

Stewardship Best Practices for Improved Discovery and Reuse of Heterogeneous and Cross-Disciplinary Earth System Data

Peng Ge¹, Smith Deborah¹, Wingo Stephanie¹, and Ramachandran Rahul²

¹University of Alabama in Huntsville

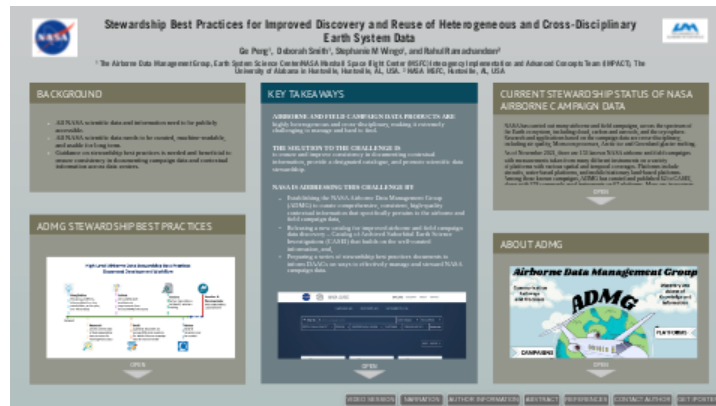
²NASA Marshall Space Flight Center

November 16, 2022

Abstract

Some of the Earth system data products such as those from NASA airborne and field investigations (a.k.a. campaigns), are highly heterogeneous and cross-disciplinary, making the data extremely challenging to manage. For example, airborne and field campaign measurements tend to be sporadic over a period of time, with large gaps. Data products generated are of various processing levels and utilized for a wide range of inter- and cross-disciplinary research and applications. Data and derived products have been historically stored in a variety of domain-specific standard (and some non-standard) formats and in various locations such as NASA Distributed Active Archive Centers (DAACs), NASA airborne science facilities, field archives, or even individual scientists' computer hard drives. As a result, airborne and field campaign data products have often been managed and represented differently, making it onerous for data users to find, access, and utilize campaign data. Some difficulties in discovering and accessing the campaign data originate from the incomplete data product and contextual metadata that may contain details relevant to the campaign (e.g. campaign acronym and instrument deployment locations), but tend to lack other significant information needed to understand conditions surrounding the data. Such details can be burdensome to locate after the conclusion of a campaign. Utilizing consistent terminology, essential for improved discovery and reuse, is also challenging due to the variety of involved disciplines. To help address the aforementioned challenges faced by many repositories and data managers handling airborne and field data, this presentation will describe stewardship practices developed by the Airborne Data Management Group (ADMG) within the Interagency Implementation and Advanced Concepts Team (IMPACT) under the NASA's Earth Science Data systems (ESDS) Program.

Stewardship Best Practices for Improved Discovery and Reuse of Heterogeneous and Cross-Disciplinary Earth System Data

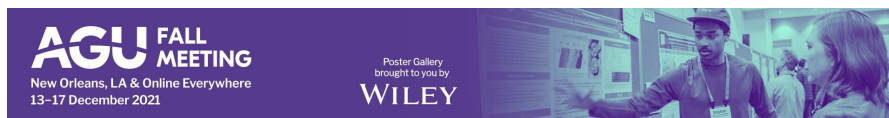


Ge Peng¹, Deborah Smith¹, Stephanie M Wingo¹, and Rahul Ramachandran²

¹ The Airborne Data Management Group, Earth System Science Center/NASA Marshall Space Flight Center (MSFC) Interagency Implementation and Advanced Concepts Team (IMPACT), The University of Alabama in Huntsville, Huntsville, AL, USA. ² NASA MSFC, Huntsville, AL, USA



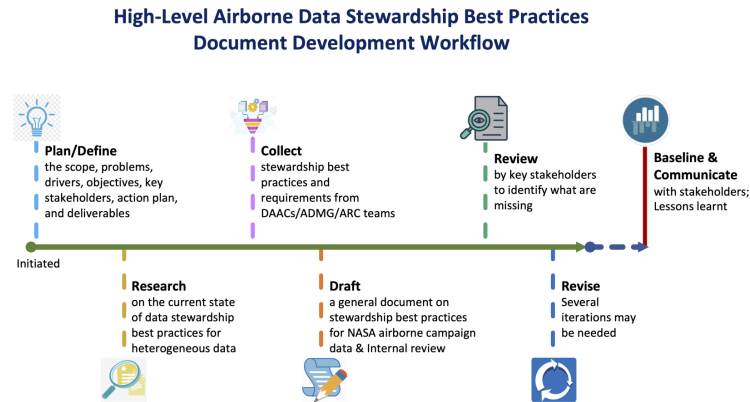
PRESENTED AT:



BACKGROUND

- All NASA scientific data and information need to be publicly accessible.
- All NASA scientific data needs to be curated, machine-readable, and usable for long term.
- Guidance on stewardship best practices is needed and beneficial to ensure consistency in documenting campaign data and contextual information across data centers.

ADMG STEWARDSHIP BEST PRACTICES



The following high-level recommended practices apply to all types of data and information from Earth science airborne and field campaigns that NASA supports or participates in:

- Communicate clearly and effectively with campaign data producers and users across multiple disciplines. Specifically, it is recommended to:
 - Interact with and support the science team during the campaign to gain better understanding of campaign information,
 - Use and encourage campaign scientists' use of controlled vocabularies as appropriate,
 - Develop a DAAC* data management plan to clearly communicate how the DAAC will care for the data products after they are obtained from the data producers,
 - Ensure important contextual information is obtained from the science team before team members move on to other campaign efforts or field archives are removed.
- Maintain a comprehensive and consistent set of high-quality information and documentation for every campaign. Specifically, it is recommended to:
 - Be consistent in the organization of data and information across campaigns,
 - Establish a discoverable and consistent campaign "landing page",
 - Ensure the DAAC landing page content is machine accessible.
- Actively manage NASA campaign data and information as a valuable agency asset. Specifically, it is recommended to:
 - Adhere to NASA standards and conventions in data stewardship,
 - Ensure and maintain quality of ingested and released data and information,
 - Ensure cross-connection of information and use of networking tools to improve information access and education for user communities.

Additional complementary, domain specific documents are under development for specific types of the airborne and field campaign data; these include but are not limited to:

- Atmospheric chemistry data,
- Ground and airborne radar and lidar data,
- Unmanned Aircraft System (UAS) data.

* DAAC - Distributed Active Archive Centers

Peng, G., D. Smith, S.M. Wingo, and R. Ramachandran, 2021: Stewardship Best Practices for Documenting and Representing NASA Airborne and Field Campaign Data and Information. *Airborne Data Management Group*. Version: 2021-11-15.

KEY TAKEAWAYS

AIRBORNE AND FIELD CAMPAIGN DATA PRODUCTS ARE

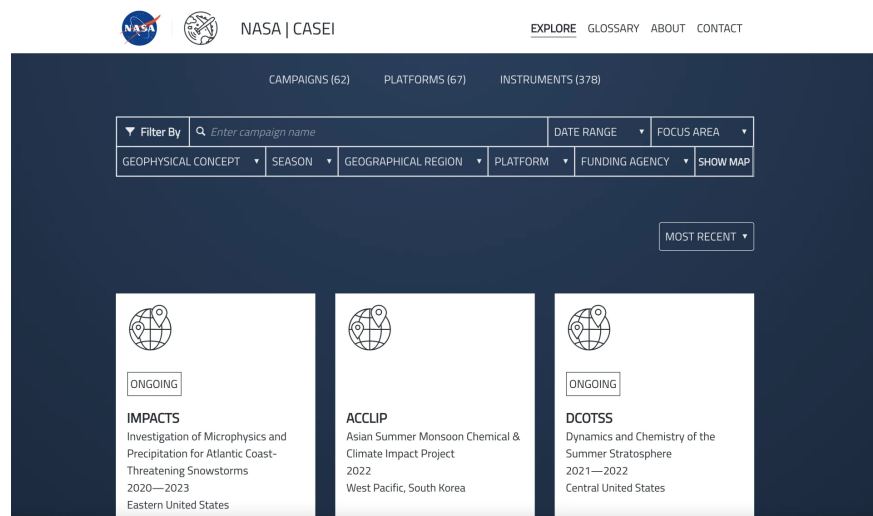
highly heterogeneous and cross-disciplinary, making it extremely challenging to manage and hard to find.

THE SOLUTION TO THE CHALLENGE IS

to ensure and improve consistency in documenting contextual information, provide a designated catalogue, and promote scientific data stewardship.

NASA IS ADDRESSING THIS CHALLENGE BY

- Establishing the NASA Airborne Data Management Group (ADMG) to curate comprehensive, consistent, high-quality contextual information that specifically pertains to the airborne and field campaign data,
- Releasing a new catalog for improved airborne and field campaign data discovery – Catalog of Archived Suborbital Earth Science Investigations (CASEI) that builds on the well-curated information, and,
- Preparing a series of stewardship best practices documents to inform DAACs on ways to effectively manage and steward NASA campaign data.



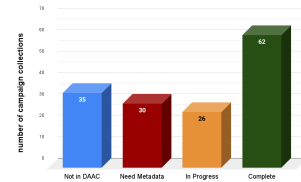
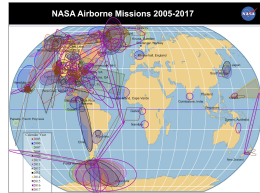
Related CASEI Resources:

- CASEI website: <https://impact.earthdata.nasa.gov/casei/about/> (<https://impact.earthdata.nasa.gov/casei/about/>)
- AGU21 CASEI Overview Abstract: <https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/957784> (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/957784>)
- AGU21 CASEI Training Abstract: <https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/965948> (<https://agu.confex.com/agu/fm21/meetingapp.cgi/Paper/965948>)

CURRENT STEWARDSHIP STATUS OF NASA AIRBORNE CAMPAIGN DATA

NASA has carried out many airborne and field campaigns, across the spectrum of the Earth ecosystem, including cloud, carbon and aerosols, and the cryosphere. Research and applications based on the campaign data are cross-disciplinary, including air quality, Monsoon processes, Arctic ice and Greenland glacier melting.

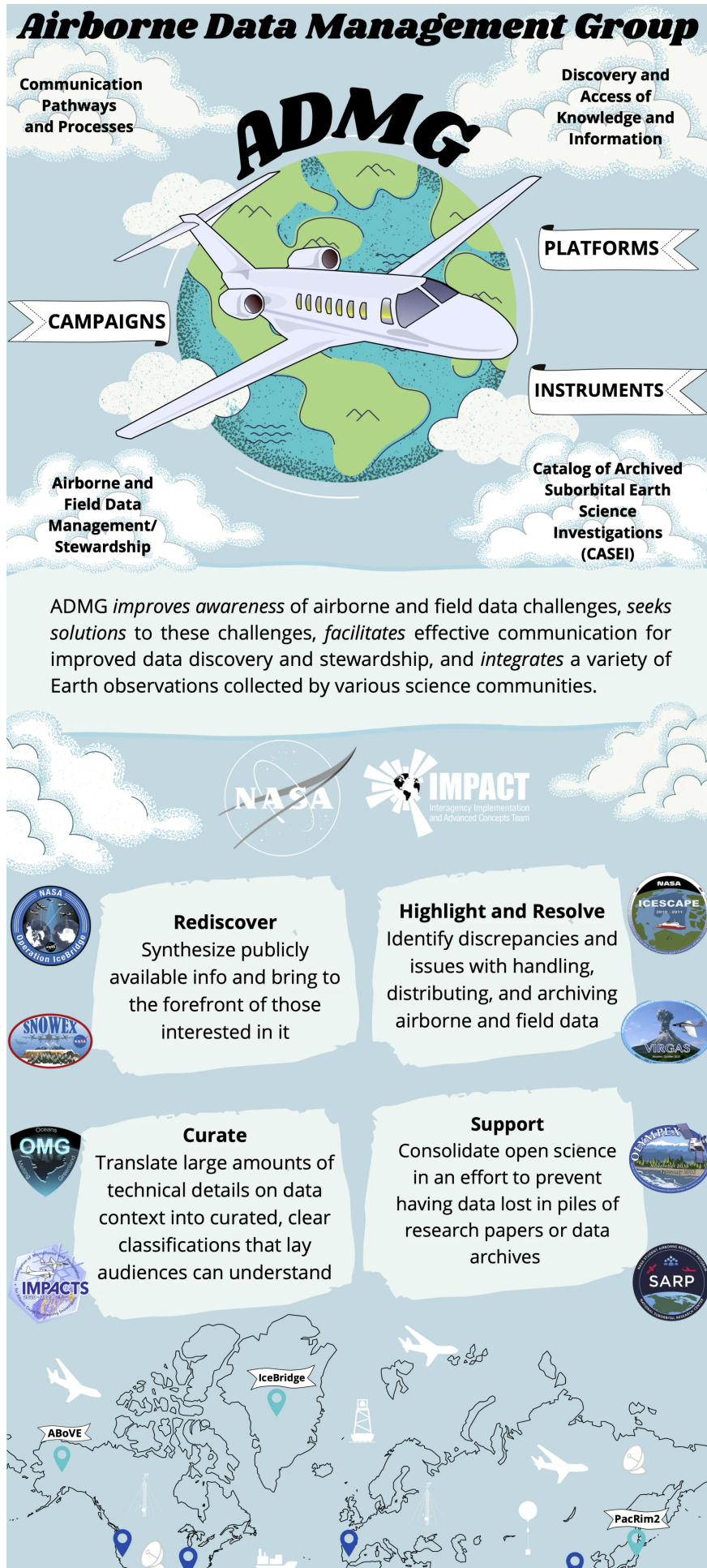
As of November 2021, there are 153 known NASA airborne and field campaigns with measurements taken from many different instruments on a variety of platforms with various spatial and temporal coverages. Platforms include aircrafts, water-based platforms, and mobile/stationary land-based platforms. Among those known campaigns, ADMG has curated and published 62 to CASEI, along with 378 commonly used instruments on 67 platforms. More are in progress and will be published to CASEI when they are ready.

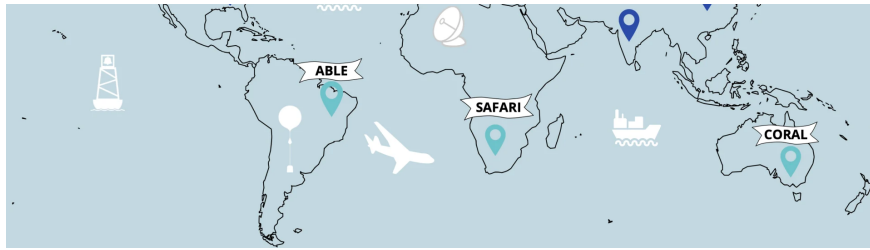


Left: NASA airborne missions from 2005 to 2017.

Source: https://airbornescience.nasa.gov/sites/default/files/documents/NASA_Airborne_Science_Mission_Map.pdf

Right: The current (November 2021) status of 153 NASA campaigns that have been identified by ADMG. Complete denotes that campaign metadata have been curated and published to CASEI. Not in DAAC denotes that the campaigns have not yet been assigned to a NASA Distributed Active Archive Centers (DAAC) for data publication. Some of these campaigns are currently being assigned to DAACs.





(Courtesy of Camille Woods and Lucia Alonso Guzman)

ADMG Website

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

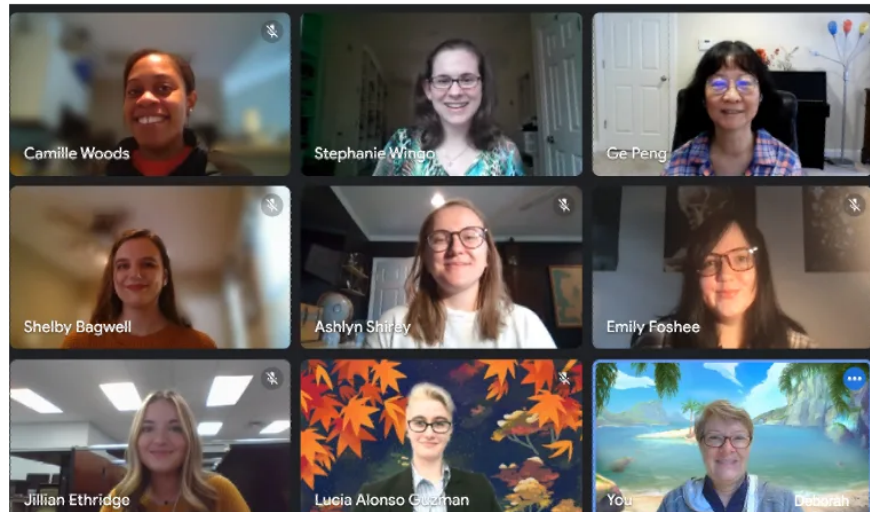
Acknowledgement

ADMG is supported by NASA through the MSFC IMPACT project under a cooperative agreement with the University of Alabama in Huntsville (NASA Grant NNM11AA01A). Feedback from NASA DAAC data managers and other personnel who work with airborne and field campaign data, ADMG and ARC (Analysis and Review of CMR) team members, was extremely beneficial.

Contact Information

Ge.Peng@uah.edu

The Current ADMG Team Members



Top row: Camille Woods, Dr. Stephanie Wingo, Dr. Ge Peng

Middle row: Shelby Bagwell, Ashlyn Shirey, Dr. Emily Foshee

Bottom row: Jillian Ethridge, Lucia Alonso Guzman, Deborah Smith

Not pictured: Dr. Danielle Groenen



AUTHOR INFORMATION

Ge Peng, PhD

Sr. Principal Research Scientist

Earth System Science Center (ESSC)/NASA MSFC IMPACT

The University of Alabama in Huntsville

Ge.Peng@uah.edu

<https://orcid.org/0000-0002-1986-9115>

Deborah Smith

Principal Research Scientist

Airborne Data Management Group Lead

ESSC/NASA MSFC IMPACT

University of Alabama in Huntsville

deborah.smith@uah.edu

<https://orcid.org/0000-0003-0312-8352> (<https://orcid.org/0000-0003-0312-8352>)

Stephanie M. Wingo, PhD

Research Scientist

ESSC/NASA MSFC IMPACT

University of Alabama in Huntsville

stephanie.m.wingo@nasa.gov

<https://orcid.org/0000-0002-1300-5682> (<http://https://orcid.org/0000-0002-1300-5682>)

Rahul Ramachandran, PhD

NASA MSFC, Huntsville, AL 35899

rahul.ramachandran@nasa.gov

<https://orcid.org/0000-0002-0647-1941> (<https://orcid.org/0000-0002-0647-1941>)

ABSTRACT

Some of the Earth system data products such as those from NASA airborne and field investigations (a.k.a. campaigns), are highly heterogeneous and cross-disciplinary, making the data extremely challenging to manage. For example, airborne and field campaign measurements tend to be sporadic over a period of time, with large gaps. Data products generated are of various processing levels and utilized for a wide range of inter- and cross-disciplinary research and applications. Data and derived products have been historically stored in a variety of domain-specific standard (and some non-standard) formats and in various locations such as NASA Distributed Active Archive Centers (DAACs), NASA airborne science facilities, field archives, or even individual scientists' computer hard drives. As a result, airborne and field campaign data products have often been managed and represented differently, making it onerous for data users to find, access, and utilize campaign data.

Some difficulties in discovering and accessing the campaign data originate from the incomplete data product and contextual metadata that may contain details relevant to the campaign (e.g. campaign acronym and instrument deployment locations), but tend to lack other significant information needed to understand conditions surrounding the data. Such details can be burdensome to locate after the conclusion of a campaign. Utilizing consistent terminology, essential for improved discovery and reuse, is also challenging due to the variety of involved disciplines.

To help address the aforementioned challenges faced by many repositories and data managers handling airborne and field data, this presentation will describe stewardship practices developed by the Airborne Data Management Group (ADMG) within the Interagency Implementation and Advanced Concepts Team (IMPACT) under the NASA's Earth Science Data systems (ESDS) Program.

REFERENCES

Peng, G., D. Smith, S.M. Wingo, and R. Ramachandran, 2021: Stewardship Best Practices for Documenting and Representing NASA Airborne and Field Campaign Data and Information. *Airborne Data Management Group*. Version: 2021-11-15.