



Improving the Discoverability of NASA Airborne and Field Investigation Data

January 26th, 2022

Deborah Smith, Stephanie Wingo, Camille Woods, Shelby Bagwell, Carson Davis, Ashlyn Shirey, Lucia Alonso Guzman, Danielle Groenen, Emily Foshee, Jillian Ethridge, Ge Peng, Anthony Lukach, Alice Rühl, Necoline Hubner, Rahul Ramachandran



Discoverability of Data

Looking for something?

Data can be much harder to find and use when:

- Distributed across many locations
- Inconsistently organized
- Is very heterogeneous and varied
- Available in different formats
- Has missing or incorrect metadata
- Is stored separately from other needed data
- Is not publicly available or searchable

This was the case for NASA's airborne and field data



Image source: [Azurtri](#)

Creating Organization Isn't Easy



NASA airborne and field data have historically suffered from inattention compared to NASA satellite data

Our goals are to:

- **Bring more consistency** to how scientists find and obtain airborne and field data from NASA
- **Provide important contextual details** of the data collection environment
- Ensure the **availability, discovery** of and access to these valuable NASA data



Image source: [Azurtri](#)

Airborne Data Management Group (ADMG)

The **Airborne Data Management Group** (ADMG) was officially established by the NASA Earth Science Data System management in Sep 2018 as part of IMPACT located at NASA MSFC in Huntsville, AL

ADMG Primary Tasks:

- Take a **full assessment** of NASA Airborne Earth Science data
- Construct a public, centralized, **metadata-rich inventory** of airborne and field investigations, platforms, instruments, and data product access
- Develop systematic **approaches and best practices** that bring **consistency and expediency** to airborne and field data stewardship
- Maintain a **knowledge center** containing important information and document access, and simplified access to **airborne tools** and **use cases**
- Improve communication between the DAACs, airborne campaign investigators, ADMG, and other stakeholders

Assessment of NASA Airborne and Field Data

Assessment Goals:

- To identify and **locate all** NASA airborne Earth science data
- To **identify existing metadata** and metadata issues
- To **identify issues** in airborne data archival, discovery and access



To date, we have identified over 150 airborne/field investigations

Identify all airborne data at DAACs

Data that are discoverable in CMR do not represent all DAAC data

Identify Data at other NASA Repositories

Access any data external to DAACs that require publication

Google Scholar Searches

Publications mention airborne field campaign data that exists somewhere

NASA Ames Airborne Science

Work with Airborne science program to find what Earth science to publish

NASA LaRC Atmospheric Chemistry

Data from many atmospheric chemistry airborne campaigns are at LaRC

A Holistic View of Campaigns

- “**Relating to or concerned with wholes** or with complete systems rather than with the analysis of, treatment of, or dissection into parts” *Merriam Webster*
- Looking at the **BIG PICTURE**
- Campaign data products do not exist in isolation, but are part of a **complete investigation** with a scientific purpose
- **Contextual Information is essential** - details connecting all the parts of a campaign and explaining the data collection environment are needed for best data use and reuse



Image source: NASA

CASEI: The Catalog of Archived Suborbital Earth Science Investigations

- CASEI contains detailed information of airborne and field campaigns and points to data access using data product DOIs
- Additional contextual details provide users with **flexible discovery** and a greater understanding of campaigns, platforms, and instruments
- Built for scientists, users can explore and discover the inventory via the **highly-linked web-based user interface** to find information across all campaigns, data centers, people, topics, etc.
- Filters provide users with **many different ways to search** through the linked information



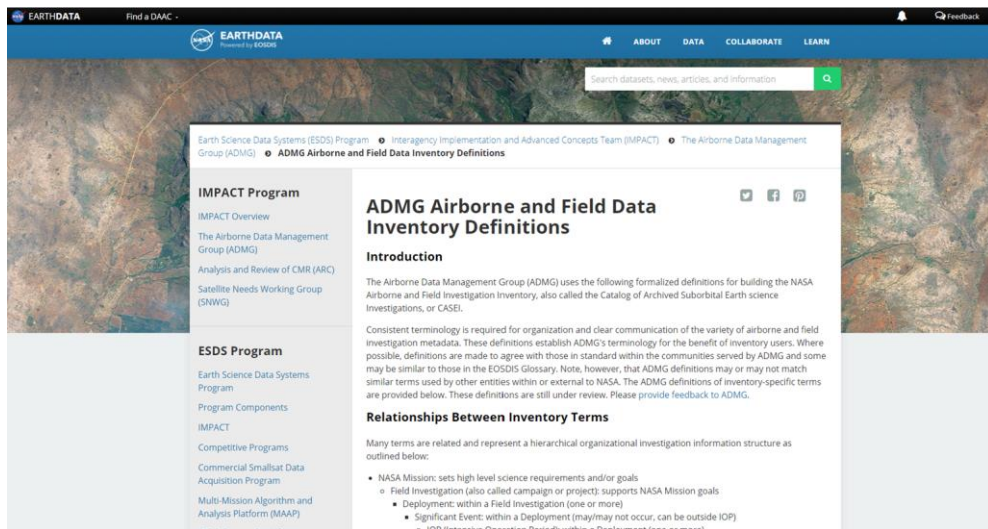
CASEI was developed by the ADMG with help from Development Seed team and other IMPACT team members



CASEI Step 1: Definitions

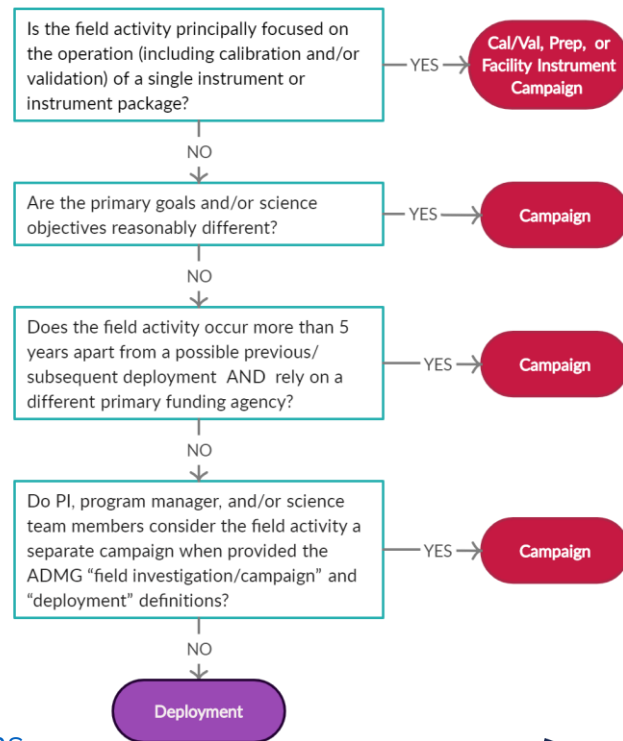
Definitions and Decision Trees:

Consistent use of carefully defined terms in all decision making



Definitions of Common Terms:

- <https://earthdata.nasa.gov/esds/impact/admg/admg-definitions>



CASEI Step 2: Holistic Metadata Model

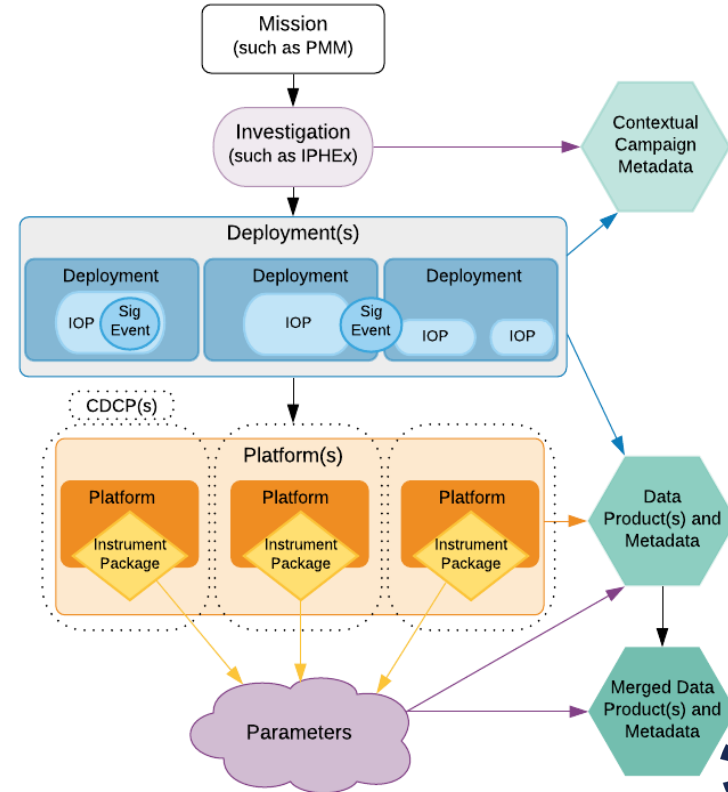
For each campaign, CASEI provides:

- Overviews, acronyms, aliases, logos
- Dates, season, region, coverage details
- Descriptions of deployments and intensive operating periods (IOPs)
- Significant events
- Leadership and management details, funding
- Data product DOIs

- **All information is highly inter-linked**

CASEI also provides metadata for the platforms and instruments used during the campaigns

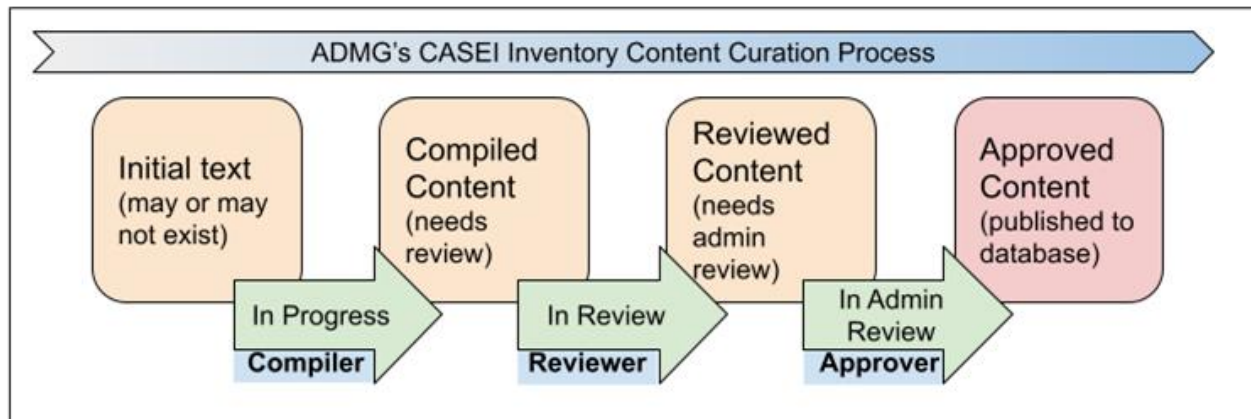
- Platforms
 - type
 - owner
 - image (if available)
 - GCMD name
- Instruments
 - measurement style/type
 - owner
 - measurement region
 - resolutions
 - GCMD name



CASEI Step 3: Team Curation

Information Curation Process

All metadata curated and closely reviewed by three curation staff



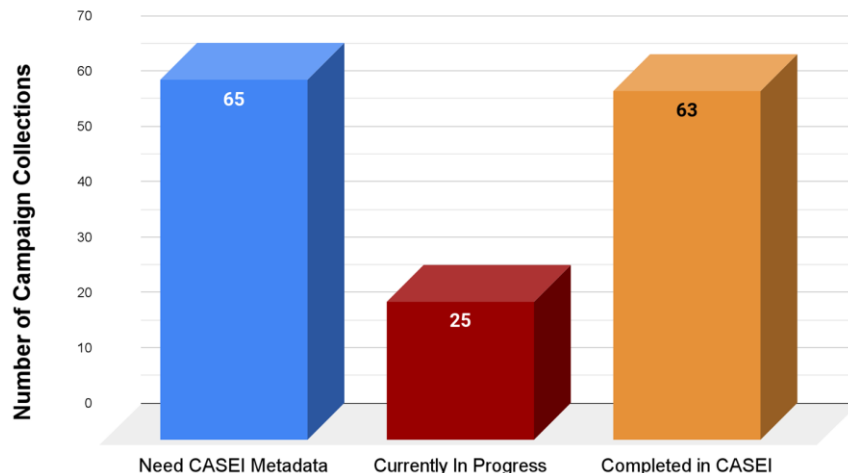
Where is CASEI metadata found:

- Campaign reports
- Peer-reviewed journal articles describing the campaign
- NASA websites (such as the Airborne Science Program)
- NASA DAAC (Distributed Active Archive Center) landing pages for campaigns
- Annual reports
- Other credible and authoritative resources

Inventory Progress

- Current list of known campaigns contains 153 named campaigns or NASA airborne activities
- CASEI will ultimately contain detailed metadata for all known campaigns regardless of data publication status

Total Known Campaigns as of Dec 2021: 153



The above numbers include campaigns assigned to a DAAC and those not yet assigned to a DAAC

Contains 28%
of known NASA
campaigns

Includes 50%
more contextual
metadata than CMR

Over 500
instrument
descriptions

3 sets of eyes
on every bit of
added metadata

Step 4: CASEI Structure

User Portal - public user interface and API

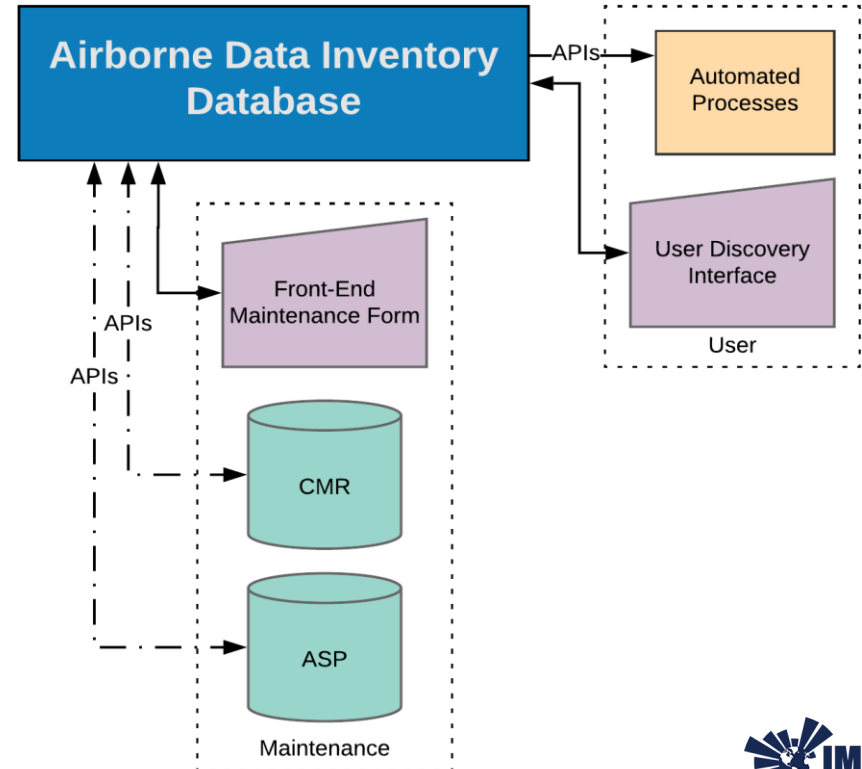
- submit queries
- obtain information
- access data
- find campaign details

Maintenance Interface - information curation

- add new information
- validate existing information with routine checks for errors
- add missing data and fix broken links
- access inventory user statistics

APIs for external database access

- NASA Common Metadata Repository (CMR)
- NASA Airborne Science Program database



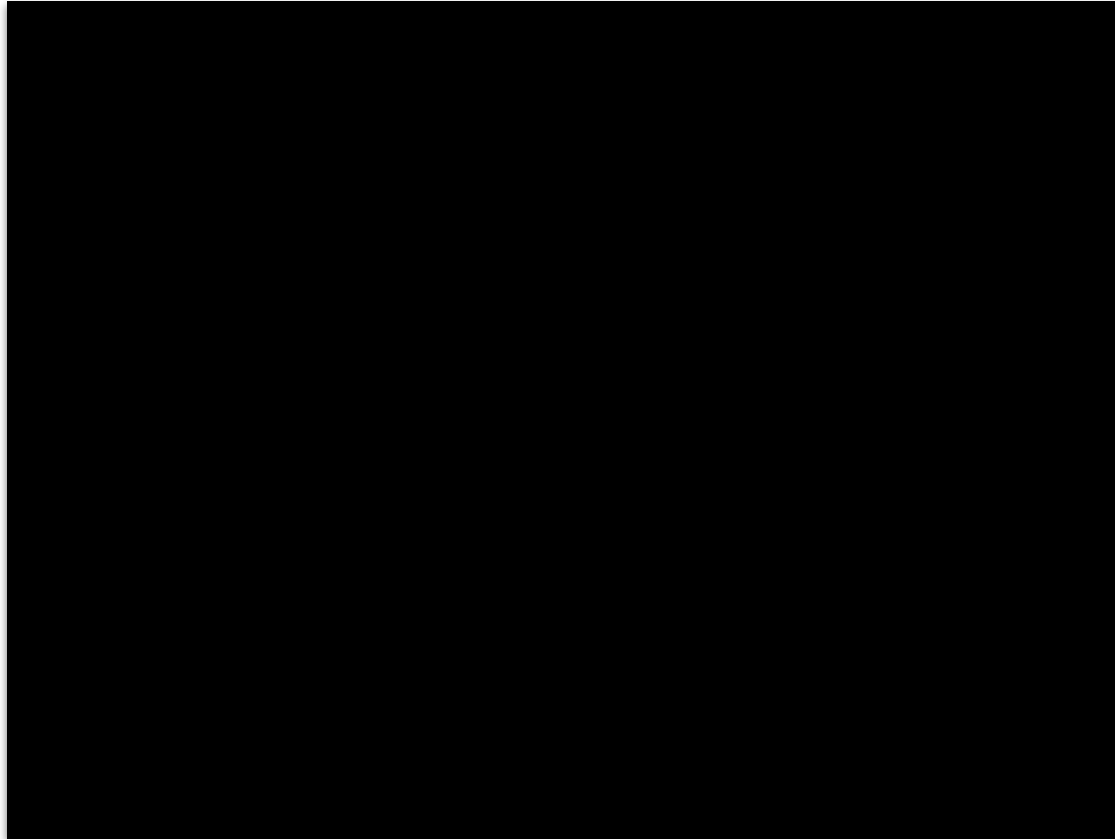
Step 5: CASEI Design and Function

The image displays three overlapping screenshots of the NASA CASEI (Catalog of Archived Suborbital Earth Science Investigations) web application. The leftmost screenshot shows the main search interface with filters for DATE RANGE, FOCUS AREA, GEOGRAPHICAL CONCEPT, SEASON, GEOGRAPHICAL REGION, PLATFORM, and FUNDING AGENCY. Below the filters is a world map with several rectangular boxes indicating search areas. The middle screenshot shows the 'Olympic Mountains Experiment' (OLYMPLEX) campaign details, including a map of the study area, statistics (1 DEPLOYMENTS, 58 COLLECTION PERIODS, 36 DATA PRODUCTS), and an overview section. The rightmost screenshot shows the 'Douglas DC-8' instrument details, including a photo of the aircraft, statistics (9 CAMPAIGNS, 11 FLIGHTS), and a detailed 'Instrument Details' table.

INSTRUMENT TYPE	N/A	LEAD INVESTIGATOR	Paul Lawson, Ralf Fiedler
INSTRUMENT RESPONSIBILITY	CLOUD, ATMOSPHERE, CLIMATE, MICROPHYSICS, PARTICULATE SIZE DISTRIBUTION, CLOUD, ATMOSPHERE, CLOUD MICROPHYSICS	TECHNICAL CONTACT	Paul Lawson
INSTRUMENT RESOURCES	14 customer profile	INSTRUMENT MANAGER	Paul Lawson
NO. OF DEPLOYMENTS	0 / 1	PLANNED DATES	10 years of 2000-2010
INSTRUMENT FREQUENCY	10 years of 2000-2010	DATA LOCATIONS	301.8 TIG
FORM DETAILS	https://www.nasa.gov/.../DC-8/	INSTRUMENTS	https://www.nasa.gov/.../DC-8/
MISSION	https://www.nasa.gov/.../DC-8/	INSTRUMENTS	https://www.nasa.gov/.../DC-8/

- Single information page approach
- Interconnected campaigns, platforms, and instruments
- Heterogeneous data products connected via 1 interface
- Paid careful attention to:
 - What are the common user goals?
 - What are the search paths to prioritize?
 - Where do users run into informational dead ends?

CASEI Design and Function



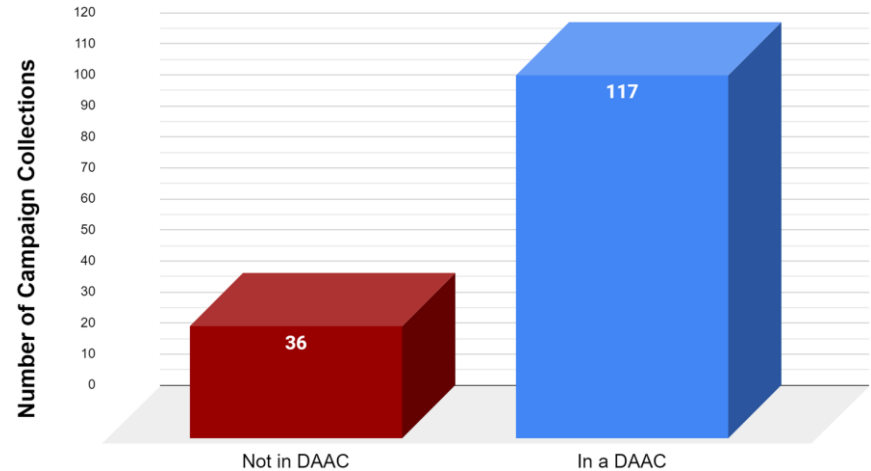
Data Archeology

ADMG aids in providing open access to **all** NASA airborne and field data

We **locate historical data** and drive the historical data publication process workflow

- Identify and summarize historical campaigns
- Locate data and personnel with data knowledge
- Recommend data center assignments
- Work with data centers to ensure effective data transfer and management
- Function as science team proxy if needed
- Ensure detailed information added to CASEI

Total Known Campaigns as of Dec 2021: 153



Making Change for the Future

The key to improving future airborne and field data is to change data management

ADMG is building a comprehensive set of best practice documents that

- encourage consistent term usage
- consistent organization and archival decisions
- good communication
- known responsibilities and requirements



Image source: Ross Findon, Unsplash

Improvements can already be seen from recent efforts!

Final Thoughts

Via CASEI and the other efforts of ADMG, NASA's airborne and field data are becoming **more open and easier to find**

The **holistic collection of needed contextual information** improves understanding and use of data

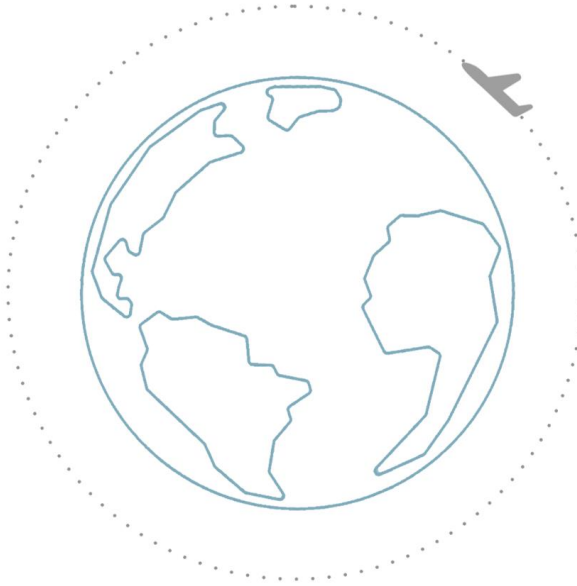
These ADMG efforts **facilitate data reuse** and enhance the return on NASA's investment in suborbital research



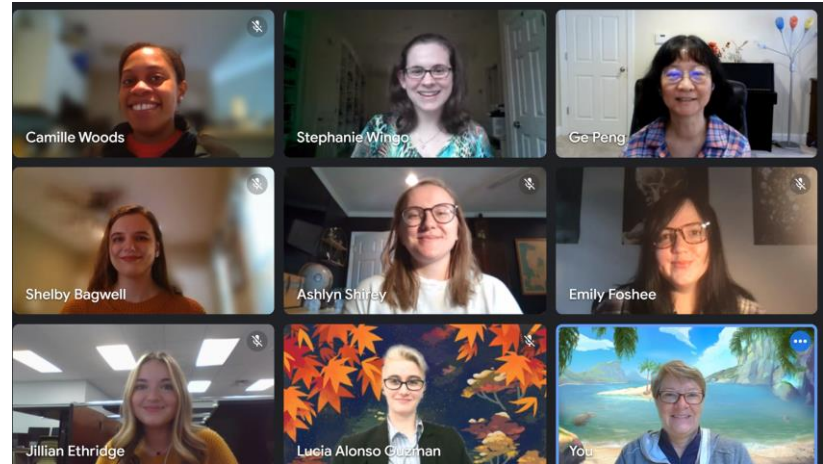
Explore CASEI: <https://impact.earthdata.nasa.gov/casei/>



NASA | CASEI



This work is supported by NASA Grant NNM11AA01A as part of the Interagency Implementation and Advanced Concepts Team (IMPACT).



UL to LR: Camille Woods, Stephanie Wingo, Ge Peng, Shelby Bagwell, Ashlyn Shirey, Emily Foshee, Jillian Ethridge, Lucia Alonso Guzman, Deborah Smith, Not pictured: Dr. Danielle Groenen