Breaking Barriers: Earth Science Data for Interdisciplinary Users

Binita KC, NASA GSFC GES DISC (Adnet Inc.)
Elizabeth Joyner, NASA Goddard Space Flight Center (SSAI, Inc.)
Zhong Liu, NASA GSFC GES DISC (CSSIS, George Mason University)

AGU

December 9, 2024

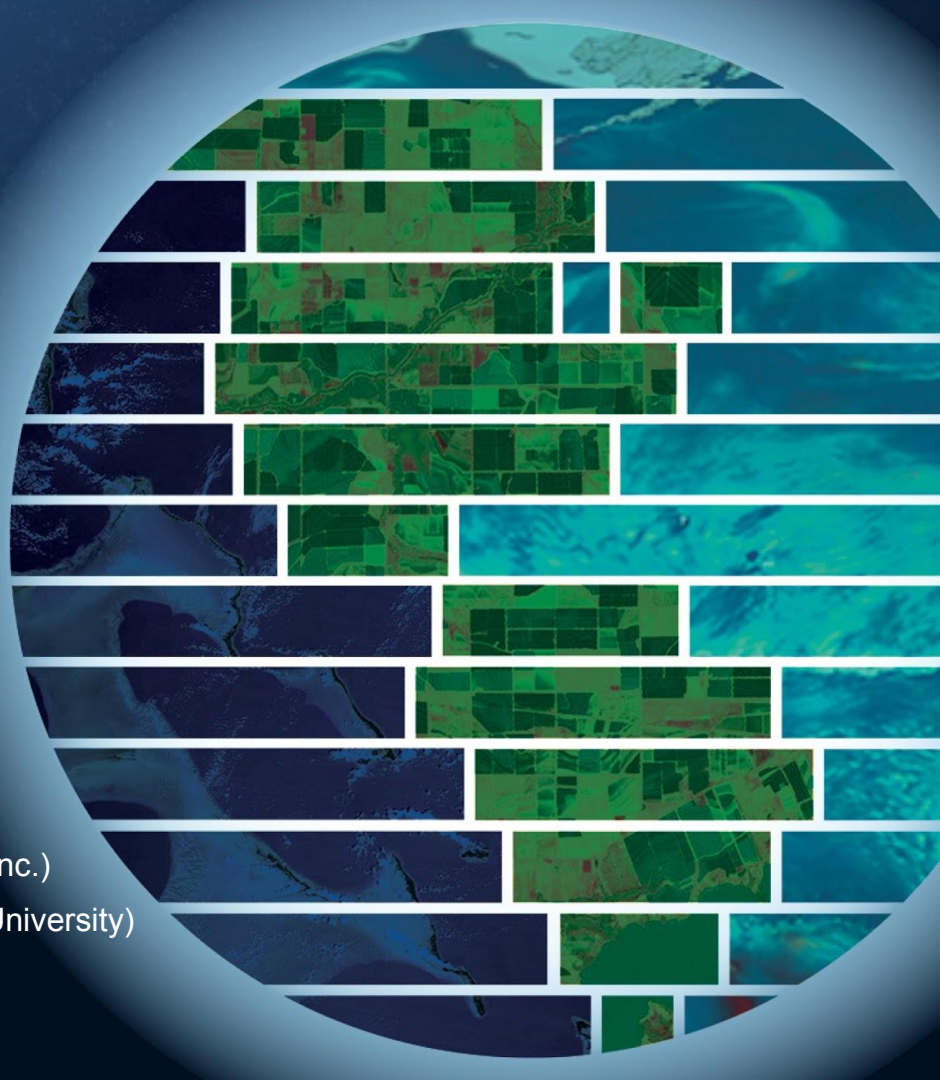
National Aeronautics and Space Administration

EARTHDATA

Breaking Barriers: Earth Science Data for Interdisciplinary Users

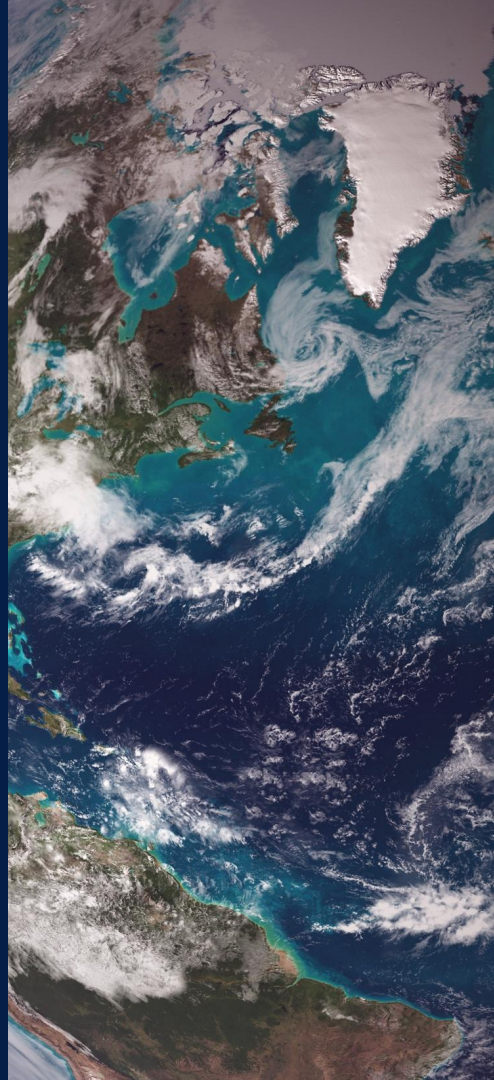
December 9, 2024

Binita KC, NASA GSFC GES DISC (Adnet Systems, Inc.)
Elizabeth Joyner, NASA Goddard Space Flight Center (SSAI, Inc.)
Zhong Liu, NASA GSFC GES DISC (CSSIS, George Mason University)



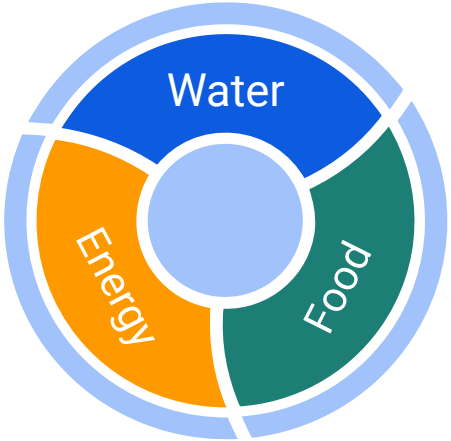
Outline

- Introduction
- User Journey
- Challenges
- Recommendations



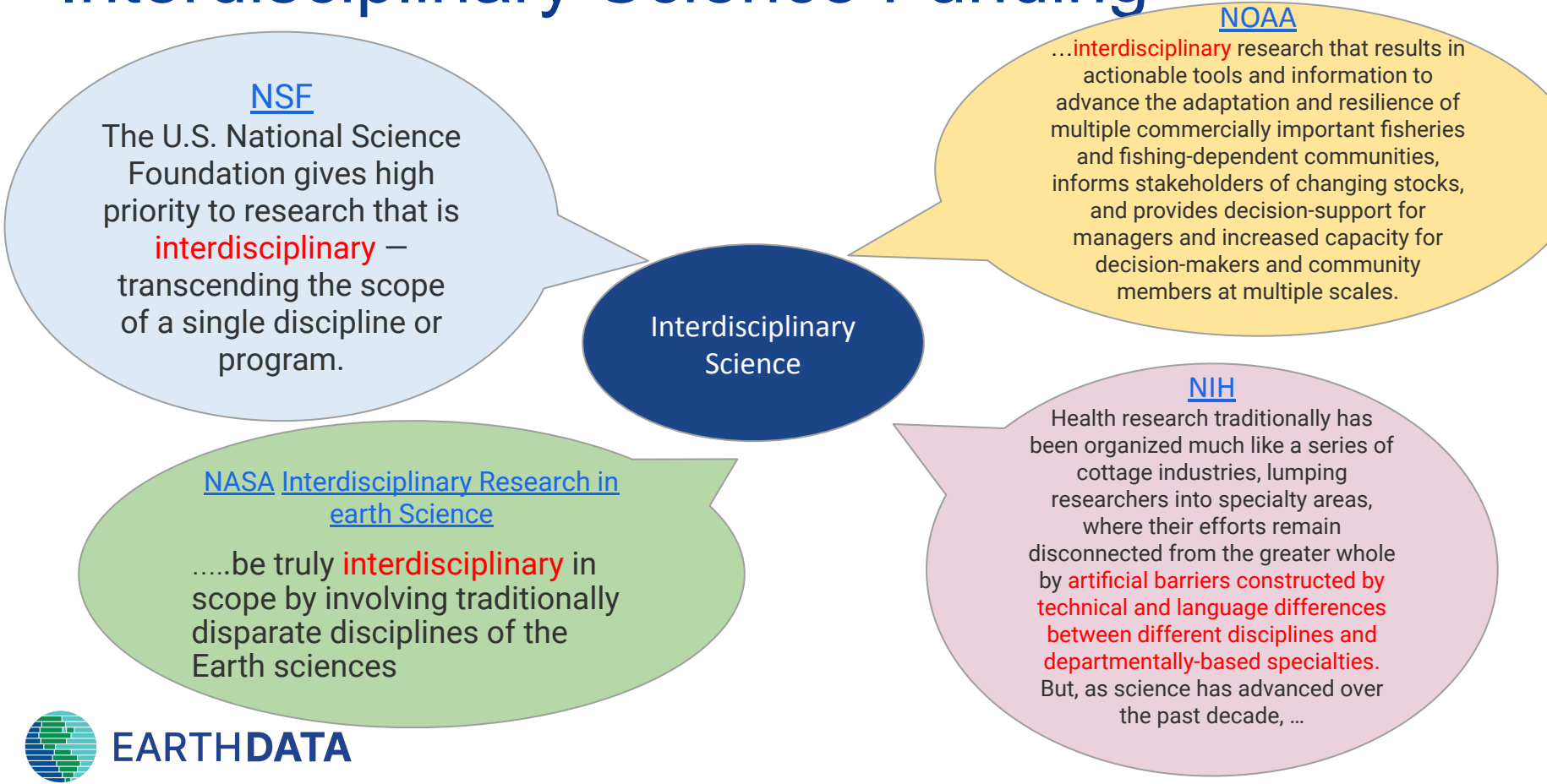
Interdisciplinary Users: Changing Data Needs

- Global challenges like climate change, biodiversity loss, and pollution require expertise from multiple disciplines (e.g., environmental science, socio-economics, policy, engineering).
- Users from different fields need varied datasets, including environmental data, socio-economic data, and policy-related data. For example, water, energy and food
- Combining data from various disciplines allows for a more comprehensive understanding of environmental problems.



EARTHDATA

Interdisciplinary Science Funding



EARTHDATA

Your Data Journey

YOU

Start Here

Explore

Narrow Down

Customize Data

Access Data

Analyze

Publish/ Document

Develop Applications

A Researcher's Journey

Indicates steps for an advanced researcher

1 Explore:

- Review literature (publications)
- Identify the major variable of interest and search for data
- Quickly visualize the primary variable and co-locate
- Find introductory information on data
- Explore available resources- tools, service
- Explore related data through publication
- Search for data and documentations cross-DAAC

2 Narrow Down:

- Determine spatial, temporal, resolution, time span, coverage, latency, processing level, platform
- Determine data quality, limitations
- Subset and download sample data

3 Access Data:

- Use services and tools
- Computing resources available
- Determine Data volume/granules needed
- Tools, software available
- Use case/tutorials

Access Data:

- Cloud access
- Web/API services
- Programmatic Chunking/Subsetting
- Metadata/Pathfinders
- Downlinking/Data logger

4 Customize Data:

- Processing to convert to grid (for non gridded)
- Data format conversion
- Regrid/reproject
- Fill missing data
- Bias correction

Customize Data:

- Cleaning-consistency checks
- Transformation and Reformatting
- Data enrichment
- Prepare data for interoperability with different systems

5 Analyze:

- Co-location
- Time series
- Change Detection
- Correlation
- Regression
- Machine learning/AI
- Trend Analysis
- Zonal mean

6 Publish/Document:

- Cite the Data (DOI), documents, service/tools
- Mention data limitations and caveats
- Data quality

7 Develop Applications

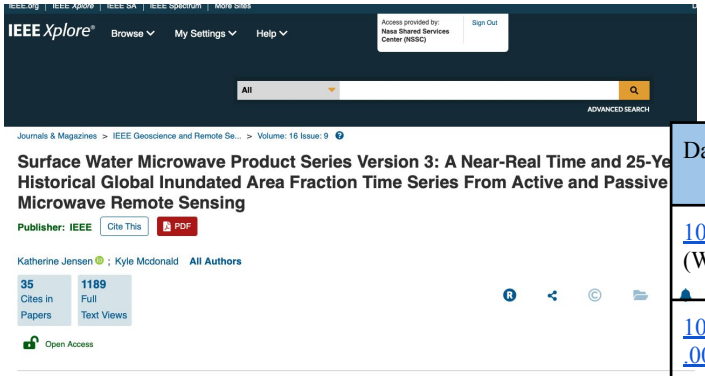
- Web/Mobile/Cloud ready
- Open Science and FAIR
- 508 compliant
- Operational support if required
- In support of data and business continuity for stakeholders
- For Internal-NASA and External stakeholders

Publish/Document:

- Publish research with data

EARTHDATA

Use Case



Jensen, K., & Mcdonald, K. Surface Water Microwave Product Series Version 3: A Near-Real Time and 25-Year Historical Global Inundated Area Fraction Time Series From Active and Passive Microwave Remote Sensing. *IEEE Geoscience and Remote Sensing Letters*, **16**(9), 1402–1406 (2019). https://doi.org/10.1109/LGRS.2019.2898779

- Use case from a published journal article referenced seven EOSDIS data collections from five DAACs: PO.DAAC, ASF DAAC, ORNL DAAC, LP DAAC, and NSIDC DAAC .
- This indicates that interdisciplinary researchers often need to source data from multiple DAACs for their studies, necessitating familiarity with various instruments, methods, tools, formats, and software packages tailored for that particular data.

Data Collection DOI	DAAC	Data Format	In CMR	Processing Level	Available Data Services
10.3334/ORNLDAAC/1247 (Wieder, 2014)	ORNL DAAC	NetCDF-4	yes	4 - Gridded Model Output	THREDDS
10.5067/MODIS/MOD10C2.006 (Hall, 2016)	NSIDC DAAC	HDF-EOS2	yes	3 - Gridded Observations	Spatial Subsetting, GeoTIFF reformatting, Reprojection, Resampling
10.5067/J0CJG7RZW3IJ (Knowles, 2004)	NSIDC DAAC	Binary	no	unknown	EASE Grids
10.5067/KJVERY3MIBPS (Armstrong, 2005)	NSIDC DAAC	Binary	no	unknown	EASE Grids
10.5067/MODIS/MOD13C1.006 (Didan, 2015)	LP DAAC	HDF-EOS2	yes	3 - Gridded Observations	OPeNDAP
10.5067/SWAMPS3.2 (McDonald, 2018)	ASF DAAC	GeoTIFF	no	unknown	none
10.5067/OSXXX-L1B02 (JPL QuikSCAT Project, 2006)	PO.DAAC	HDF4	yes	1B - Radiance, Sensor Coordinates	none

EARTHDATA

Challenges in Data Discovery

- **Citation and Traceability Issues:** Challenges in locating datasets from journal publications due to improper citation practices such as missing DOI, making it difficult to trace and verify data sources.
- **Complex Dataset Naming Conventions:** Difficulty navigating datasets with complex and technical naming conventions, which can be especially challenging for users without domain-specific expertise.
- **Lack of Interdisciplinary Connections:** Limited linkages between datasets at the variable level (e.g., connecting "NO2, population" or "evaporation, crop yield") hinder cross-disciplinary research and analysis.
- **Limited Advanced Visualizations:** Limited support for 3D visualizations and challenges in comparing variables at a location and and lack of sample files for data preview.
- **Simplified Dataset Information:** Difficulty finding simplified information about datasets to support understanding and usability especially for users without domain knowledge **Limited Search Functionality:** Limited options to search for data based on application areas and cross discipline.

EARTHDATA

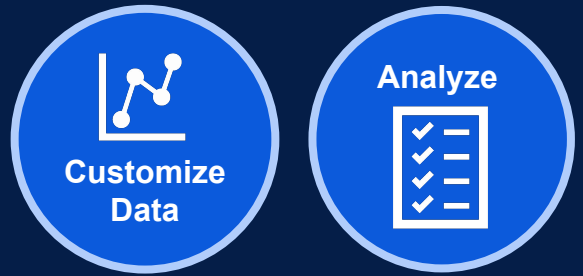
Challenges in Accessing Data

- **Data Accessibility vs. Availability:** Discrepancies between data being available online and its actual accessibility, particularly in low-bandwidth areas where downloading large datasets is impractical.
- **Fragmented Tools and Services:** Interdisciplinary users must rely on multiple tools and services, often switching between them, as these tools are not designed to cater to the diverse needs of interdisciplinary users or support seamless, end-to-end processes.
- **Niche Tool Design:** Many tools are created for specialized, domain-specific users, limiting their usability and applicability for cross-disciplinary research and collaboration.
- **Device Dependence:** Data access is predominantly restricted to desktop and laptop computers, with limited options for mobile-friendly or cloud-based access, further hindering flexibility and accessibility.

EARTHDATA

Challenges in Data Customization

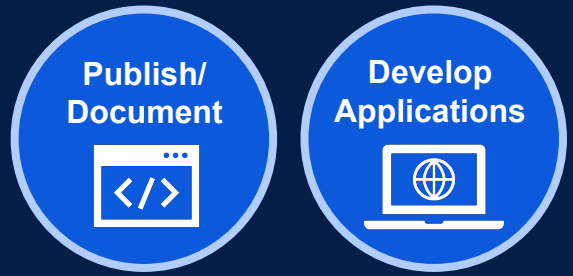
- **Navigate Diverse Data Formats and Structures:** Datasets are available in varying grid sizes, projections, and formats (e.g., NetCDF, HDF, GRIB, GeoTIFF/COG), making integration and analysis challenging for users who lack technical expertise.
- **Limited Tools for Non-Coding Users:** User-friendly tools that support data customization and analytics for non-programmers, creating a barrier for broader data use.
- **Subsetting Data for planning:** Tools often lack the capability to subset datasets based on administrative boundaries or specific planning scales which are used for decision-making.
- **Complexity of Level 2 Data:** Level 2 datasets are complex and require expertise for co-locating with gridded data.
- **Basic Analytics Gaps:** Limited tools offer fundamental analytics like basic statistics, time series, trends, accumulation, and zonal means etc.
- **Scalability Issues:** Existing tools often struggle to handle large-scale datasets efficiently, leading to performance bottlenecks and limiting their usability for high-resolution or long-term data analysis.



EARTHDATA

Challenges with Publication and Building Applications

- **Improper Data Citation:** Data are often not properly cited using DOI, making difficult to trace the datasets back to its original source/version.
- **Long Term Availability of data:** Developing an application requires that the dataset is continuously available and, when the mission ends, proper instructions for the follow-on mission are needed.
- **Integration Across Disciplines:** Applications often struggle to integrate datasets from different fields due to inconsistencies in formats and standards.



EARTHDATA

Recommendations

Steps

Recommendations

Explore Data

1. **Improve Dataset Traceability: L**ocate the datasets from publications citing the datasets

2. **Link Related Variables:** based on usage patterns in journal databases and publication metrics

3. **Develop Data Pathways for Interdisciplinary Use Cases**

4. **Enhance Metadata Standards Catering Interdisciplinary Users**

Access Data

1. **Develop an Enterprise Data Tool:** Visualize, subset and analyze the data irrespective of data platform, formats, levels (L2, L3). End-to-End services

2. **Unified User Interface**

3. **Simplify Data Documentation:** reduce technical requirements

4. **Optimize Data File Sizes for Accessibility:** ensure accessibility for users in low-bandwidth regions

5. **Support mobile Access**: Expand data accessibility beyond computers

Customize

1. **Services for Non-Technical Users:** no-code or low-code interfaces, interactive visualizations, and guided workflows

2. **Convert L2 data to gridded Form with options provided:** including validation tools to ensure the accuracy and usability of gridded data for analysis

3. **Administrative Boundary, Tract, and Point Subsetting:** subsetting by regions or planning scales for policy and decision-making

Analyze

1. **Co-location of geophysical observations from across disciplines**

2. **Perform Analytics-Rapid time series and trend analysis:** trend analysis, zonal statistics, anomaly detection and time-series

3. **Provide code snippets for easy access to data**

3. **API Access for Custom Analysis:** programmatically access datasets and perform analytics, integrating with custom workflows and external applications.

Publish

1. **Citation of Data and services using DOI**

2. **Open sourcing and community contribution**

3. **Data Provenance and reproducibility:** robust protocols for documenting data origins, transformations, and processing steps to establish clear data provenance.

4. **Foster Collaboration and Feedback:** Forums to share best practices, seek guidance, and provide feedback to improve tools and workflows

EARTHDATA

earthdata.nasa.gov

Thank you

binita.kc@nasa.gov

Acknowledgement:
Interdisciplinary Community Needs (ICN)
Earth Science Data System Working Groups (ESDSWG)