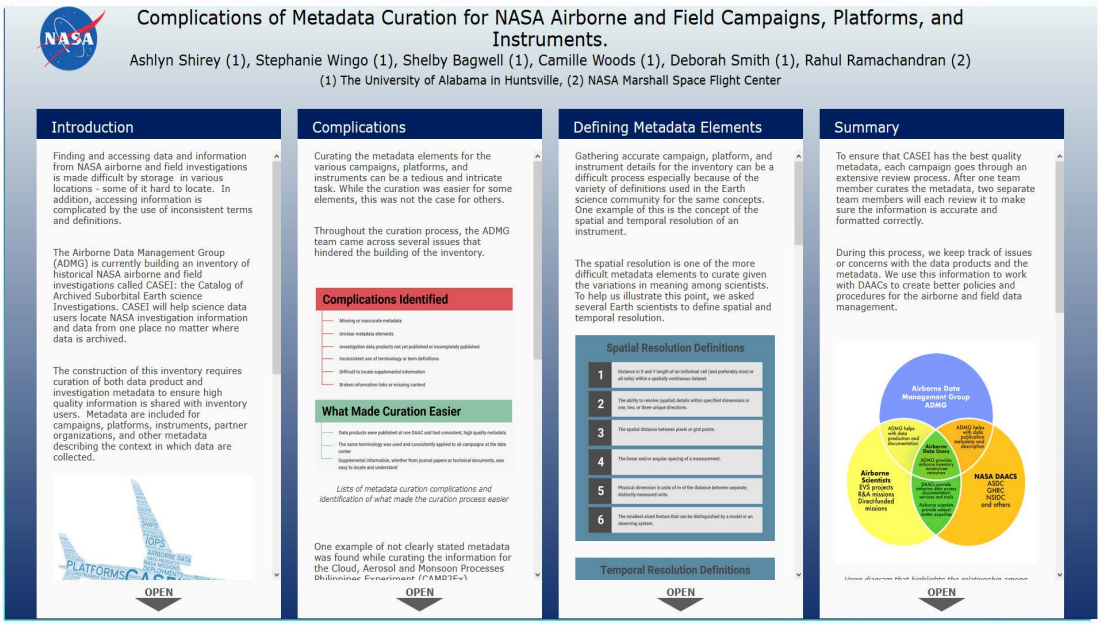
Complications of Metadata Curation for NASA Airborne and Field Campaigns, Platforms, and Instruments.



Ashlyn Shirey (1), Stephanie Wingo (1), Shelby Bagwell (1), Camille Woods (1), Deborah Smith (1), Rahul Ramachandran (2)

(1) The University of Alabama in Huntsville, (2) NASA Marshall Space Flight Center

PRESENTED AT:

AGU FALL MEETING
Online Everywhere | 1-17 December 2020

INTRODUCTION

Finding and accessing data and information from NASA airborne and field investigations is made difficult by storage in various locations - some of it hard to locate. In addition, accessing information is complicated by the use of inconsistent terms and definitions.

The Airborne Data Management Group (ADMG) is currently building an inventory of historical NASA airborne and field investigations called CASEI: the Catalog of Archived Suborbital Earth science Investigations. CASEI will help science data users locate NASA investigation information and data from one place no matter where data is archived.

The construction of this inventory requires curation of both data product and investigation metadata to ensure high quality information is shared with inventory users. Metadata are included for campaigns, platforms, instruments, partner organizations, and other metadata describing the context in which data are collected.

Word cloud for CASEI with terms related to items in the inventory

During the curation process, ADMG identifies many metadata issues that require improvement to contextual campaign and data product metadata. These issues slow the ADMG effort and complicate the work.

One of the main complications was the use of various definitions for terms in use by science teams across the different scientific fields that carry out airborne and field investigations.

To address this issue, ADMG identified a set of formalized definitions for metadata terms to help make the process of building the inventory easier. This enables more efficient discovery and access to NASA observations by allowing science data users to search for certain clearly defined metadata values.

COMPLICATIONS

Curating the metadata elements for the various campaigns, platforms, and instruments can be a tedious and intricate task. While the curation was easier for some elements, this was not the case for others.

Throughout the curation process, the ADMG team came across several issues that hindered the building of the inventory.

Complications Identified

- Missing or inaccurate metadata
- Unclear metadata elements
- Investigation data products not yet published or incompletely published
- Inconsistent use of terminology or term definitions
- Difficult to locate supplemental information
- Broken information links or missing content

What Made Curation Easier

- Data products were published at one DAAC and had consistent, high quality metadata
- The same terminology was used and consistently applied to all campaigns at the data center
- Supplemental information, whether from journal papers or technical documents, was easy to locate and understand

Lists of metadata curation complications and identification of what made the curation process easier

One example of not clearly stated metadata was found while curating the information for the Cloud, Aerosol and Monsoon Processes Philippines Experiment (CAMP2Ex) campaign. The cloud probes that were used were generally referred to as "SPEC Cloud Probes" on the instrument list on the campaign website.

RSP	NASA P-3	Research Scanning Polarimeter	Brian Cairns (PI)
SPEC Cloud Probes	NASA P-3 SPEC Lear Jet		
SSFR	NASA P-3	Solar Spectral Flux Radiometer	Peter Pilewskie (Co-I)

Part of the instrument list from CAMP2Ex

Probes is plural, so this indicates that multiple cloud probes were used, but it is not clear what the exact probes were. The only information we have is that they are manufactured by SPEC, which is an instrument company. However, SPEC has many different instruments identified as probes, so again it is not clear what instruments were used.

Without more information, we can not determine which specific SPEC cloud probes were used during the campaign. More research is needed to find the answer.

Issues like this make curating the metadata more difficult, but it is beneficial for science data users that we provide exact details to the inventory.

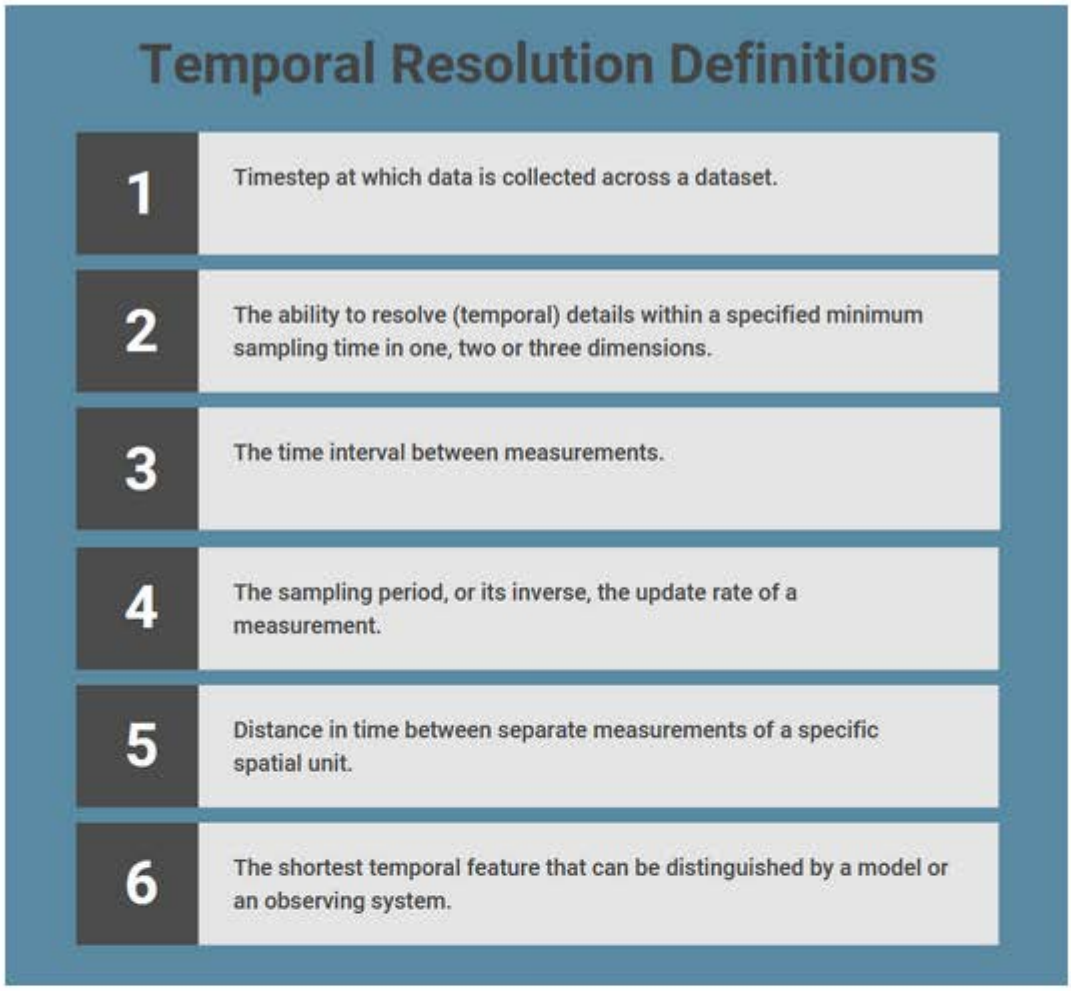
DEFINING METADATA ELEMENTS

Gathering accurate campaign, platform, and instrument details for the inventory can be a difficult process especially because of the variety of definitions used in the Earth science community for the same concepts. One example of this is the concept of the spatial and temporal resolution of an instrument.

The spatial resolution is one of the more difficult metadata elements to curate given the variations in meaning among scientists. To help us illustrate this point, we asked several Earth scientists to define spatial and temporal resolution.

Spatial Resolution Definitions

1	Distance in X and Y length of an individual cell (and preferably most or all cells) within a spatially continuous dataset.
2	The ability to resolve (spatial) details within specified dimensions in one, two, or three unique directions.
3	The spatial distance between pixels or grid points.
4	The linear and/or angular spacing of a measurement.
5	Physical dimension in units of m of the distance between separate, distinctly-measured units.
6	The smallest-sized feature that can be distinguished by a model or an observing system.



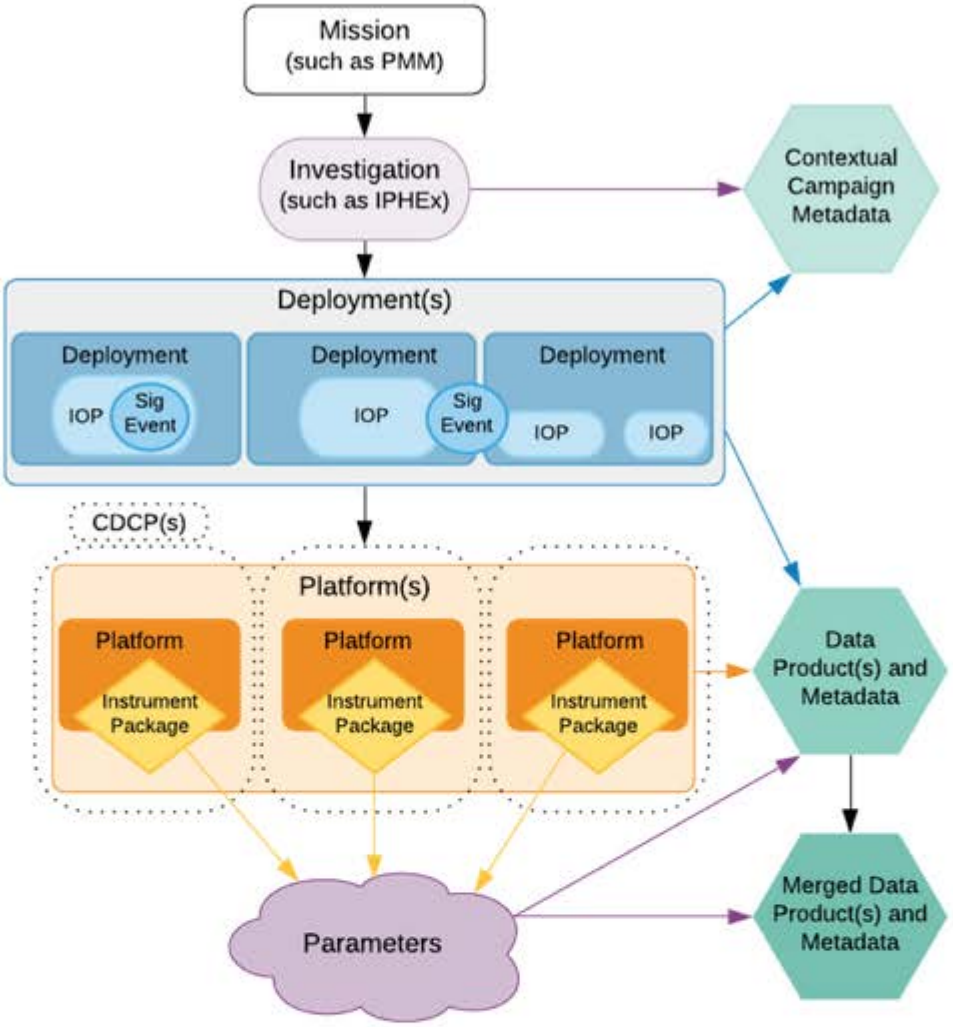
Spatial and temporal resolution definitions provided by several Earth scientists in response to an ADMG survey

For spatial resolution, some scientists stated that it was the spatial distance or spacing between each point, while others said that it was the size of the actual cell. Some responses define it as the smallest feature resolved from the observing system. While these characteristics are related, the multiple descriptions make it difficult to convey to users which is meant by the one term.

For temporal resolution, most scientists defined it as the time between each measurement. Others referred to it as the temporal features or details that are resolved.

Spatial and temporal resolution is often referred to as spatial and temporal coverage such as listed in the EOSDIS Glossary (https://earthdata.nasa.gov/learn/user-resources/glossary#ed-glossary-s). However, many feel coverage is not the same concept as resolution.

These examples highlight the need for more clarified definitions to enable consistent information sharing. In building CASEI, the ADMG has set formalized definitions and built an information model that consistently relates and applies the metadata to the inventory. This allows for the information we collect to be handled in a more consistent way across multiple campaigns.



The ADMG information model used for the CASEI inventory

Term Definitions:

Field Investigation: An observational study during which individuals/programs/agencies/institutions utilize preselected, specific sensors or sets of sensors to acquire targeted observations in a natural, non-laboratory setting in support of common, clearly defined, science or research objectives.

Platform: A structure, item, or surface on which an instrument and/or instrument package can be placed for operation. Platforms can be moving or stationary; permanent or temporary; and airborne, land-/ground-, or water-based. ADMG has identified platform types, guided largely by GCMD's Platform keywords' Category level.

Instrument Package: All instruments operating on a platform. This may consist of a single or multiple individual instruments and all their associated technology. The instrument package used on a given platform can change over time, by deployment, or by field investigation, to support specific science/research objectives or due to individual instrument failure. For airborne platforms, the instrument package is also often referred to as the instrument payload.

Continuous Data Collection Period (CDCP): A continuous period of data collection by a platform during which the instrument

package is operated. Continuous Data Collection Periods (CDCPs) are based on the activity or operation of a platform, not on individual instruments comprising the instrument package on the platform.

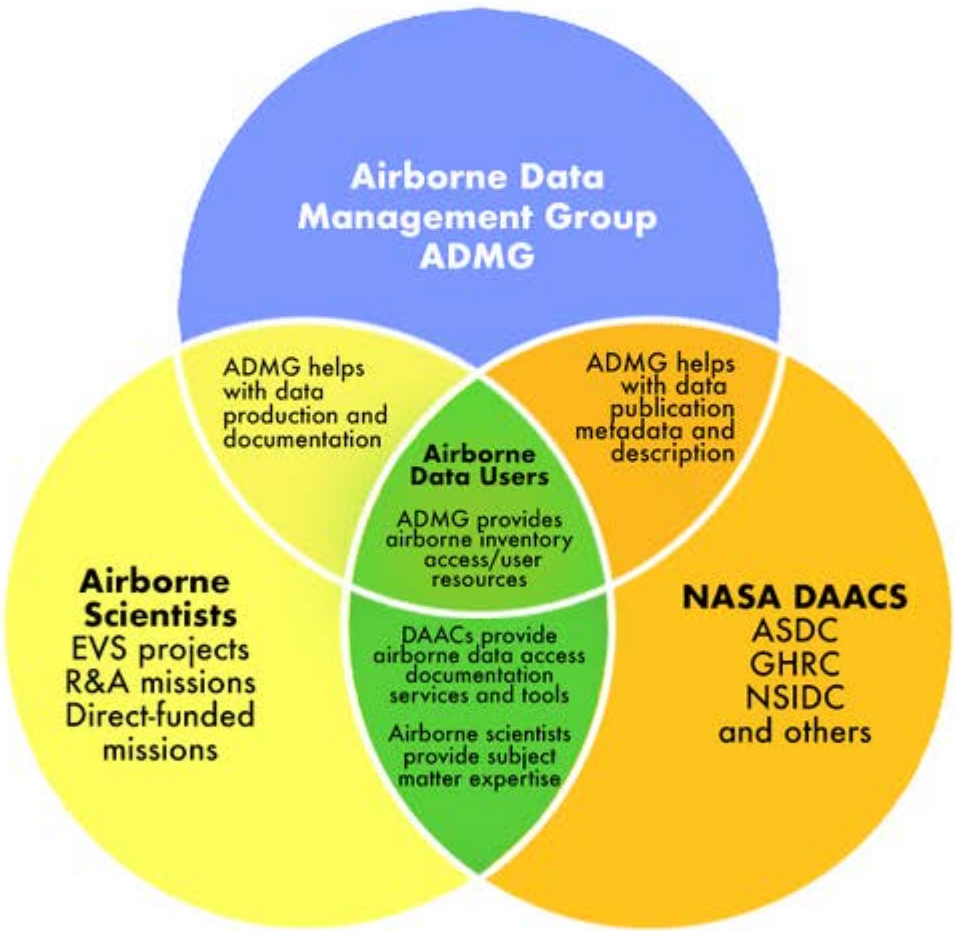
To read more about ADMG's definitions visit the definition website:

https://earthdata.nasa.gov/esds/impact/admg/admg-definitions (https://earthdata.nasa.gov/esds/impact/admg/admg-definitions)

SUMMARY

To ensure that CASEI has the best quality metadata, each campaign goes through an extensive review process. After one team member curates the metadata, two separate team members will each review it to make sure the information is accurate and formatted correctly.

During this process, we keep track of issues or concerns with the data products and the metadata. We use this information to work with DAACs to create better policies and procedures for the airborne and field data management.



Venn diagram that highlights the relationship among ADMG, NASA DAACS, and Airborne Scientists.

Several issues have been discovered during the CASEI development process carried out by the ADMG.

One major complicating issue has been the difficulty presented when metadata elements have several different terms or definitions used across various scientific fields. The example shown here of the terms spatial and temporal resolution of instruments, are just one of these cases. The definitions provided by several scientists highlight the need for the use of more universal definitions across the Earth Science scientific fields.

The ADMG team is tackling this issue by deriving and using formalized definitions and decision trees in order to improve metadata curation and information communication for airborne and field investigations, platforms, and instruments. Through this effort, the metadata collected for the inventory will be consistent across every campaign and will enable more efficient data discovery and access to NASA observations by allowing CASEI users to search for certain clearly defined criteria and metadata values.

The CASEI inventory user interface will be available to the public in 2021.

To learn more about what ADMG does, visit this website:

https://earthdata.nasa.gov/esds/impact/admg (https://earthdata.nasa.gov/esds/impact/admg)

And check out these other ADMG presentations at AGU:

- Improvements to Open Data Sharing of NASA Airborne and Field Investigation Data - SY028-03, 9 December 2020 19:07 - 19:10 PST
- Development of a Suborbital Portal with End-Users and Data Archivers in Mind - IN044-03, 16 December 2020 07:08 - 07:12 PST
- Improving Data User Accessibility to Investigation Notable Events - H140-0002, 14 December 2020 04:00 - 20:59 PST

ABSTRACT

The Airborne Data Management Group (ADMG) curates metadata that describe NASA's airborne and field campaigns, platforms and instruments. This activity is vital to building a useful inventory of sub-orbital Earth science data that improves data discovery and access. During the curation process, many metadata issues were identified that required improvement to campaign and data product metadata. In some cases, locating the needed metadata to add to the inventory was a simple process. For other cases, the information was hard to find. In addition, identifying accurate investigation and instrument details to add to the inventory was especially complicated because of the variety of definitions used in the Earth science community for the same concepts. One example of this is the concept of instruments' spatial and temporal resolution. The spatial resolution is one of the more difficult elements to curate given the variations in meaning across various disciplines. Clarified definitions are needed to enable consistency of information across campaigns and instruments. In this presentation, we introduce results from a survey of scientists from various fields in which we asked for definitions of spatial and temporal resolution. Our survey results highlight the importance of creating more universally acceptable definitions for certain metadata elements. By curating sub-orbital field campaigns and instrument metadata, ADMG is enabling more efficient discovery and access to NASA observations by allowing science data users to search for certain clearly defined criteria and metadata values.