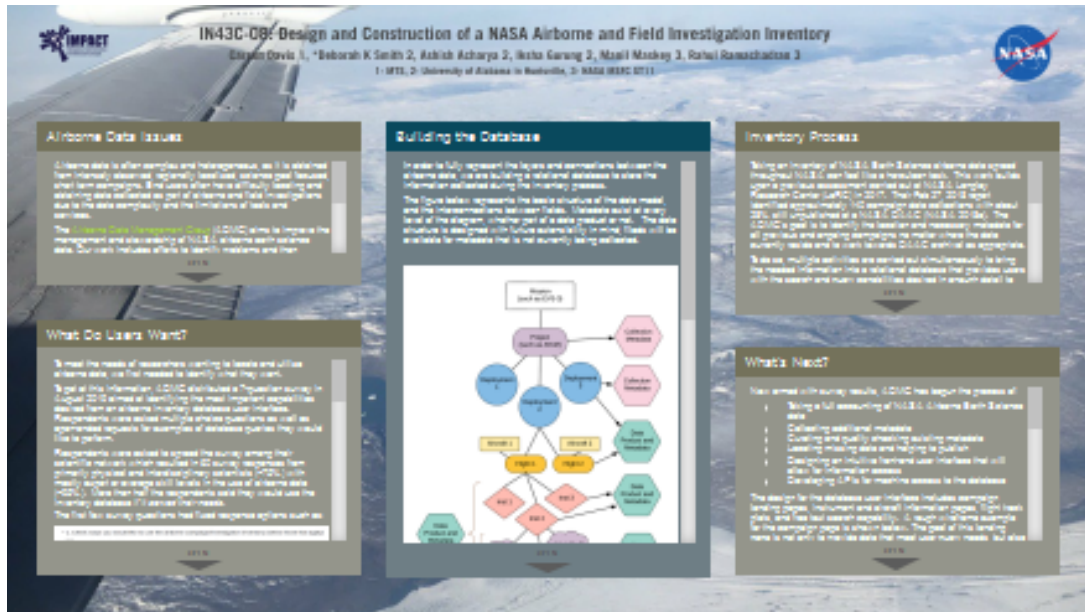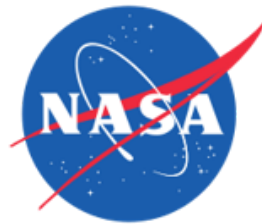# IN43C-08: Design and Construction of a NASA Airborne and Field Investigation Inventory

Carson Davis 1, *Deborah K Smith 2, Ashish Acharya 2, Iksha Gurung 2, Manil Maskey 3, Rahul Ramachadran 3

1- MTS, 2- University of Alabama in Huntsville, 3- NASA MSFC ST11

PRESENTED AT:

AGU 100 ADVANCING EARTH AND SPACE SCIENCE

FALL MEETING

San Francisco, CA | 9–13 December 2019

AIRBORNE DATA ISSUES

Airborne data is often complex and heterogeneous, as it is obtained from intensely observed, regionally localized, science goal focused, short term campaigns. End users often have difficulty locating and obtaining data collected as part of airborne and field investigations due to the data complexity and the limitations of tools and services.

The Airborne Data Management Group (https://earthdata.nasa.gov/esds/impact/admg) (ADMG) aims to improve the management and stewardship of NASA airborne earth science data. Our work includes efforts to identify problems and then improve metadata content, data access, usability, and the discovery of airborne data.
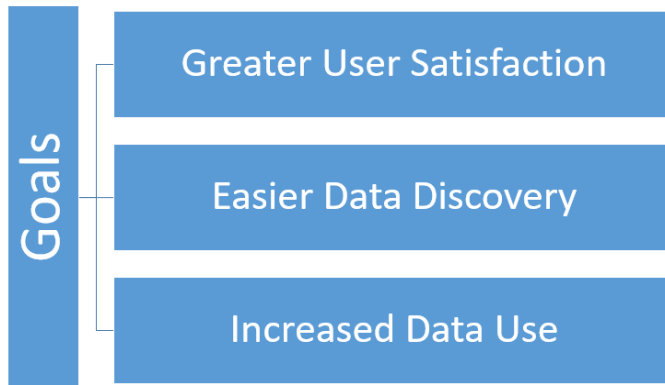
ADMG Goals with Respect to Airborne Data

The primary issues with airborne data include:

- Many users identify dissatisfaction with the tools and services for accessing airborne data.
- There is insufficient metadata for answering scientist queries.
- Valuable data is scattered throughout NASA (even on personal hard drives!)
- Data complexity impedes metadata organization for maximum benefit and ease of use.

Solving airborne issues can include a variety of solution spaces, including:

Improve Airborne Metadata

(make data more discoverable)

Require Use of Standard Data Formats

(make data easier to use)

Develop Better Tools and Services

(make data easier to visualize)

Improve Communication

(make data better understood)

Improve Data Awareness and Access

(increase data users)

Potential Solution Spaces for Improving Airborne Issues

WHAT DO USERS WANT?

To meet the needs of researchers wanting to locate and utilize airborne data, we first needed to identify what they want.
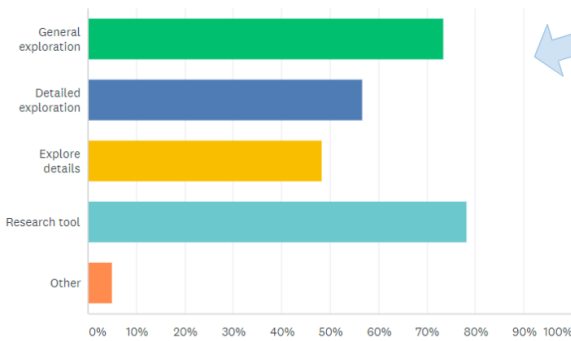
To get at this information, ADMG distributed a 7-question survey in August 2019 aimed at identifying the most important capabilities desired from an airborne inventory database user interface. Respondents were asked multiple choice questions as well as open-ended requests for examples of database queries they would like to perform.

Respondents were asked to spread the survey among their scientific network which resulted in 60 survey responses from primarily physical and interdisciplinary scientists (>75%) with mostly expert or average skill levels in the use of airborne data (>95%). More than half the respondents said they would use the inventory database *if it served their needs*.

The first few survey questions had fixed response options such as:

Survey question Example with Analysed Results

The survey also asked "How important are the following features". The bars with the greatest portion of red and brown showed that time/space data location, links to data, and searchable flight tracks were a priority.
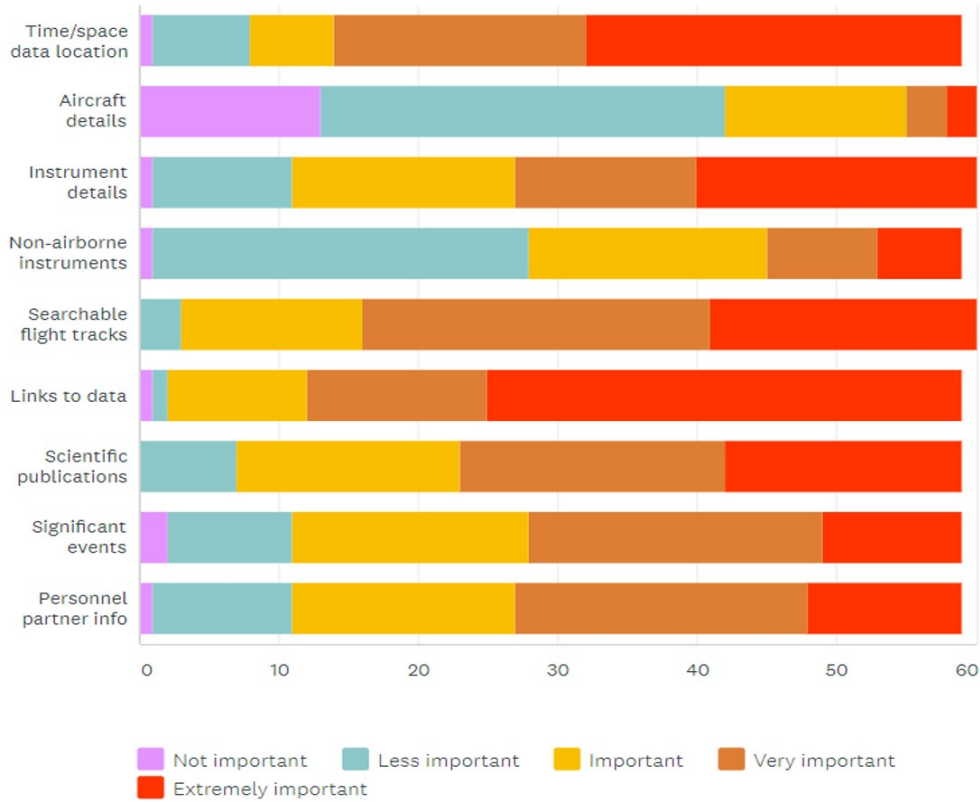
The Importance of Specific Features/Capabilities for the Airborne Inventory

The survey then requested for examples of queries. More than 100 queries were categorized by type and assessed for implementation difficulty. Most prevalent were multi-part questions that sought to access the intersection of two types of information, such as "what campaigns studied aerosol atmospheric content over ocean regions?"

What airborne missions measured air-quality variables?
What airborne studies sampled the tropical Pacific troposphere?
What aircraft data were collected over Alaska from 2015 to 2019?
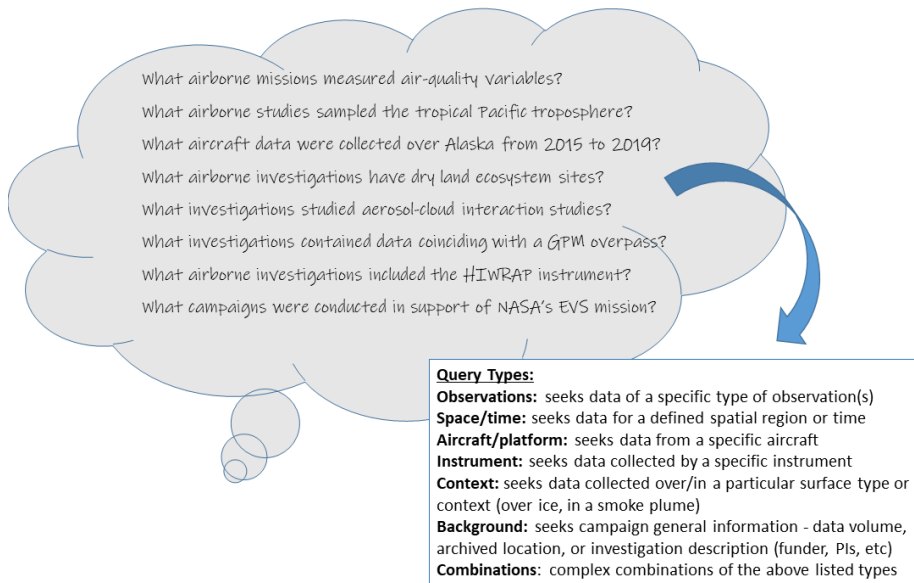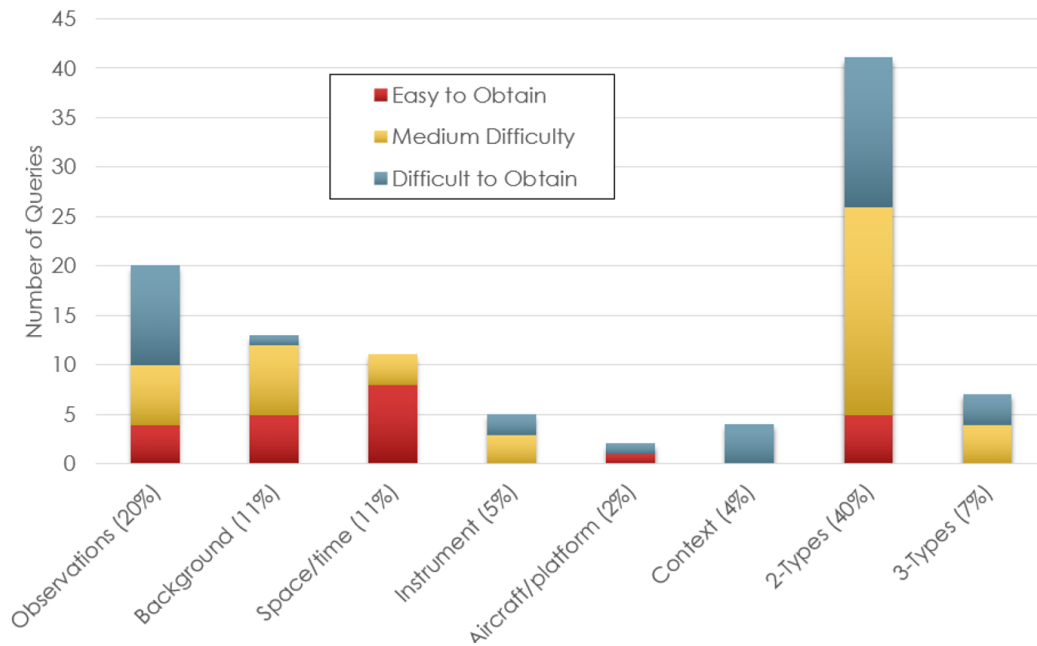What airborne investigations have dry land ecosystem sites?
What investigations studied aerosol-cloud interaction studies?
What investigations contained data coinciding with a GPM overpass?
What airborne investigations included the HIWRAP instrument?
What campaigns were conducted in support of NASA's EVS mission?

Query Types:
Observations: seeks data of a specific type of observation(s)
Space/time: seeks data for a defined spatial region or time
Aircraft/platform: seeks data from a specific aircraft
Instrument: seeks data collected by a specific instrument
Context: seeks data collected over/in a particular surface type or context (over ice, in a smoke plume)
Background: seeks campaign general information - data volume, archived location, or investigation description (funder, PIs, etc)
Combinations: complex combinations of the above listed types

Example User Queries and Query Categories Used

Number of Queries per Category with Implementation Complexity Indicated

Findings from the analyzed results have led to changes in the inventory process, metadata curation, and database construction plans.

The most important survey take-aways are:

- Put a higher priority on adding searchable flight tracks
- Put a lower priority on supplying aircraft details
- Invest time in curating existing metadata to improve quality and science relevance
- Add more metadata to allow for the type of discovery users want

An important question we later wanted to determine, but did not ask in the survey, was if the desired queries can already be performed using some tool or service, why are they asking for them? Or why are the tools or services not being used?

We decided to identify which queries can presently be performed using NASA's Earthdata Search (https://search.earthdata.nasa.gov/search). Our assessment shows that ~60% of identified queries are already possible using Earthdata Search. The planned additional inventory metadata will make it possible to answer about 20% more queries. The remaining 20% of survey queries either require significant additional metadata that is either too costly (in time or effort) to add or is just not possible as it involves some type of detailed data analysis.

If we assume that these are _desired_ queries, then one may also assume respondants either do not know to use Earthdata Search for discovery or do not find it easy to use (indicating usability issues may exist). Since we did not ask about Earthdata Search, we can not determine which is the case. However, we are designing the airborne inventory database user interface to make it easy for scientists to intuitively locate desired information. We also intend to broadly announce the inventory interface once completed.

BUILDING THE DATABASE

In order to fully represent the layers and connections between the airborne data, we are building a relational database to store the information collected during the inventory process.

The figure below represents the basic structure of the data model, and the interconnections between fields. Metadata exist at every level of the diagram, whether part of a data product or not. The data structure is designed with furture extensibility in mind; fileds will be available for metadata that is not currently being collected.

Information Layers of Airborne Investigation Data used in the Development of the Database
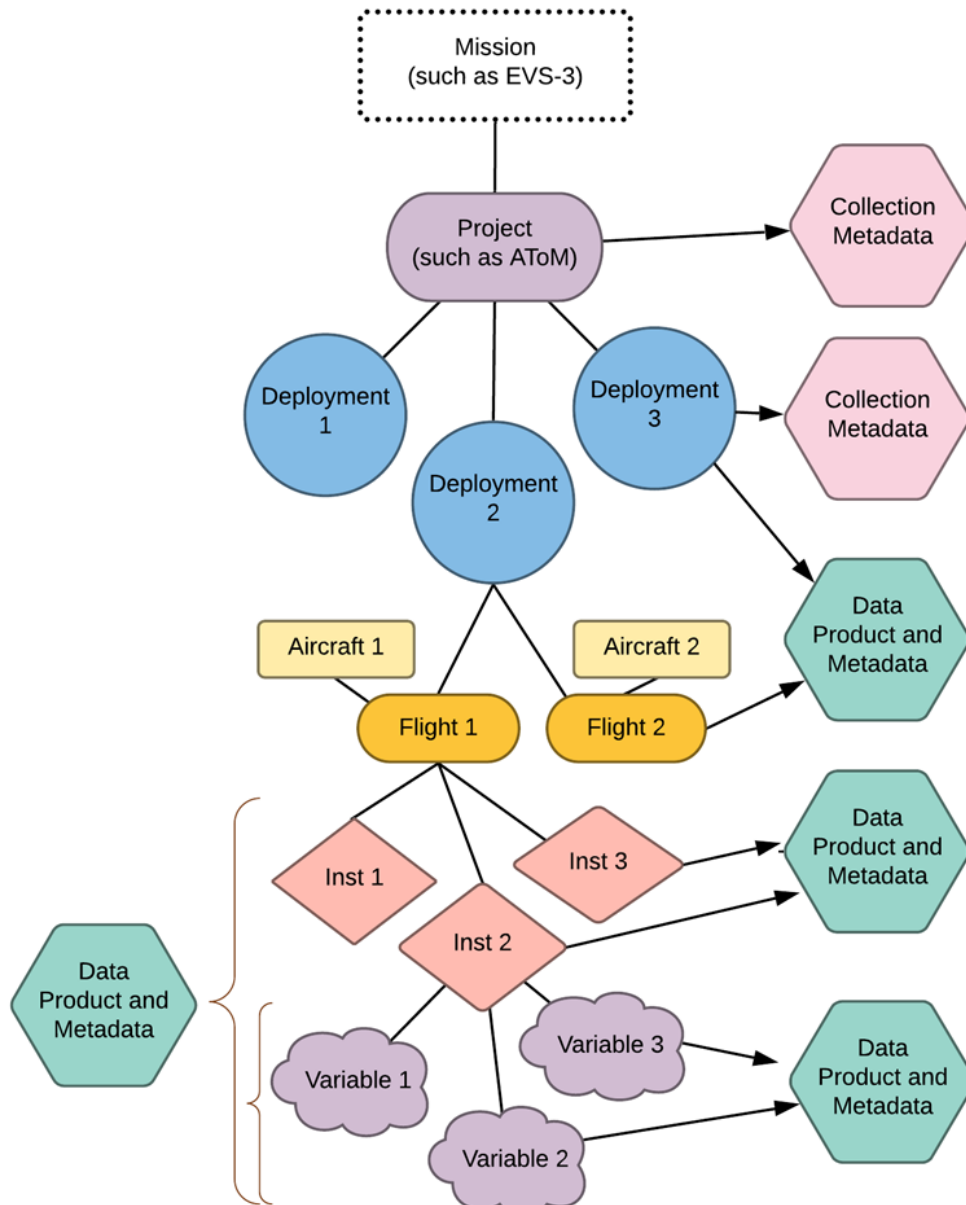
Some of our desired metadata already exists in the NASA Earth Science Data System (ESDS) Common Metadata Repository (https://earthdata.nasa.gov/eosdis/science-system-description/eosdis-components/cmr) (CMR) and will be incorporated into our database as we add the additional information needed to thoroughly describe the investigation. The additional metadata will more fully describe the deployment information, significant event details, aircraft and instrument details, measured parameters, scientific goals, and responsible personnel.

The inventory curation process will collect this additional metadata for every NASA Earth Science airborne investigation in order to allow for the complex database inquiries desired by users.

One of the greatest challenges to gathering and organizing NASA's airborne campaign data and information into one resource is the different definitions in use by the community and the different approaches employed by various DAACs when publishing the data products. We found it necessary to apply consistent definitions and requirements to existing metadata.

The following definitions are used:

Term	Definition
Mission	A NASA-funded program containing one or more satellites and/or field projects. Missions target a relatively broad set of specific science requirements, while supporting projects each tend to investigate a more narrow set of mission-supporting objectives and/or research questions. Examples include the Aqua mission, the Precipitation Measurement mission, and the Earth Venture-Suborbital missions
Field Investigation*	An observational study during which individuals/programs/agencies utilize preselected, specific sensors or sets of sensors to acquire targeted observations in a natural, non-laboratory setting to support common, clearly defined science or research objectives, usually over a designated geographical space and/or period of time. Sensors can include either or both in-situ and remote sensing instruments operating on airborne, ground-based, and/or other non-satellite platform(s). Satellite-based observations may be included, but are not collected specifically as part of, a *field investigation*.
Deployment	A previously scheduled, continuous time period during which a *field investigation*'s platforms and/or sensors are dedicated for use in support of the *investigation* objective(s). The *deployment* time frame typically consists of the calendar date range during which there are mission and/or instrument scientists and/or NASA ESPO personnel stationed on-site for the *investigation*. There may be time periods within a deployment during which individual platforms and/or instruments do not operate due to unfavorable environmental conditions, mechanical issues/maintenance, and/or prescribed personnel rest time. There may be one or more *deployments* within a single *field investigation*.
Intensive Operation Period (IOP)	An individual time frame, within a *deployment*, during which all or a subset of the *field investigation*'s instrument(s) and platform(s) are operated to observe phenomena in support of the *field investigation* science or research objective(s). *IOPs* are the primary measurement periods within a single *deployment*. A single *IOP* may contain observations of one or more events or cases, and these may or may not be considered *significant events* or *golden cases*. There may be one or more *IOPs* within a single *deployment*.
Significant Event	An event observed during a *deployment* that is notable for the occurrence of one or more specific phenomena relating to the *field investigation* science or research objective(s). Significant events may occur within or outside of an *IOP*.
Golden Case	An event during an *IOP* that is notable for having all or nearly all of a *field investigation*'s selected instrument(s) or platform(s) operating successfully, and/or a *significant event* which has data available from most of the deployed instruments/platforms. It is possible for a *significant event* to also be a *golden case*.
Partner Organization	An institution that participates in a *field investigation* with separate funding resources (not NASA) allocated specifically for the *field investigation*.

*Field investigations are also called airborne investigations, campaigns, projects, missions, deployments, or IOPs. These terms have not been consistently used over time or research fields

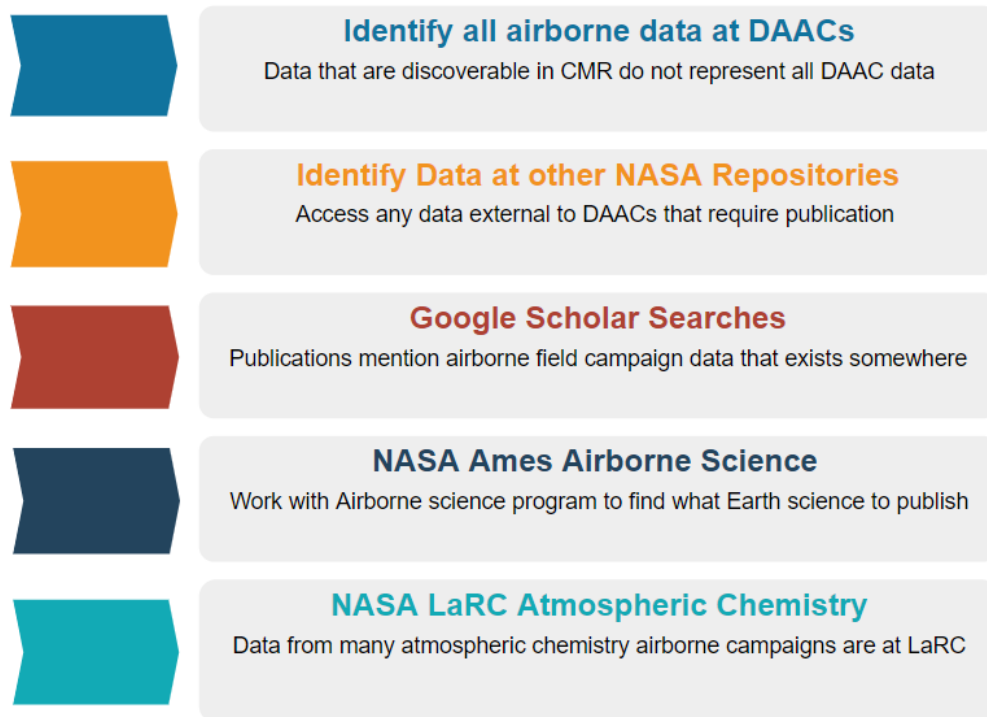Definitions of Utilized in the Airborne Inventory

The relational database information schema is quite detailed and contains over 100 metadata items. APIs are being developed to incorporate existing CMR data product metadata and also information from the NASA Airborne Science program (https://airbornescience.nasa.gov/ (https://airbornescience.nasa.gov/)). ADMG team members add additional metadata through the use of a form and maintenance interface.

INVENTORY PROCESS

Taking an inventory of NASA Earth Science airborne data spread throughout NASA can feel like a herculean task. This work builds upon a previous assessment carried out at NASA Langley Research Center (LaRC) in 2017. Their Feb 27, 2018 report identified approximately 140 campaign data collections with about 28% still unpublished at a NASA DAAC (NASA 2018a). The ADMG's goal is to identify the location and necessary metadata for all previous and ongoing campaigns no matter where the data currently reside and to work towards DAAC archival as appropriate.

To do so, multiple activities are carried out simultaneously to bring the needed information into a relational database that provides users with the search and query capabilities desired in enough detail to improve data discovery and user satisfaction. Multiple sources are utilized during the process and existing metadata are collected and curated to make sure the quality meets the needs of the inventory and consistency is maintained across all data.

We first need to identify what airborne data should exist and where it is located. We can do so using some of the following approaches:

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

Approaches to Locating NASA Airborne Data

The primary effort is to:

- Locate all available public data and curate and expand the metadata by focusing on the details of campaigns, deployments, significant events, instrumentation, and context
- Collect and curate additional metadata by reviewing publications, campaign documentation, and other authoritative sources
- Shepard unpublished data through data publication workflows to bring data located during the inventory process into a DAAC
- Locate existing DAAC data missing from CMR and work with the DAAC to incorporate it.

WHAT'S NEXT?

Now armed with survey results, ADMG has begun the process of:

- Taking a full accounting of NASA Airborne Earth Science data
- Collecting additional metadata
- Curating and quality checking existing metadata
- Locating missing data and helping to publish
- Designing an intuitive front-end user interface that will allow for information access
- Developing APIs for machine access to the database

The design for the database user interface includes campaign landing pages, instrument and aircraft information pages, flight track plots, and free text search capability. A rough wireframe example for the campaign page is shown below. The goal of this landing page is not only to provide data that meet user query needs, but also to provide detailed campaign metadata and links to existing public data products.
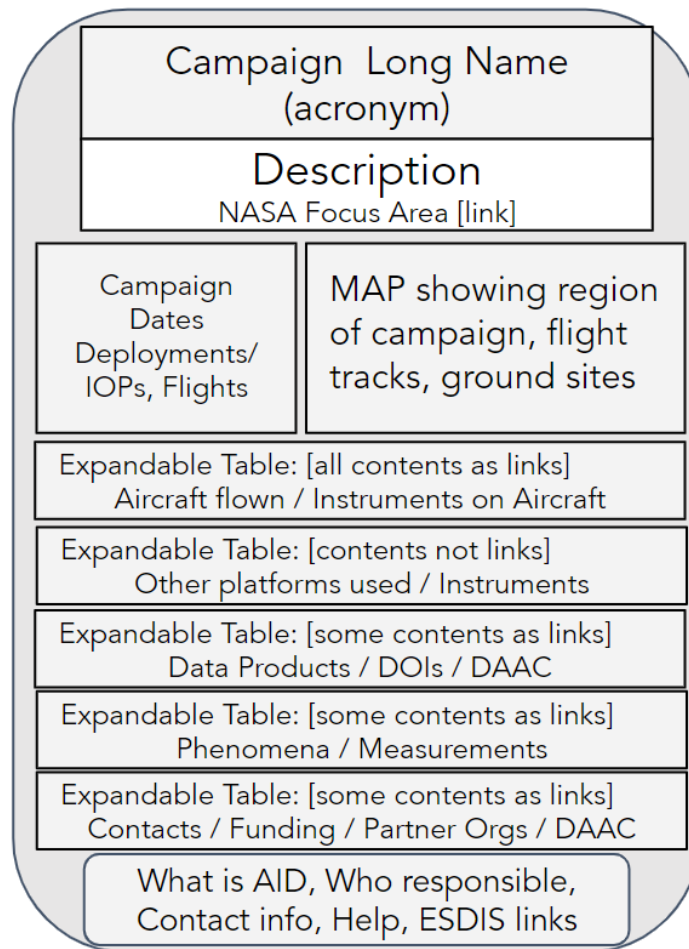
Wireframe Example of Possible Campaign Description Page

After significant testing, including a soft release to a select group of scientists, we will release the database interface to users later in 2020. Survey responses identified how important it is to build an intuitive user interface and provide users with clear documentation, webinars, and user videos to help them feel comfortable with interface features and capabilities. We plan for users to have the ability to provide ADMG with feedback so we can continue to make improvements.

What will be the long term location for the inventory database and user access? It is hoped that with the proven value of the additional metadata that either the database interface will be moved to NASA Earthdata website, or that Earthdata Search improvements will be made to incorporate the capabilities demonstrated in the inventory database user interface.

We Welcome Your Feedback!

CV

Deborah K. Smith is the presenting author. She leads the Airborne Data Management Group (ADMG) at the University of Alabama in Huntsville.

NASA's Airborne Data Management Group (https://earthdata.nasa.gov/esds/impact/admg) was established to support the NASA Airborne Science community, Distributed Active Archive Centers (DAACs), and airborne data users by working to improve access to resources and data, supplying information about existing and past airborne projects, and taking a full accounting of NASA data. The ADMG is located within the Interagency Implementation and Advanced Concepts Team (IMPACT) and operates under the direction of NASA's Earth Science Data Systems. Learn more about the ADMG at https://earthdata.nasa.gov/esds/impact/admg (https://earthdata.nasa.gov/esds/impact/admg)

If you have questions or suggestions, please contact deborah.smith@uah.edu

ABSTRACT

NASA conducts airborne and field investigations that produce a wealth of valuable research data. Unfortunately, this data is often scattered across individual scientist hard drives or NASA Distributed Active Archive Centers and it can be difficult to locate and retrieve. Although satellite data has been successfully consolidated by tools such as EarthData Search, airborne and field investigation data present unique challenges stemming from the variability of temporal, spatial, platform, and instrument metadata. To address these difficulties with data retrieval and metadata variability, the Interagency Implementation and Concepts Team established an Airborne Data Management Group to improve airborne data search, understanding, access, and use. Surveys have been conducted of end users in order to build query lists that will drive the augmentation and standardization of existing metadata. Detailed metadata was then laboriously compiled from present and historic airborne and field investigations to build a database that will enable intelligent data search and retrieval. The inventory structure and function will be described and demonstrated. The purpose of this presentation is to bring awareness to this effort, to highlight and describe the issues and complications in development, and to increase user interest prior to public release in 2020.

REFERENCES

NASA, 2018a, *Management of Airborne Data Products: Challenges and Recommendations*, February 27, 2018. Available from ADMG (https://earthdata.nasa.gov/esds/impact/admg).

NASA, 2018b, *NASA Airborne Data Holdings Overview,* Presentation to NASA Earth Science Data Systems, Winter 2018. Available from ADMG (https://earthdata.nasa.gov/esds/impact/admg).

SWITCH TEMPLATE