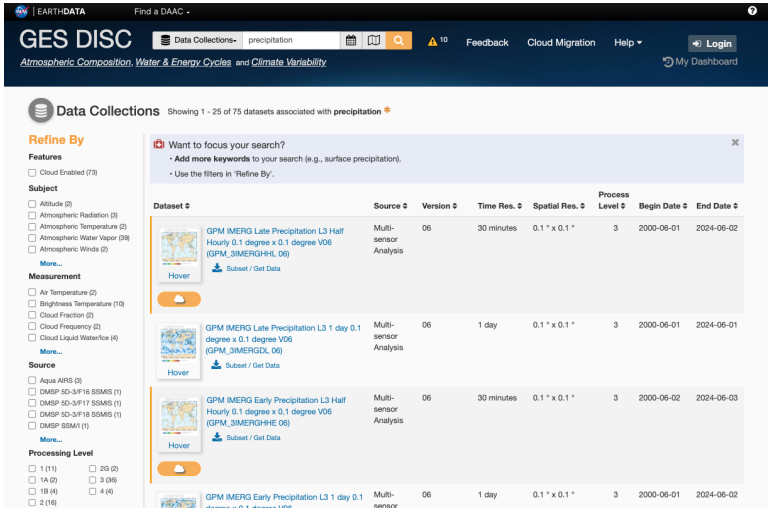
Data Silos – NASA Earthdata Cloud Evolution

- 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)

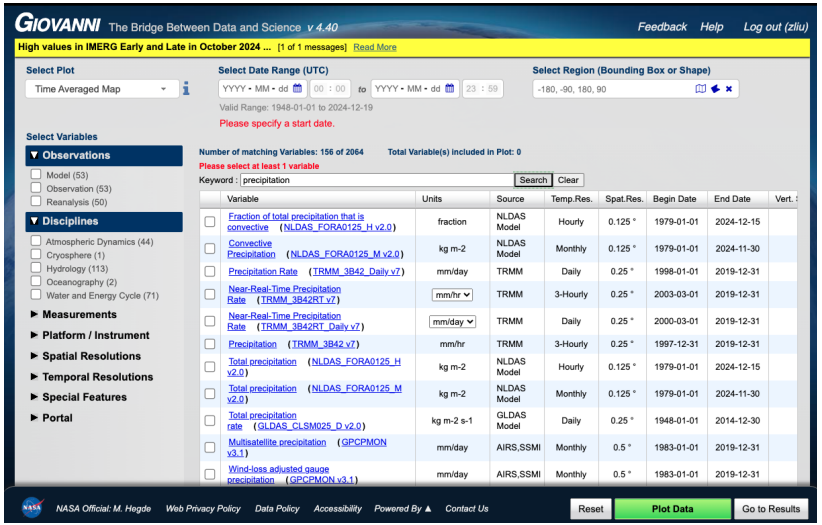
Source: https://www.earthdata.nasa.gov/about/earthdata-cloud-evolution

Information Overload

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



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

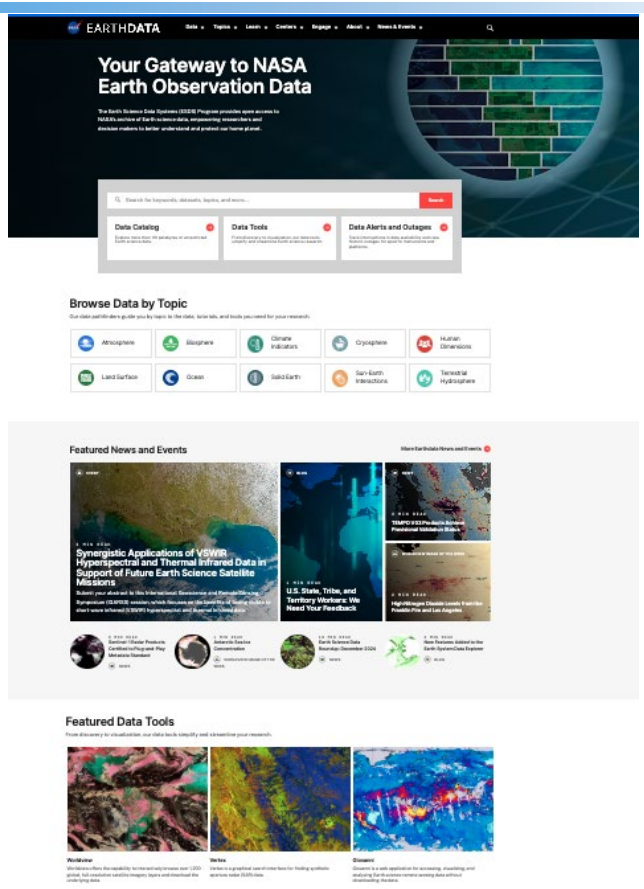


Search results from GES DISC Giovanni: 156 data variables are associated with "precipitation".

GESDISC

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.

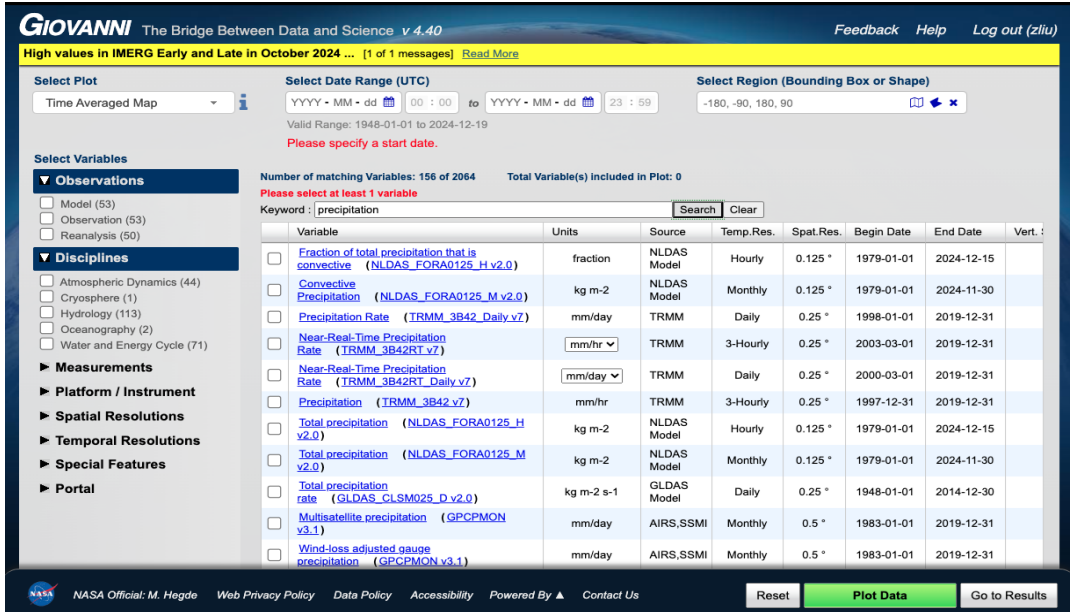


https://www.earthdata.nasa.gov/

GESDISC

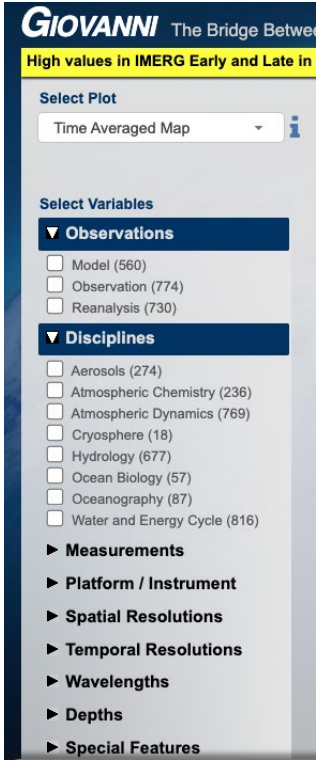
NASA Giovanni

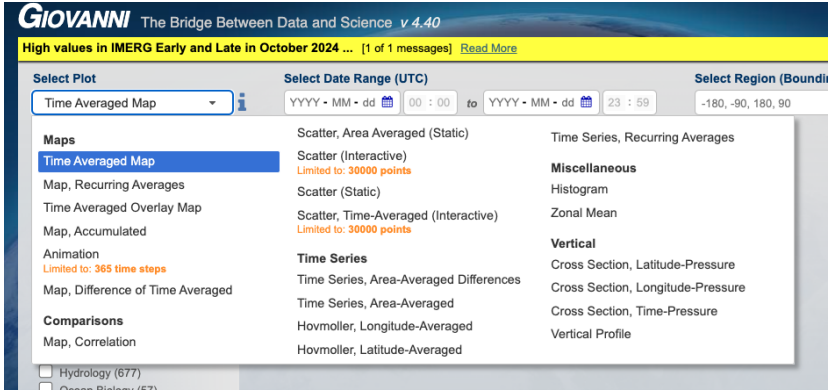
- 2000+ Earth science multidisciplinary variables.
- 22 plot types (time - averaged maps, comparison plots, vertical plots, climatology, and animations.
- Shapefiles (countries, US states, watersheds, world regions).
- Data and plots can be downloaded for further analysis.
- Results (e.g., plots) can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

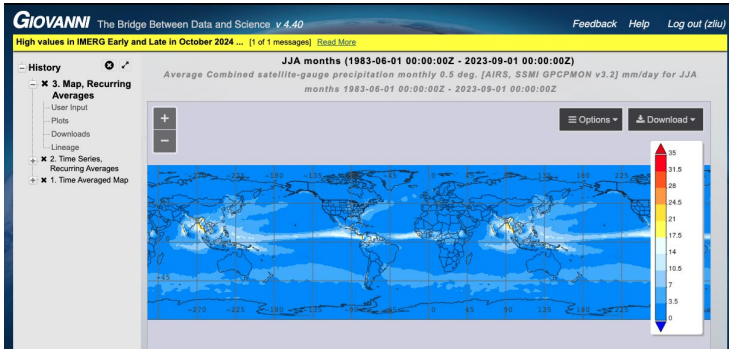


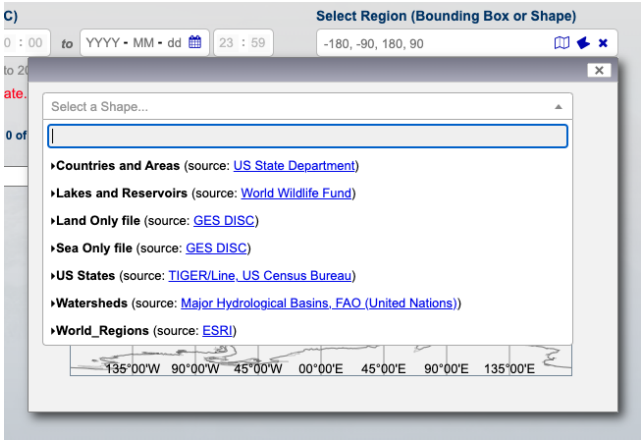
GESDISC

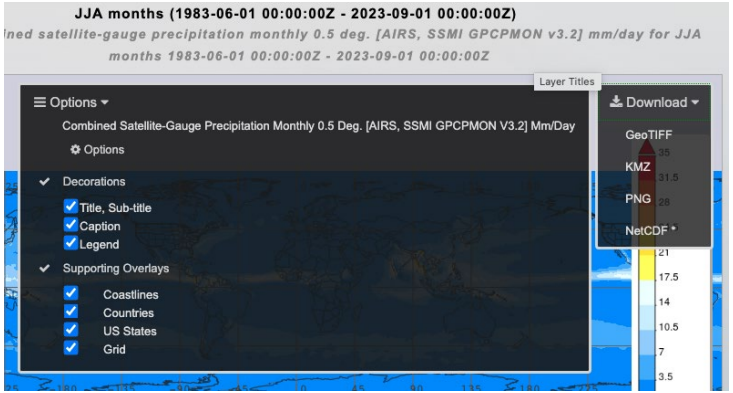
NASA Giovanni (Cont.)







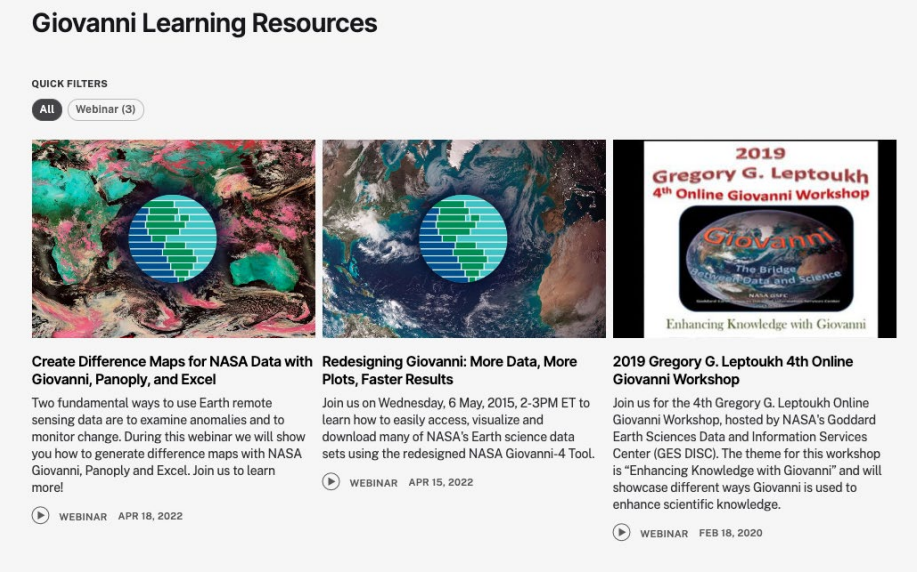




GESDISC

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: https://www.earthdata.nasa.gov/data/tools/giovanni

GESDISC

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*