NASA Airborne and Field Data Workshop (2022) Workshop Report and Recommendations

Workshop Report Authors: Sara Lubkin (ESDIS), Deborah Smith (ADMG), Michele Thornton (ORNL), Bruce Wilson (ORNL), Megan Buzanowicz (ASDC), Leigh Sinclair (GHRC)

Executive Summary

For over 50 years, NASA's Earth Science Division has invested in the collection of airborne and field Earth observations. However, these data have often not been given the same care and attention as NASA's satellite data. In March 2022, members of NASA's Airborne Data Management Group (ADMG), Earth Science Data and Information System (ESDIS), and the Distributed Active Archive Centers (DAACs) co-organized and hosted a two-day workshop aimed at obtaining a better understanding of concerns of both data users and data producers and at receiving comprehensive, constructive suggestions for improving the discoverability, accessibility, and (re)usability of NASA's airborne and related field data. The workshop's goal was to identify how NASA can make strategic improvements to help researchers realize the full value of all available agency airborne and field data holdings. This virtual workshop was widely announced and drew more than 100 participants. The vast amount of feedback obtained has been analyzed and discussed openly over the past year.

This report contains recommendations for NASA to significantly improve the ability for data producers and data users to meet NASA science objectives. Of the recommendations presented, the organizers suggest immediately implementing the easy wins - those that have low cost and high benefit. The two most productive next steps are already underway:

1) assign DAACs to campaigns (investigations) earlier to improve communication and collaboration, and
2) build a comprehensive Airborne and Field Data Resource Center on Earthdata using existing and new materials available from ADMG, DAAC web pages, and Earthdata.

Other important recommendations include improving the amount of contextual information available to users to enhance data discovery. This would require changes to the associated Common Metadata Repository (CMR) as well as the Earthdata Search tool.

In our rapidly changing open-source science environment, there is a great need for improved information and resources. This includes the use of modern data formats - for both new and historical airborne and field data, systems for providing both the user and data producer with data recipe notebooks and code, and better documentation ensuring consistency in information and

data stewardship across DAACs and disciplines. It is imperative to hold another virtual workshop in two years to measure progress and get the input needed to further refine NASA's work with this type of data. One finding of the initial workshop was clear - the airborne and field community is eager for more support and for opportunities to provide input.

Introduction

The NASA Earth Science Data Systems (ESDS) Airborne and Field Data Working Group (AFDWG) is an *ad hoc* collaboration among NASA's Airborne Data Management Group (ADMG), Earth Science Data and Information System (ESDIS) Project, and Distributed Active Archive Centers (DAACs). This group meets monthly to discuss and advance the stewardship of NASA's airborne and field data. As a part of this ongoing effort to improve the accessibility and (re)usability of NASA's Airborne data holdings, the AFDWG planned and hosted a two-day virtual workshop on March 29-30, 2022. The workshop's purpose was to listen to data users and data producers as they shared comprehensive, constructive suggestions for improving the discoverability, accessibility, and (re)usability of NASA's Earth Science airborne and related field data. The workshop goal was to understand how NASA can help researchers realize the full value of its airborne and field data holdings. A workshop organizing committee (consisting of a subset of the AFDWG members, see Appendix A) met weekly over four months to plan the two-day workshop. The first day of the workshop focused on the needs of the various airborne and field data user communities and the second day focused on the needs of airborne and field data producers.

Two previous articles[1], published in May and July 2022, shared preliminary workshop outcomes with a broad audience. In addition, the organizing committee hosted a session at the summer Earth Science Information Partners (ESIP) July 2022 meeting to share with the ESIP community user stories derived from the workshop feedback. The ESIP session attendees were asked to categorize the user stories with respect to cost, effort, and priority. A second ESIP session held in January 2023 focused on utilizing the ESIP community to assist with devising a solution to one of the immediate workshop recommendations: the development of an airborne and field resource center tailored to meet various user needs. Both ESIP sessions proved fruitful in providing additional feedback and guiding the workshop committee to best serve airborne and field data user needs.

---

[1] The Earth Observer: May-Jun, 2022, " Summary of the 2022 NASA Airborne and Field Data Workshop", pp 19, available at https://eospso.nasa.gov/earthobserver/may-jun-2022 also an Earthdata article: "Improving NASA Airborne and Field Campaign Data", May 18,2022, available at https://www.earthdata.nasa.gov/learn/articles/admg-workshop-summary

This white paper discusses the need for the workshop, provides participant metrics, introduces the methods and formats used for eliciting feedback from participants, summarizes workshop outcomes, and provides the prioritized AFDWG recommendations.

Need for Workshop

The AFDWG has been meeting monthly for nearly two years to bring awareness to differences in how DAACs handle field and airborne data, discuss various data use issues, and improve the users' experience with NASA's airborne and field data resources. As DAACs strive to move toward a more cohesive archive of data, the group identified a need to more fully understand community requirements, from both data user and data producer perspectives, to better serve both these communities across all DAACs handling airborne and field data. DAACs can play a greater role in support of airborne and field campaigns (also called investigations) to ensure high quality data is available for current and future use.

The following motivations for the workshop were identified:
- Airborne data has value beyond the work associated with individual campaigns and deployments, and simply archiving data at a very basic level of service is generally insufficient for capturing the enduring value of these data. Airborne and field measurements provide important, irreplaceable information about the state of the Earth at a particular place and time and at a variety of resolutions not represented by satellite data. However, when data are presented for archival years after the data collection is finished, a very basic level of service may be the only practical option.
- Multiple groups, including the 2016 Satellite Needs Working Group, have requested NASA make it easier for anyone to find, access, and use airborne and field data. Gaining a clearer understanding of the use cases for this data is necessary to ensure this is done in an efficient and effective manner.
- There is a need to extend the use of NASA's airborne and field data assets, but this work is expensive. It is therefore important to identify the needs that are most critical to the broadest range of users in order to prioritize the most important elements of this work.
- Airborne and field data are inherently different from satellite data. Unlike typical satellite data collections, data from a single airborne and field campaign can include a wide variety of data formats, data processing levels, and many other heterogeneities. This variety, as well as the large number of coincident and related data products associated with any particular study, increases the complexity of data holdings. Airborne and field measurements focus on meeting specific scientific objectives at the various resolutions needed for scientific understanding. There can also be long periods of data collection that are less directly relevant to the specific science objectives of the campaign, but which may have value in other scientific contexts. The differences in science objectives can result in different methodologies used for data collection, and all differences and details are critical aspects of the campaign documentation. Unfortunately, documentation

processes may be inconsistent across DAACs, as well as across time within a given DAAC.

- Each DAAC handles airborne data in different ways based on their understanding of community needs and the data management practices appropriate to the types of data they steward, but subsequent airborne and field data users and producers often need to use data stewarded by multiple DAACs. Because these differences create barriers to data usability and can lead to data user frustration, it is essential for DAACs and EOSDIS to work together to understand the reasons for these differences and find ways to eliminate those that create unnecessary barriers to use.

As discussed in the next section, the organizing committee designed a workshop to solicit the information needed to address these concerns.

Workshop Description

This workshop had over 250 registrants: over 115 attendees on Day 1 and over 70 attendees on Day 2. Attendees included data users, data producers, those who consider themselves both data users and producers, as well as personnel from various DAACs, NASA Headquarters, ADMG, and ESDIS (Appendix C). Attendees represented a wide range of scientific fields, including atmospheric science, terrestrial ecology, physical oceanography, ocean biology, and cryospheric science. Roughly 40% of attendees identified as students or early career researchers and 35% identified as mid-career. The attendee demographics clearly show that the workshop drew the attention of the intended target audience.

Workshop Overview: Day 1

The workshop began with welcomes from Kevin Murphy [NASA Science Mission Directorate Chief Science Data Officer] and Melissa Martin [NASA Deputy Airborne Science Program Director/Earth System Science Pathfinder Program Office (ESSPPO) Earth Venture Suborbital (EVS) Mission Manager] to thank all attendees for joining and to thank the speakers for offering their insight to the conversations.

Session 1: The Airborne Data User Community Shared their Experience
The first day of the workshop focused on the needs and concerns of data users. Data users were given an opportunity to share their experiences using NASA airborne and field data and to provide feedback about what works well and what needs to be, or can be, improved. To set a solid foundation of understanding for all attendees, ADMG Lead Deborah Smith introduced the purpose and activities of ADMG, and ESDIS DAAC Operations Engineer Frank Lindsey provided an overview of ESDIS.

The first session highlighted four invited speakers who were asked to share their experiences using NASA airborne and field data. Speakers from NASA missions with significant airborne components were chosen who represented the broad range of airborne instrumentation, measurements, and data. Speakers introduced themselves and described how they use NASA airborne and field data in their scientific research. Each speaker answered a series of previously provided questions aimed at identifying what works well for data users and what improvements might facilitate discovery, access, and use of NASA's airborne and field data. Questions included:

- From your perspective, what does it mean to you to be a "User of Airborne and Field Data"?
- What airborne and field data are you using?
- What does a generalized access-to-data workflow look like for you, *and* what are the challenges of data access within that workflow?
- What are your (or students'/colleagues') data challenges? Discoverability? Availability? Access? Data Format(s)? Standardization?
- With respect to airborne and field data, what is working for you that you would want to share with your research community?
- What are the challenges of Earthdata Search/DAAC data discoverability?

The first invited speaker, Phil Townsend [University of Wisconsin Environmental Spectroscopy Laboratory] uses Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG) data in combination with data from many different sources including field and lab spectra and chemical component analysis to characterize fine-scale vegetative traits and function at the continental scale. He provided valuable insight into using large, complicated datasets from cloud sources. Townsend was followed by Qing Liang [NASA Goddard Space Flight Center (GSFC), an atmospheric modeler who combines NASA, National Oceanic and Atmospheric Administration (NOAA), and National Science Foundation (NSF) data in 3-D chemistry and climate models. The third presenter was Mark Tschudi [University of Colorado, Boulder]; he works with Operation IceBridge datasets for both calibration/validation and research. Timothy Lang [NASA Marshall Space Flight Center (MSFC)] wrapped up the session with a description of his use of campaign data for statistical validation, precipitation and microphysics analysis, and case studies. Lang also provided input from his experience as a data producer working with multiple DAACs. Each speaker described what currently works well, shared their pain points, and made suggestions for improvement.

Uses of airborne data for applied science were highlighted by Kenton Ross [NASA Langley Research Center (LaRC)], chief scientist of NASA's DEVELOP Program who described several 10-week feasibility projects conducted by DEVELOP participants. In these projects, DEVELOP

participants used airborne data to detect invasive species, monitor ecosystem health, identify permafrost subsidence, and locate methane emissions.

The DEVELOP session was followed by an invited Airborne Data User Panel comprised of Elizabeth Hoy [NASA GSFC], Sean Serbin [Brookhaven National Lab], Stephen (Joe) Munchak [Tomorrow.io], Owen Cooper [University of Colorado, Boulder], K. Fred Huemmrich [University of Maryland, Baltimore County (UMBC)], Alexey Shiklomanov [NASA GSFC], Tim Bailey [The Watershed Center], and Rebecca Hornbrook [National Center for Atmospheric Research (NCAR)]. Each discussed how they access and use NASA data and the challenges they experience while doing their work.

Session 2: NASA Tools and Services for Airborne Data Users

The second half of Day 1 focused on NASA's existing and planned tools and services for airborne and related field data. Bruce Wilson [Oak Ridge National Laboratory (ORNL) DAAC Manager] led off this session with a presentation on the Earthdata Cloud Evolution. NASA has recently migrated the most heavily used Earth Observing System Data and Information System (EOSDIS) data into Earthdata Cloud. Wilson discussed NASA's use of tools to enhance access to NASA Earth science data in the commercial cloud. These efforts are focused on creating capabilities for users to have direct access to data, improving the efficiency of data system operations, and enabling tools that work across DAACs. Workflows will continue to exist that provide download and use of data locally, but users will also gain the ability to access and analyze data without downloading and the ability to use tools like Harmony that work with data from multiple DAACs. While relatively little airborne data had been migrated at the time of the workshop, the migration efforts continue and are in progress for several airborne instruments, notably AVIRIS and MASTER.

At the end of Day 1, Deborah Smith [ADMG] facilitated a series of talks providing overviews of various NASA tools for airborne and field data discovery and use. Each tool presented is summarized in Table 1.

Session 2 concluded with an interactive activity where workshop participants highlighted their data discovery and data use needs on a Jamboard (Appendix D).

Table 1: *List of relevant airborne-and-field-data tools presented at the Airborne and Field Data Workshop, March 2022*

*Please note: with the ongoing web unification, links below may be in transition and/or no longer be active when reading this report.

Tool	URL	Description
Catalog of Suborbital Earth science Investigations (CASEI)	https://impact.earthdata.nasa.gov/casei/	Explore an inventory of NASA's airborne and field campaigns for Earth Science in this tool built by ADMG
Sub-Orbital Order Tool (SOOT)	https://asdc.larc.nasa.gov/soot/	Discover and access selected airborne and field campaign data archived at the ASDC
Soil Moisture Visualizer (SMV)	https://airmoss.ornl.gov/visualize/	Visualize integrated soil moisture data from AirMOSS flights, SMAP, and selected ground sensors at the ORNL DAAC
Airborne Data Visualizer (ADV)	https://daac.ornl.gov/tools/airborne-data-visualizer-project-list/	Visualize flight paths and atmospheric composition data from ACT-America, ATom, and CARVE missions at the ORNL DAAC
Field Campaign Explorer (FCX)	https://ghrc.earthdata.nasa.gov/fcx/index.html	Discover and access selected airborne lightning data archived at the GHRC
Spatial Data Access Tool (SDAT)	https://webmap.ornl.gov/ogc	Visualize and subset gridded (GeoTIFF) data from missions held at the ORNL DAAC, including CARVE and AfriSAR.

Workshop Overview: Day 2:

Day 2 of the workshop focused on the requirements and pain points of airborne and field data producers. With increasing demands on NASA primary investigators (PIs) and little subsequent increase in funding, there is a need for DAACs to provide greater support to airborne and field campaigns than has historically occurred for these types of data. Participants started the day in randomly assigned breakout rooms where data producers discussed Day 1 activities and identified lessons learned from data user feedback. The purpose of these discussions was to

gauge if the appropriate "data user" topics were covered, identify the knowledge or understanding gained from Day 1 in comparison to the start of the workshop, and determine which topic areas within Day 1 were most applicable or meaningful to the data producer workshop participants. Discussion notes highlighted data producer perspectives about the previous day.

Session 3: The Airborne Data Producer Community Shares their Experience

This session offered an opportunity for data producers to share their perspectives. Deborah Smith [ADMG] moderated a series of flash talks given by NASA-funded scientists involved in airborne and field campaigns that contribute significant airborne and field data (Table 2). Workshop organizers provided a series of questions to guide the Session 3 flash talks which allowed data managers and producers to share their experiences. Guided questions included:

- Which datasets have you produced?
- What DAAC have you worked with?
- What worked well?
- What pain points did you find when working with the DAAC?
- What suggestions do you have for improvement?
- Have you tried to use data in the cloud?
- What do you think is needed to encourage and support future use of your data product(s)?
- What support would be needed from ADMG?

Table 2: Session 3 Flash Talk Speakers

Panelist/Scientist	Mission	Description
Stacy Brodzik, Software Engineer, University of Washington	IMPACTS	Investigation of Microphysics and Precipitation for Atlantic Coast-Threatening Snowstorms (IMPACTS)
Emma Yates, Research Scientist, NASA Ames Research Center	AJAX	Alpha Jet Atmospheric eXperiment (AJAX)
Daniel Jensen, Postdoc, NASA	Delta-X (AVIRIS-NG)	Delta-X / Airborne Visible InfraRed Imaging Spectrometer - Next Generation (AVIRIS-NG)
Fred Bingham, Physical Oceanographer, University of North Carolina at Wilmington	SPURS-1, 2, S-MODE, SASSIE	Sub-Mesoscale Ocean Dynamics Experiment (S-MODE)

| *Gao Chen*, Physical Scientist, NASA Langley Research Center	ACT-America (and 15+ NASA airborne field study deployments)	Atmospheric Carbon and Transport America (ACT-America)

Directly following the flash talks, Kasey Phillips [Atmospheric Science Data Center (ASDC)] wrapped up the session with a moderated fishbowl panel where all the flash-talk speakers responded to questions/comments from workshop participants.

Session 4: Working Together to Make Data Accessible

The final session of the workshop emphasized how the broad airborne and field community can work together to improve data accessibility, including accessing data in the NASA Earthdata Cloud environment.

Presentations by Amanda Leon [National Snow and Ice Data Center (NSIDC) DAAC Manager], Deborah Smith [ADMG], and Bruce Wilson [ORNL DAAC] focused on the project life cycle, ADMG, and DAAC processes and responsibilities, and enhancing communication among and between stakeholders. The presentations emphasized common themes:

- Working together through the *entire* project lifecycle: Project Initiation Phase, Planning and Data Management Plan (DMP) Preparation Phase, Active Data Collection/ Delivery Phase, and Closeout Phase;
- Promoting two-way communication (continuous and clear discussions regarding data and open science requirements). Science teams and DAACs need to work collaboratively throughout the entirety of the campaign;
- Formalizing the roles and responsibilities of the science team, ADMG, and DAAC (Figure 1);
- Utilizing the potential capabilities offered by a cloud environment (enabling user access to large volume data, removing barriers to cross-DAAC tools and data access, enabling analysis, etc.);
- Making NASA Earth Science data as FAIR as practical given evolving DAAC responsibilities (data publication, data access, and user support).
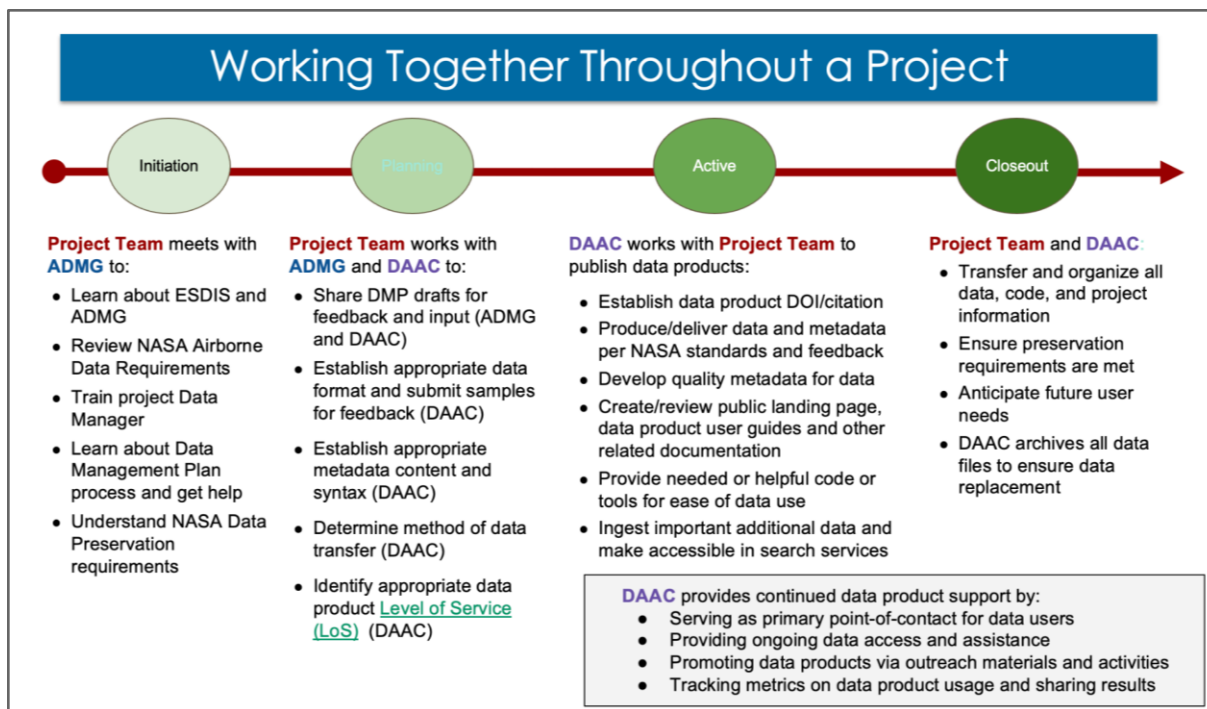
Figure 1. An outline of steps, roles, and responsibilities of the science project team, ADMG, and DAACs for communicating with each other throughout the life cycle of a project. Credit: Amanda Leon and Deborah Smith

Amanda Leon provided information regarding Earthdata Pub, a data publication workflow intended to enhance interactions between the DAAC and the data producer during the data publication process and to provide consistency across DAACs. Key discussion points included the use of Earthdata Pub to provide a data publication workflow in Earthdata Cloud with a common terminology and user interface to act as a centralized location for resources for data producers to help support the data publication process (i.e., enabling data producers to find instructions on how to publish, communicate with DAAC personnel, how to provide data and related information to DAAC, etc.).

Participants revisited the topic of collaboration in moderated breakout sessions that provided an opportunity for small-group discussions. Breakout sessions focused on the following topics:

- Data Formats/ Data Standards (commonly used data formats, pros and cons of specific formats, etc.)
- Earthdata Pub (transitioning to Earthdata Pub, the need to train DAAC staff for using Earthdata Pub, etc.)
- Airborne in the Cloud (primarily regarding datasets in the cloud, hybrid approach discussion, etc.)
- Metadata (including discussions regarding CF Conventions and GCMD keywords)

- Working with the DAACs (highlighting the importance of open lines of communication between the Principal Investigators and the DAACS to help streamline the data publication process and understand nuances within a project)

Lastly, Joe Koch [ASDC] and Siri Jodha Khalsa [NSIDC] from the EOSDIS Standards Coordination Office (ESCO) gave a presentation regarding the role of ESCO, the critical nature of standards throughout the entirety of the data product lifecycle, community standards (data formats and general recommendations), and the process of submitting a document for review through the ESCO process. The meeting wrapped up with an overview of "Open-Source Science at NASA" presented by Elena Steponaitis [NASA Headquarters Earth Science Data Systems (ESDS) Program]. This presentation included a discussion on open science and open-source science, as well as related principles and best practices. The Scientific Information Policy (SPD-41[2]) and the [Transform to Open Science](#) (TOPS) initiative/ objectives were highlighted as current and future efforts.

Workshop Findings

Over the two-day workshop, more than 100 attendees shared their experiences using, producing, and stewarding NASA airborne and field data. Participants of varied backgrounds and experience levels (Appendix C) suggested improvements through community members' presentations, breakout discussions, chat comments, panel discussions, Jamboard comments, and Slido polls (see examples in Appendix D). The variety of communication methods used during this fully virtual workshop helped ensure that everyone attending had a chance to participate and have their ideas heard.

Day 1 feedback included common themes. For example, the data user panelists in the first session all agreed that discovering data can be a problem. They suggested that keyword searches with greater flexibility than is currently available in Earthdata Search would be helpful, as well as the ability to search for data by location or by attribute. In addition, the data-user panelists suggested that additional tools for airborne data are needed. Available tools are often dataset specific and thus are not useful for researchers who want to utilize data from several sources. *All* the advanced users that presented detailed descriptions of how they utilize airborne and field data in their work mentioned issues with accessing and using large data volumes, and *all* agreed that more mature (e.g., higher level) data products are needed for non-experts.

During the afternoon session of Day 1, a key challenge was recognized - data organization and presentation vary substantially across projects, based on when data were collected, the project that collected the data, the length of time between collection and archival, and the DAAC that archived the data (if the data was even archived at a DAAC). Participants stated that they find it

---

[2] Since the workshop in March 2022, the SPD-41 Open Science Policy has been updated and replaced with [SPD-41a](#)

difficult to locate airborne data relevant to a particular location and time, as well as to align airborne data (in space and time) with the related and concurrent ground and satellite measurements. Except for CASEI, the tools available for airborne and field data are limited in scope to a subset of the data held at the DAAC that developed the tool. There is a clear need for tools that work across DAACs and for tools that enable the integration of ground, airborne, satellite, and modeling data. Participants also expressed a desire for better visualization of flight tracks; including the ability to see where tracks overlap (particularly where flight tracks overlap with satellite instrument swaths) and the ability to zoom in to obtain data. There was also some discussion of file formats, recognizing that a given format may be common for a portion of airborne users and an impediment to use by others, and that some formats are more amenable than others to the analysis-in-place supported by Earthdata Cloud.

Day 2 feedback focused more on communication issues experienced by data producers. Overall, the data managers and producers reported positive and responsive interactions when working with NASA's DAACs. Meaningful suggestions for areas of improvement in DAAC-scientist collaborations were presented. Several speakers highlighted the importance of making communication a top priority between the NASA science team members (e.g., the data producers) and the DAACs, as good communication leads to a better understanding of standardization in terms of file formats, naming conventions, metadata, and the evaluation of merged or higher-level products. Furthermore, a strong communication cycle has already provided the opportunity for the development of web-based, DAAC-hosted, dataset catalogs and other enhanced discovery methods.

Data producers also commented that there is often a Field Data Repository (or archive) run by a project data manager used as intermediate data storage for sharing and accessing data within the project science team before the DAACs publicly release the publication-quality data and metadata. Sharing these intermediate repositories with the DAACs facilitates open and effective communication between the science team and DAAC. Such early communication ensures data product quality, supports research activities both during missions and after data publication, and from the perspective of the data providers, improves the long-term curation in the form of DOIs, collaborative user guides, and NASA Earthdata outreach.

Finally, it was recognized that many datasets are not widely supported with standard software, code, or tools. A suggestion was that the DAACs and/or the ADMG should find a way to support code and tools that data producers develop while conducting their primary research activities to make archived datasets more accessible and usable by the broader research community.

Participants in the final Day 2 breakout rooms identified several challenges that hinder airborne and field data access and (re)use, especially for legacy data. These challenges include:
- Difficulty in finding historical airborne and field data;
- Delays in the assignment of a DAAC for airborne projects limiting public data access;

- The proliferation of heterogeneous, domain-specific data file standards and formats makes it difficult to use the data for multi- and cross-disciplinary applications;
- Lack of standard data download protocols and tools, with different protocols requiring different scripts for downloads at different data sources;
- Need for holistic contextual information from the start of a campaign, regardless of what formats of metadata and data files are used, to ensure appropriate reuse of data.

The organizing committee spent more than four months sifting through the entirety of all workshop suggestions and feedback, removing duplicates, sorting the feedback into twelve categories, and summarizing what was learned. Similar comments were combined keeping count of comment frequency. The categories with the most comments were: Tools, Communication, Consistency, and Data Access. From this effort, the following generalized findings were identified:

- *Communication among all stakeholders is important.*
  - Participants acknowledged the high quality of communication between DAACs and science teams. They emphasized the importance of early DAAC involvement with the campaign science team. Both the DAACs and the science teams noted that early DAAC assignment results in faster publication times, lower effort associated with data management and data publication (and therefore lower costs), and greater consistency of data products within and among projects.
  - Instrument and field teams would like additional assistance from DAACs concerning data management plan (DMP) development and metadata and data format checking, as science teams have tight budgets for science, and data management is often new to them.
  - The community would like information about open-source science and SPD-41a requirements and would appreciate guidance on the requirements for implementation within airborne and field projects.
- *More cross-DAAC consistency in documenting, formatting, and presenting data is needed.*
  - Data users would like to see consistency in terms, metadata, data formats, filenames, and data variable names.
  - Recognizing disparate disciplines have different needs and standards, participants cautioned data centers to consider those needs and standards and not attempt one-size-fits-all approaches.
  - There is a key need for an assessment of the differences in discipline practices determining which are important and necessary for serving those specific disciplines vs legacy practices.
- *Both data producers and data users need more information and resources, including:*
  - Campaign, instrument, and variable "landing pages" that summarize the data that is available and provide links to all related documents in one place;

- More mature data products that provide scientifically reviewed complex modeled data, improved clarity, easier-to-use formats, and better documentation;
- Make camera or video environmental information available to aid in interpreting data collected during the flight;
- Lower barriers to data access and use for non-experts and those new to airborne and field data - this is especially important for interdisciplinary scientists, modelers, and students, as well as others who may not be familiar with the data;
- Provide services to assist with accessing and analyzing cloud data as well as guidance documents to identify and explain available tools and services; simply putting data in the cloud is not enough.
- *Data Producers, data users, and DAACs have suggestions for improving data discovery, access, and use.* Some suggestions include:
  - Simplified user interfaces;
  - Data and documentation consistency;
  - Improved data and documentation access;
  - Improved bulk data access;
  - Improved access to legacy data;
  - Unique data product filtering;
    - Data search by variable, frequency, altitude, or flight,
    - Data search by time/space.
    - Data search by measured variable, which is particularly important for the atmospheric composition community, and
    - Ability to find coincident data: where airborne measurements were made over coincident ground measurements and where airborne flights are underneath coincident satellite measurements.
  - Improvements in data versioning and updates, as well as enabling notifications to users when data are updated or replaced;
  - Improvements in the ability to identify specific deployments and flight lines associated with specific projects or campaigns, particularly for facility instruments;
  - Ability to normalize data collected across space and time.
  - Data integration with other satellites and with model data;
  - Co-location of data rather than disbursement across all DAACs;
  - Tools that allow conversion of data to different formats (e.g., ICARTT to NetCDF).

The committee used all findings to construct a set of 60 user stories (Appendix E). Some of these user stories were used in an interactive July 2022 Summer ESIP session in which the participants placed each user story in one of four categories: Must Have, Need, Want, or Not Important. The user stories ranked by attendees as a "Must Have" were then further discussed in a follow-on

activity to determine the potential level of effort (easy, medium, or hard) required to address the need. This activity helped to identify which user stories might have high-impact/low-effort solutions.

After the ESIP session, the committee continued to meet weekly and discussed all user stories (not just those examined in the ESIP session). They performed a separate assessment like that carried out in the ESIP session and brainstormed potential solutions. The committee extensively discussed the level of effort and priority for each user story and potential solutions (Appendix E) to determine if a solution was already in place, was in progress, existed in part, or did not exist at all, depending on its current status. The result of this work is the set of recommendations below.

Committee Recommendations

The two-day workshop provided significant amounts of feedback to guide future efforts to advance the stewardship of NASA's airborne and field data. As we follow up with additional workshops, on a recommended two-year cycle, we will ensure the airborne and field data communities continue to have opportunities to provide important feedback to the DAACs and ADMG. Future workshops will continue to be virtual ensuring the greatest participation at a lower cost. Future workshops may also focus on specific topics of interest, such as applications, heritage data reuse, or cloud analysis, in addition to providing opportunities for training specific to airborne and field data.

After careful consideration of all feedback obtained during the March 2022 workshop, the organizing committee has developed the following list of the most important recommendations. This list is prioritized with the most immediate recommendations at the top.

1. <u>Assign DAACs to campaigns as soon as possible after the award</u>: Early DAAC assignment is essential for all NASA projects that collect or generate data - and critical for the success of airborne and field projects. NASA's Research and Analysis (R&A) program scientists should notify ESDS when campaign solicitations are issued so DAAC selection can occur as early as possible. This low-cost, low-effort change will improve DAAC/data producer communication by initiating the partnership between the DAAC and the science team early in the data lifecycle. This has already been successfully implemented for the Earth Venture Suborbital program's third round of funding (EVS-3). By bringing in the DAAC early, the DAAC and science team were able to work out data details, especially regarding standardization, that substantially reduced the level of effort for publication and ensured interoperability and access to tools. To function effectively as a project partner; however, DAACs must participate in science team meetings and open science workshops. This ensures DAACs gain the detailed knowledge needed to create various outreach materials (e.g., micro articles, story maps, and data recipes), ensure information is captured and shared on the DAAC website, and provide the most accurate

and complete metadata and user guides for data products. These efforts help data users find and use data and lower barriers to entry for interdisciplinary and applied science data users.

2. Develop an Airborne and Field Data Resource Center (AFRDC) on Earthdata: Construction of an airborne and field resource center on Earthdata with links to resources at ADMG and the DAACs (Appendix B), as well as information and tutorials for the airborne and field communities, will help data users and data producers by providing all needed resources in one location. Content would include a variety of airborne pathfinders and a library of Jupyter notebooks demonstrating data access and usage. The AFRDC would also include standards guidance and templates for data producers. Since the speed of response to the identified need is important, building this resource center collectively and ensuring it is highly linked to all available resources at ADMG and the DAACs is imperative.

3. Improve DAAC Campaign, Instrument, Flight, and Event Information: All DAACs should include flight catalogs or campaign catalogs with detailed information about what platform was flown, where, when, and why and with what instruments and where the instruments were placed on the platform. Providing these types of details at DAACs ensures users have the information necessary for appropriate future use and interlinks users to related resources at other locations or DAACs. The catalogs should be accessible from the new AFDRC or through CASEI to provide centralized access (not DAAC-specific access)

4. Make modifications to CMR: Users have difficulty finding NASA airborne and field data in Earthdata search. This is partly due to the limited metadata required for CMR. Without additional metadata, the search for airborne and field data is often ineffective in meeting user needs. To make airborne and field data more searchable by type of observation, campaign, platform, instrument, geographic region, and multiple variables via Earthdata Search, new metadata fields for airborne-type data would need to be added to CMR and required for input by DAACs. Adding access to CASEI or flight catalogs would enhance searchability for data users.

5. Improve search capabilities and ensure all airborne and field data can be found in Earthdata: This effort requires metadata improvements to data already in the CMR, including additional metadata beyond what CMR presently supports to enable greater interoperability. Completing the archival of legacy data not yet in CMR will allow all data to be searchable rather than hidden from users. CASEI has been built to provide users with additional metadata and, therefore, enhanced searchability. CASEI content should either be absorbed into Earthdata or permanently provided to users as a free-standing tool.

6. <u>Improve usability and interoperability of disparate data formats</u>: This involves the conversion of legacy data into new formats, developing and maintaining services to return data in common formats, and increasing the use of formats that are cloud optimized. These steps will improve both data usability and data user satisfaction.

7. <u>Address inconsistent DAAC user experiences for both data users and data producers</u>: Greater cross-DAAC consistency or DAAC-agnostic experiences are needed. This includes consistency in processes, metadata, formats, communications, actions, and interface look and function. Consistency in tools such as Earthdata Pub for submitting data to a DAAC and the Algorithm Publication Tool for authoring Algorithm Theoretical Basis Documents (ATBDs) have already been developed and are beginning to be put into operation. However, there remain issues with users accessing data from non-DAAC, and sometimes non-NASA, websites. Early DAAC assignment and working through DAACs to provide early public access is essential in furthering both open-source science and the effective ongoing reusability of these extremely valuable airborne and field research data. Providing an alternative to research group or instrument websites ensures that users obtain appropriate data and pertinent contextual information for responsible research use and publication and reduce confusion that occurs when multiple dataset versions are made public. The user experience should also be consistent regardless of DAAC. However, it must still be recognized that disciplines may require unique tools and services and that DAACs are experts in serving their discipline communities. While there are many reasons to utilize the valuable information and support within the topical domains of the DAACs, bringing consistency to user guides and landing pages across DAACs will reduce user frustration and time.

8. <u>Expand and improve an FCX-type tool</u>: Some tool capabilities already exist that allow users to visualize airborne and field data. However, existing tools, such as the Field Campaign eXplorer (FCX), have varying capabilities, are DAAC-specific, and provide visualization for only a few campaigns. Data users have indicated a need to see flight tracks, satellite swaths, instrument data, flight reports, video recordings, etc. A tool that works for all campaign flights, that is available from a centralized location, and that supports working with data from any DAAC would provide these services. However, this can be a difficult delivery given the current differences among DAACs and among data products.

9. <u>Address large airborne and field file needs</u>: While campaign data files are often small, there is a need to address user issues by using larger data files such as radar (both airborne and ground-based radar), lidar, or hyperspectral data products. The AFDWG recommends two approaches to help ease user concerns:

    a. <u>Provide the ability to subset large data</u>: Subsetting based on location, date, altitude, instrument, platform, or other means would improve data access by allowing users

to download only the data they need. Identification of the most valuable subset capabilities and implementation of cost-effective changes allowing the additional subsetting will improve data user satisfaction.

b. Provide cloud-optimized data products: Major airborne data products, such as facility instrument data, should be made available in cloud-optimized formats (especially important for hyperspectral and other large datasets that are unwieldy when downloaded). For more information about airborne and field data in the cloud, see Appendix F.

10. Provide the ability to search for multiple variables in the same data product: Many workshop participants identified the need to search for data products by variables. For example, a user may want data products with both ozone and carbon monoxide. Depending on the needs of the user and their discipline, the following types of searches may be needed:

- In situ air samples (i.e., targeted compounds)
- Air column (i.e., total column or specific altitude minus surface noise)
- Remote sensing of surface - (location, bands, altitude)
- Remote sensing of the atmosphere (e.g., radar variables)
- Ocean surface
- Ocean subsurface

11. Identify and resolve intercloud data transfer issues: Larger Earth Venture Suborbital projects often have field archives that are set up by project-specific data managers for use by the research team and collaborators. Currently, there are no requirements for these field archives. Some field archives are located in a different cloud service than the AWS cloud used by NASA DAACs. This requires extra effort to move the campaign data products and information from one cloud environment to another. More information about these types of issues is needed so that solutions can be developed to reduce effort, potential errors from data transfer, and data producer frustration.

12. Increase compatibility and coordination across other US agencies: There is a need to improve information about all airborne and field data-generating activities including those funded by other organizations such as NSF National Ecological Observatory Network (NEON)), NOAA (Earth Observing Laboratory (EOL)), Department of Energy (Atmospheric Radiation Measurement (ARM)), Department of Interior (U.S. Forest Service, U.S. Department of Agriculture) and other U.S. and international organizations that also gather airborne and field data. While many NASA-funded campaigns coordinate with non-NASA co-campaigns or collaborate in multi-organizational investigations, issues identified with finding airborne and field data are not specific to NASA. Users also find it difficult to locate information about other agencies' efforts. This is a difficult problem with no clear solutions, although participation in outreach activities, standards

groups, and cross-agency collaborations can help. Higher level approval is needed for ADMG or another entity to further collaborate with other organizations.

Conclusion

The committee unanimously agrees that the two-day airborne and field workshop was a highly valuable effort that increased the awareness of airborne and field data issues for all involved. The workshop findings will continue to guide the ADMG, ESDIS, and the DAACs in the coming years as each entity works towards improving activities and efforts in order to better support data users and data producers.

The whole workshop process, from conception to final report, is a good example of open science and community engagement. At every opportunity, community input was sought and information about workshop outcomes was openly provided. The workshop web page on Earthdata continues to be a resource for data producers, data users, DAACs, and NASA management. The page contains access to the workshop recordings and information and will be the primary source of access to this report.

Of the recommendations presented in this final report, the AFDWG suggests immediately implementing the easy wins - those that have a low cost but high benefit. The two most productive next steps have already been started:

- Assign DAACs earlier to campaigns: Early DAAC assignment is the number one recommendation from the workshop. It was implemented for the third round of Earth Venture Suborbital funding, and it has proven successful, especially regarding development of more complete DMPs. AFDWG recommends expanding the practice to other funded campaigns. The AFDWG will work to educate R&A program scientists about the need for early assignment so that assignment occurs within three months of project funding for optimal outcomes. One of the biggest advantages of early DAAC assignment is the collaboration and strong relationship built between the DAAC and the science team resulting in a more complete DMP, more efficient data publication, and quicker data access for the public. Funding for DAAC participation in science team activities is the primary cost of this recommendation.

- Build the Airborne and Field Data Resource Center (AFDRC) website: The AFDRC website will be a collaborative effort between ADMG/DAACs/ESDIS to revise and improve airborne and ADMG pages on Earthdata, thereby ensuring data users have *a single comprehensive location* for NASA airborne and field resources. Some content is already available on the ADMG Earthdata pages and individual DAAC websites. In January 2023, the AFDWG hosted a Winter ESIP session where feedback and suggestions were obtained from session attendees. The AFDWG is now utilizing its regular monthly meetings to work towards AFDRC development. The primary cost of this recommendation is contributor time and Earthdata web content team efforts. Given

the upcoming web unification project and anticipated content freeze, the AFDWG will have time to develop the resource before construction can begin.

An additional concern heard frequently in the workshop, as well as individually from data users, is the need for data subscription and notification services. Several DAACs already have these services. This need has not been added as a recommendation in the lists above since it is already implemented. Currently, Earthdata Search has file (granule) and data product (collection) notification services; however, no user documentation exists. As a result, few people know about these services or how to use them. We recommend further outreach and education to guide users to subscription and notification resources, and engagement with usability experts to assess how to broaden the reach of these resources. In addition, information should be posted on the Earthdata Forum providing detailed assistance.

Lastly, the ADMG was chartered in 2018 in response to a recognizable need for improved airborne and field data stewardship, and it has become an important asset to the DAACs over the past four years, particularly with regard to science team communications. If the DAAC has communication issues with a scientist or science team (e.g., the science team is unresponsive to emails), ADMG can step in to ensure that necessary conversations occur. The ADMG link to NASA HQ is a crucial part of that effort by providing valuable insight into what is occurring at HQ regarding airborne and field data activities and plans and providing collective DAAC communication to HQ. ADMG has also played an important role in streamlining the data ingest, archive, and distribution process to help both the DAACs and science teams make the data publicly available at a faster rate. The AFDWG recommends that the ADMG continue functioning as a facilitator and communicator to enhance cooperation and collaboration between DAACs, project science teams, airborne science programs, and upper management.

Appendices

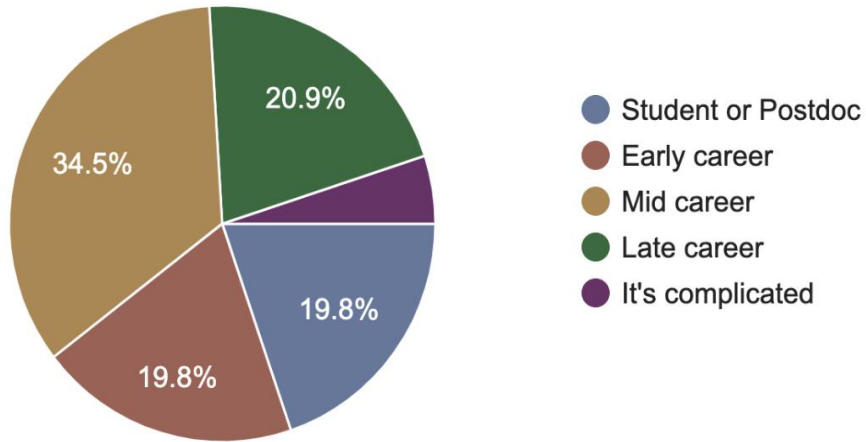## Appendix A: Organizing committee members

Sara Lubkin (ESDIS)
Deborah Smith (ADMG)
Michele Thornton (ORNL)
Bruce Wilson (ORNL)
Megan Buzanowicz (ASDC)
Kasey Phillips (formerly ASDC)
Leigh Sinclair (GHRC)
Amanda Leon (NSIDC)
Matt Tisdale (ASDC)
Geoffrey Stano (GHRC)
Jeanne Behnke (now retired)
Drew Kittel (ESDIS)
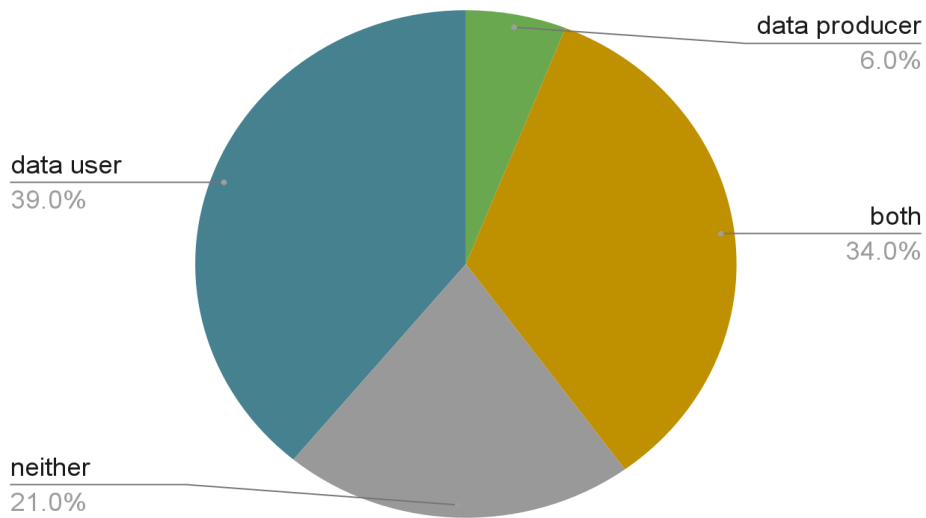Sean Bailey (OB.DAAC)

Appendix B: Resources

- Workshop web page: (contains workshop recordings, workshop report) https://earthdata.nasa.gov/esds/impact/admg/nasa-airborne-and-field-data-workshop
- ADMG web page: https://earthdata.nasa.gov/esds/impact/admg
- Tools presented in workshop:
  - CASEI: https://impact.earthdata.nasa.gov/casei/
  - SOOT Power User Interface: https://asdc.larc.nasa.gov/soot/power-user
  - Soil Moisture Visualizer: https://airmoss.ornl.gov/visualize/
  - Airborne Data Visualizer: https://daac.ornl.gov/tools/airborne-data-visualizer-project-list/
  - Spatial Data Access Tool: https://webmap.ornl.gov/ogc
  - FCX: https://ghrc.earthdata.nasa.gov/fcx/index.html
- ADMG Terms/Definitions: https://earthdata.nasa.gov/esds/impact/admg/admg-definitions
- Previously published articles:
  - The Earth Observer: May-June, 2022, *Summary of the 2022 NASA Airborne and Field Data Workshop*, pg19, https://eospso.nasa.gov/earthobserver/may-jun-2022
  - Earthdata article: *Improving NASA Airborne and Field Campaign Data*, May 18,2022, https://www.earthdata.nasa.gov/learn/articles/admg-workshop-summary

Appendix C: Workshop registration and attendee metrics

Figure C.1 Self-designated registrant status

Legend:
- Student or Postdoc
- Early career
- Mid career
- Late career
- It's complicated

20.9%
34.5%
19.8%
19.8%

Figure C.2 Self-identified registrant user type

data producer 6.0%
both 34.0%
data user 39.0%
neither 21.0%

Figure C.3 Registrant self-reported primary area of area of interest
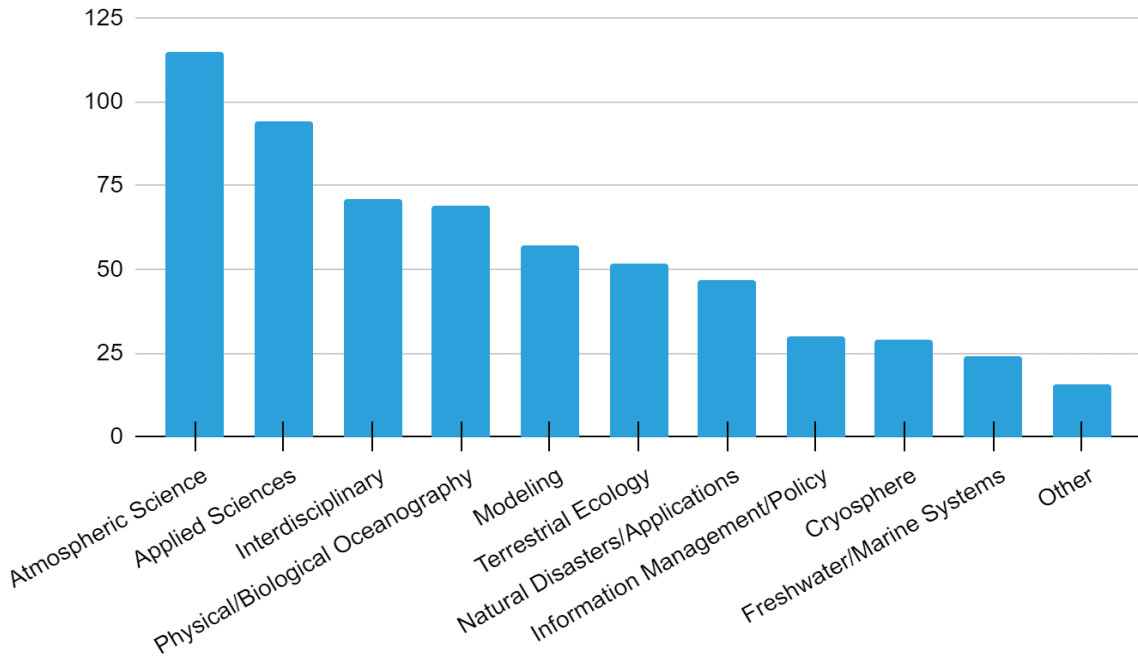
Figure C.4 Summary of workshop metrics

Workshop Attendees

- 272 people registered for the workshop
- 50 to over 100 attendees at any given time
- Maximum attendance occurred on Day 1; lowest attendance at the end of Day 2

Appendix D: Examples of workshop feedback

The workshop provided ample opportunities for participants to provide feedback. Methods used included registration questions, breakout room discussion notes, data user and data producer presentations, meeting chats, panel discussions, Slido polls and Jamboard comments. Some examples are shown below.

Figure D.1 Categorized responses to the registration question "What is your most important feedback about NASA airborne and field data discovery, access, and/or use?"
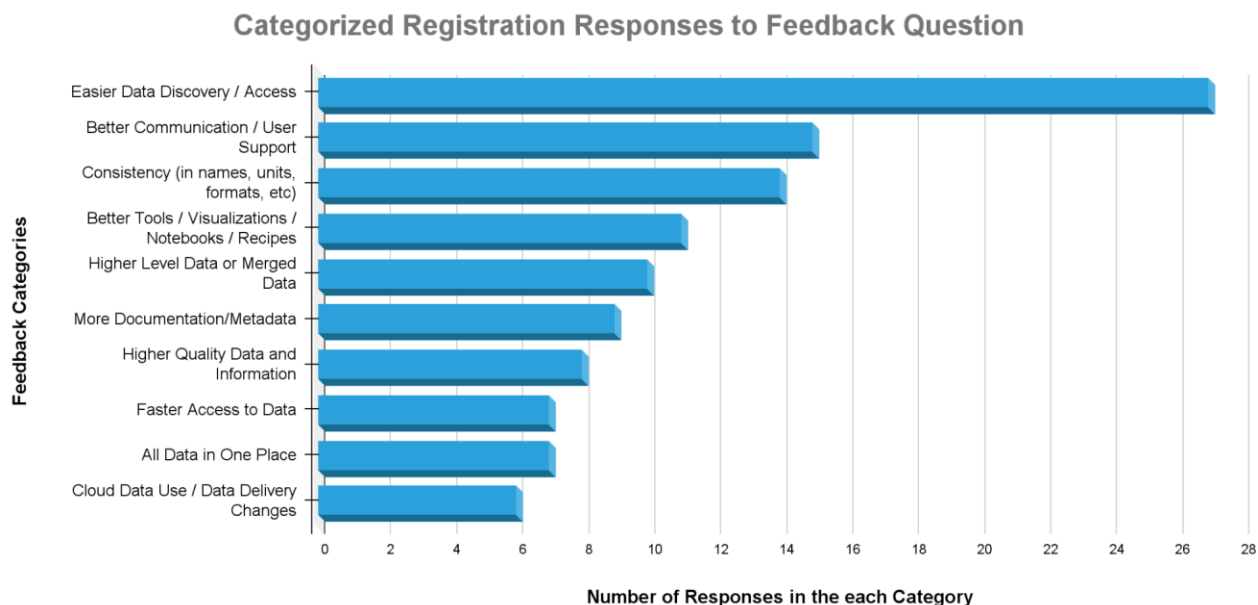
Figure D.2 Example of Jamboard feedback from the tool session on day 1 of the workshop

Figure D.3 Most mentioned tools and resources data users need to help with airborne and field data use

Audience feedback activity: What tools do you need to do your work?

Key Takeaways:

- Clear documentation
- Clear instrument metadata for differentiating campaigns
- Consistent naming conventions
- Consistent data formats
- Searching by variable, altitude, location
- Cloud data access
- APIs for data access
- How to videos
- Data available as zarr format
- User friendly metadata entry for data producers
- Alerts for new versions of data I have accessed before
- Jupyter notebook examples
- Identifying/visualizing overlapping flight tracks with satellite overpasses
- Gridding and subsetting services

Figure D.4 Results of data producer workshop Slido poll on participant knowledge and communication with DAACs and ADMG
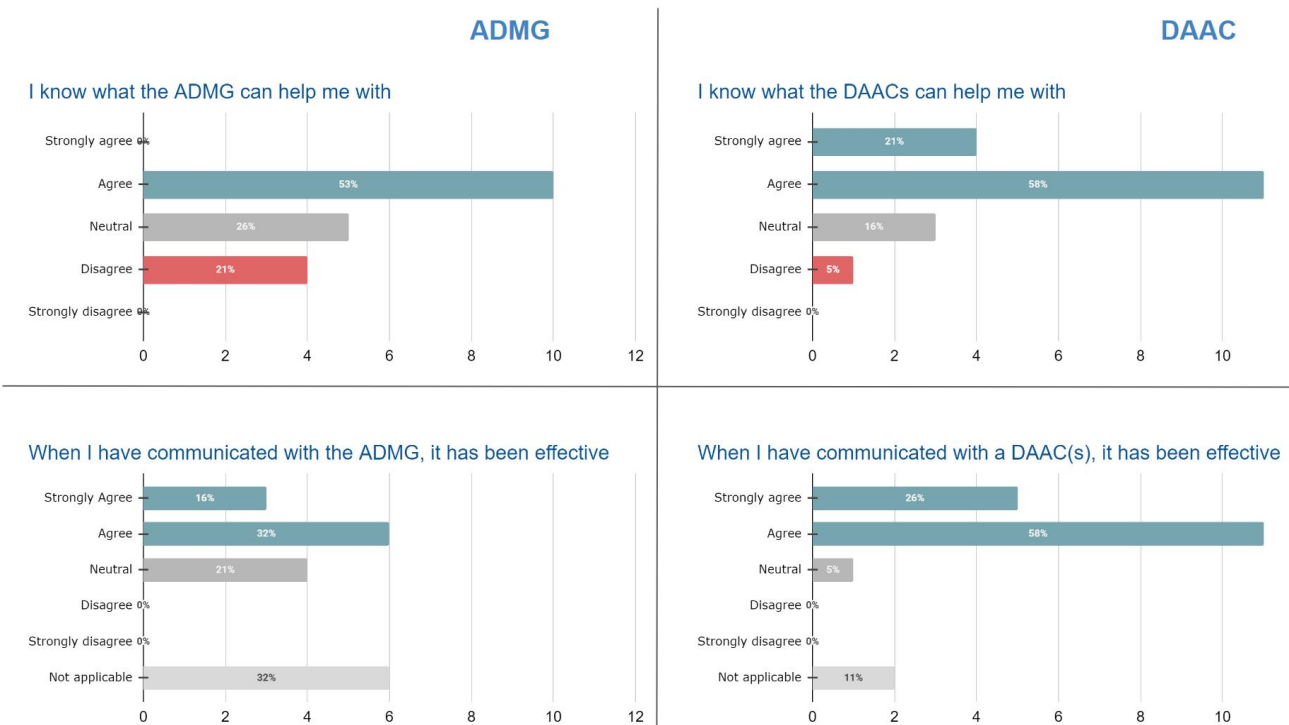
I regularly think to involve the DAAC in meetings/activities

Strongly agree	16%
Agree	37%
Neutral	26%
Disagree	21%
Strongly disagree	0%
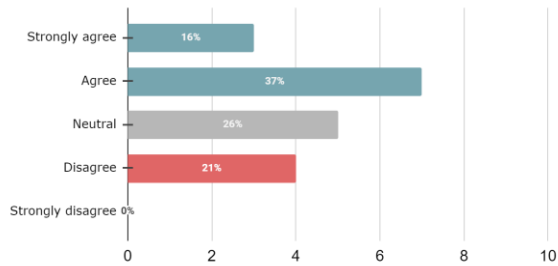
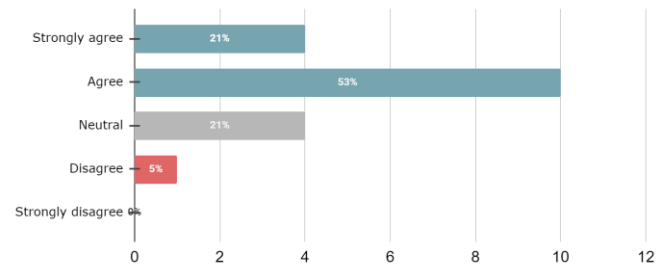**I know when to communicate with the DAAC**

Strongly agree	21%
Agree	53%
Neutral	21%
Disagree	5%
Strongly disagree	0%

Figure D.5 Example of chat questions and comments from workshop session participants

All Question text	Score	Upvotes
Not all PIs are "recalcitrant" by choice. Data providers are overworked, support has not kept up with demand or costs. Funding levels constant + inflation...	6	6
Extra funding for outreach would allow PIs to hire more team members to respond to data inquiries, host open data workshops, and advertise available data sets.	5	5
what would it take for airborne missions to produce gridded data that could be widely used by modelers?	3	3
Will NASA be funding more data analyses for instrument teams, associated with Airborne Science Data?	3	3
What's the carbon footprint of storing all this old data? Some old data has a shelf life, and we might consider what is worth keeping.	2	2
is FCX open source or will be soon/when??	2	2
There are time requirements now on EVS projects for data delivery. Is this not helping with on-time deliveries?	2	2
I think ESCO or DAACs should provide templates to help data providers to create datasets in different formats (e.g., netCDF).	2	2
Any plan to have the ESCO standards in the data management plan of NASA proposals?	2	2
Is there going to be a link built between the DAACs and the high performance computing clusters, like Pleaides at NASA Ames?	1	1
Our HySpex Mjolnir data are not publicly available just yet. But we want to distribute everything!! Right now, we do not have a way to distribute the data	1	1
'Older' data is the basis for longer-term change analyses: will there be work to update the formats of data sets, without changing this non-replicable data?	1	1
@Stephanie...if the campaign I am involved with is in CASEI, but has some outdated metadata, what is the process to update it?	1	1
Any particular reason not to allow multiple groupings using something like tags for ICARTT?	1	1
it occurs to me that oceanographic cruise data is similar to airborne data. Perhaps something can be learned from the solutions that community has come up with	1	1
given it's long history I would imagine there are numerous tools for working with AVIRIS data. ENVI is the default, of course. But it comes as binary, no?,	1	1
which means any programming language can open and read an AVIRIS file. is it free software for e.g. atmospheric correction that a limitation?	1	1
For quality information, is it most needed on the file level or will information about quality at the dataset level be sufficient as a starting point?	1	1
Do NASA DAAC(s) have, or are working, preprint dataset capability like was mentioned with ORNL?	1	1

Figure D.6 Example workshop notes from a breakout session on Day 1

Room 7 - Tammy and Tom

Megan- discoverability of data is largest concern, she works to link users to relevant datasets. Does not use the cloud much. Works with SnowX often. Does not often find that DAACs are behind in publishing Versions of data that is available from original provider. One issue she is having is the ballooning of tools and terminology as new campaigns come about. These things are only summarized in a 'sad appendix at the end' of a report. Funding goes toward the science execution of the campaign, however its key to reserve funds for data management. On improving data management plans- have set guidelines, especially when multiple people are making measurements. This helps to prevent the ballooning of terms, site names, ect.

Gabriel- New to using Airborne data. Finding products has been difficult, unsure where exactly to look.

Matteo- General interest, been working on large campaigns for 15 years. Main curiosity is uniformity in the way data is posted. He is concerned about data that is archived in multiple places and they might not be identical versions. (Tammy- given the need of posting data in a timely manner, some data is released before QA is completed and then updated following QA. One challenge we can address is to find all the places where the data is located and have those in one place, such as EarthData Search).

Megan- It is hard to keep all the repositories straight. (Tammy- EarthData Search hopes to address this by serving as a repository of repositories)

Megan- Timelines can be challenging because researchers want to use data right away, yet there needs to be time to perform QA. This leads to data being transferred to them that is not really ready for analysis'

Takeaways:

- There are a lot of repositories/ places to get data. Knowing where to look is often confusing.
- Having multiple versions of a dataset available from different repositories is a concerning in regards to data integrity
- There is this tradeoff between getting data out to researchers quickly and repositories conducting thorough QA
- The most effective Data Management Plans have clear guidelines and a team committed to following them

Figure D.7 Example workshop notes from a breakout session on Day 2

- What formats do you like to work with?
  - Needs to use ICT and NetCDF- are there any tools that make that easier? With lots and lots of variables.
    - ASDC is working on a conversion tool, and ORNL has a script to make conversions between ICT and NetCDF, lo
    - Gao was talking about a tool that they could put their information in, and see if the file had everything in it — we th https://www.air.larc.nasa.gov/missions/etc/helpFscan.html
      - Available for any campaign
      - here is the freestanding ICARTT File Format Scanning For Windows https://www-air.larc.nasa.gov/missions/et
    - PODAAC has data management best practices (https://podaac.jpl.nasa.gov/PO.DAAC_DataManagementPractices#File%20Formats),
      - `The HDF5 group

        http://www.hdfgroup.org/HDF5/

        Unidata and netCDF
        http://www.unidata.ucar.edu/netcdf/

        An overview of tools and services that work with netCDF
        https://wiki.earthdata.nasa.gov/display/ESO/netCDF+Technical+Working+Group

  - Why do we like netCDF? Universally used by modelers, and it can be odd for some to try to review... but most can figure out ICA
    - Interoperable, easy to use across different systems (same can be true for ICARTT as well)
  - Can make a netCDF as a wrapper, but doesn't necessarily make the data more useful and doesn't always lend itself to all tools (
    - Because naming conventions are not included in netCDF conventions? Yes, part of it
    - netCDF is less accessible to new users, involves more reading up top
    - HDF dump could be a good way to help newer users into a more readable format, so netCDF is not completely in
- Any other formats you wish we had greater access to?
  - COG, often useful file format for some (Cloud optimized geoTIFF, larger than standard GeoTIFF, but can look at layer types very
  - GeoTIFF is another type of data used
  - JSON file is better for communication, can easily be written and read by the broader community
- Multiple formats archived in one way, but depending on the use-case for accessing the data, could make more sense to have the files available

Appendix E: Example of categorized and prioritized user stories

All feedback from the workshop was collected, categorized, and compiled into a set of 60 user stories used to develop the recommendations provided in this report. The figure below is an example of some of the higher priority user stories. The full set of user stories is available from the organizing committee on request.

Priority (must have, need, want, not imp)	Effort (low, high)	Cost (low, high)	Category Short Title	User Story (as a user I want to....)
Must	medium	medium	Access	As a power data user, I want to easily download bulk data.
Must	medium	medium	Access	As a data user, I want to be able to access data grouped by campaign or geographic area or some other variable by accessing virtual collections
Must	medium	medium	earch / Discove	As a data user who is used to Earthdata search, I want more airborne data in EDS because I don't know which DAAC to look at for the data products I need.
Must	medium	medium	Access	As a data user I want to get all the instrument, platform, or campaign data accessible from one place.
Need	high	high	Access	As a data user I want to combine multiple coincident datasets into a single Zarr archive, with relevant coordinates automatically aligned.
Need	high	high	Access	As a data user I want to to get all the instrument, platform or campaign data accessible from one place
Need	low	low	Access	As a data user I want to have a Jupyter notebook to help me use some particular airborne or field data
Need	medium	low	Support	As a data producer, I want to understand and have guidance on providing standardized data formats that meets the protocol for incorporation into community tool resources.
Need	high	low	Support	As a tool developer, I want to understand the common data formats/standards in which airborne data are provided in order to create software products for the user community.
Need	low	low	Metadata	As a data user, I want to identify and work with all data that is part of a campaign, which spans multiple airborne platforms over discontiguous space and time/instrument used in multiple campaigns. I need metadata about these different flights/campaigns that will allow me to search by "virtual collection" or "group"
Need	medium	medium	Metadata	As a data user who wants to search by instrument/campaign, I need metadata that is more consistent, uniform(ish) and searchable for all campaigns and instruments
Need	medium	medium	Consistency	As a data user or data producer, I want consistency in file names, units, file formats and variable names
Need	low	low	Consistency	As a data user who obtains data from more than one DAAC and Earthdata, I want more consistency between archives and all Earthdata sites (e.g. download procedures for bulk download also vary between DAACs)
Need	high	high	Tools	As a data user, I want to easily find and subset different airborne instrument data products and have the data from these products available on the same grid to make it easier to get imagery from the same Earth location.

Figure E.1 User stories created from workshop user feedback prioritized and assessed for effort and cost.

Appendix F: Cloud considerations

NASA's migration to Earthdata Cloud presents many opportunities to address AFDWG workshop recommendations around airborne and field data. As described in the report, the workshop provided information to and feedback from Data Producers and Data Users on NASA's migration to Earthdata Cloud (EDC) and the relevance to airborne and field data. NASA's Earth Science Data Systems (ESDS) Program defines open science as a collaborative culture enabled by technology that empowers the open sharing of data, information, and knowledge within the scientific community and the wider public to accelerate scientific research and understanding. Regarding cloud considerations, the AFDWG is paying particular attention to open science through the needs and recommendations of the user community, cloud-optimized data structures, NASA Enterprise tooling, and in particular advantages of cloud-based discovery, access, and computing.

The AFDWG identified high priority workshop recommendations centered around improving discoverability and access by co-locating data instead dispersing it across individual DAACs. With data co-located in NASA Earthdata Cloud, interoperability of coincident data is easier to improve, as is analysis-readiness. For example, consider two Facility Instruments: MASTER and AVIRIS-NG, both image spectrometers. These instruments have each flown for multiple years supporting many research campaigns. Likewise, a large pool of field vegetation spectra measurements and relevant satellite overpass measurements exist presenting unique opportunities for scaling investigations that incorporate field, suborbital, and orbital coincident data. Similar scaling scenarios exist for in situ air sample measurements or air column estimates as well as the plethora of optical passive and active remote sensing instruments found in the NASA Airborne Science Program. An emerging solution to discovering these data is NASA's Earthdata Common Metadata Repository (CMR) and a growing ecosystem of Spatio Temporal Assess Catalog (STAC) software. The STAC specification is a common language to describe geospatial information, so it can more easily be worked with, indexed, and discovered. The CMR Cloud STAC API provides CMR STAC Catalog endpoints for cloud-hosted STAC Collections where assets are available in the cloud; for example on S3. The CMR Cloud STAC API provides an opportunity to leverage as a back-end to campaign or flight-line catalog services.

The early focus of EDC data migration was to move existing data to cloud storage. Many past and active airborne campaigns have data processed in legacy formats that are at times deprecated, sometimes proprietary, and typically not optimized for cloud access and analysis patterns. Some datasets are prohibitively large and do not lend well to "download workflows" as is the case for many airborne radar and lidar instruments as well as spectrometry (or hyperspectral) datasets. The AFDWG workshop recommendation is to examine airborne and field data holdings and provide consistent collection level cloud-optimized and analysis-ready data based on appropriate use-case access patterns. This includes, for example, transforming deprecated formats, such as HDF4 to HDF5, and considering converting proprietary ENVI

paired binary/header or legacy formats to chunked netCDF and/or Zarr, or COG mirror data stores which leverage the ability of clients issuing HTTP GET range requests to ask for the temporal or spatial file parts of interest. Other emerging cloud optimized solutions exist, such as Kerchunk, which allow zero-copy mapping while improving access patterns. Based on workshop recommendations, the AFDWG sees the need to provide guidance based on demonstrated use-case and access patterns.

AFDWG Workshop recommendations highlight that attention to standardization is critical to improving cloud-based discoverability, access, and interoperability. Variable naming conventions and data format specifications are such considerations. NASA Enterprise tools and services, such as Harmony, that provide data reduction and in-place analysis, also require that data providers follow recommended standardization practices.

Another tool example is the Field Campaign Explorer (FCX). FCX is a cloud-based three-dimensional visualization and analysis tool developed at the NASA Global Hydrometeorology Resource Center (GHRC) Distributed Active Archive Center (DAAC). It was designed to address data exploration needs in portraying diverse and co-incident datasets obtained in NASA sponsored field campaigns. The targets of those field campaigns are often at the forefront in advancing new scientific knowledge. FCX lets end-users examine the interactive nature of observations from multiple data streams, such as satellite, airborne, and ground measurements, in a coincident environment. It can vary view angle, play speed, zoom, display of in-flight instrument reading and analysis charts, as well as changing scenes to follow, unfollow, and switch the scene of interest. Multiple aircraft can be followed and unfollowed at any time to track the in-flight instrument readings. Supporting multiple datasets turns FCX into an extensible framework with a process that reads data in different file formats and generates visualization files, such as CZML, point cloud, or curtains. As part of the data exploration capabilities, users can select a specific field campaign. Once selected, FCX provides numerous options for users to select data to visualize and identify temporally coincident datasets.

In addition, there is an important need for airborne data user educational resources that are themselves organized and discoverable and that lower the barriers to open source scientific evaluation of airborne and field data. Workshop recommendations include "playgrounds" or Hubs, such as Openscapes 2i2c instances, that allow DAACs to collaborate with Science Teams to provide airborne and field tutorials that demonstrate science-based applications of airborne and field data in an environment that allows early adoption, experimentation, and success. Synergies and open code/applications between science managed cloud environments and NASA Earthdata Cloud warrant exploration.