

Search, Discovery, and Accessibility of Field Campaign Data: Status Update

Alice Cialella¹, Daniel Larkin¹, Richard Cederwall³, Richard Wagener¹, Kathy Lazar¹, and Karen Gibson³



¹Brookhaven National Laboratory, Upton, NY 11973
³Oak Ridge National Laboratory, Oak Ridge, TN 37831



1. Problem

Atmospheric Radiation Measurement (ARM) Climate Research Facility (ACRF) data collected during field campaigns (FC) throughout the year must be assigned metadata (namely instrument classes and primary measurements) before they can be “discovered” by a user through the ARM website. In the summer of 2010 a significant portion of the field campaign data received metadata assignments. New instrument classes and primary measurement types were defined to pave the way for the influx of former Aerosol Science Program (ASP) field campaign data, and for aerosol data from the largest field campaign to date, Carbonaceous Aerosol and Radiative Effects Study (CARES). However, due to lack of personnel resources neither the ASP field campaign data nor the CARES field campaign data had been assigned metadata. The current project addressed these deficiencies.

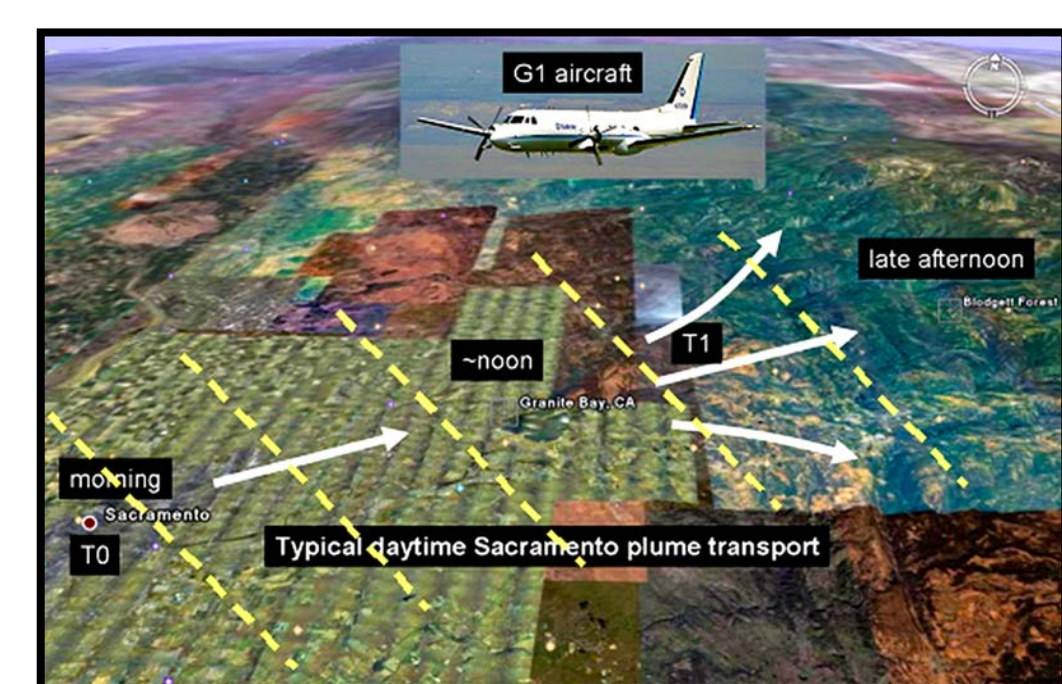


Figure 2. CARES Proposed ground sites (T0 & T1) and G-1 airborne instrumentation strategy.



Figure 3. Tomlinson UHSAS deployed on G-1 Aircraft during CARES FC.

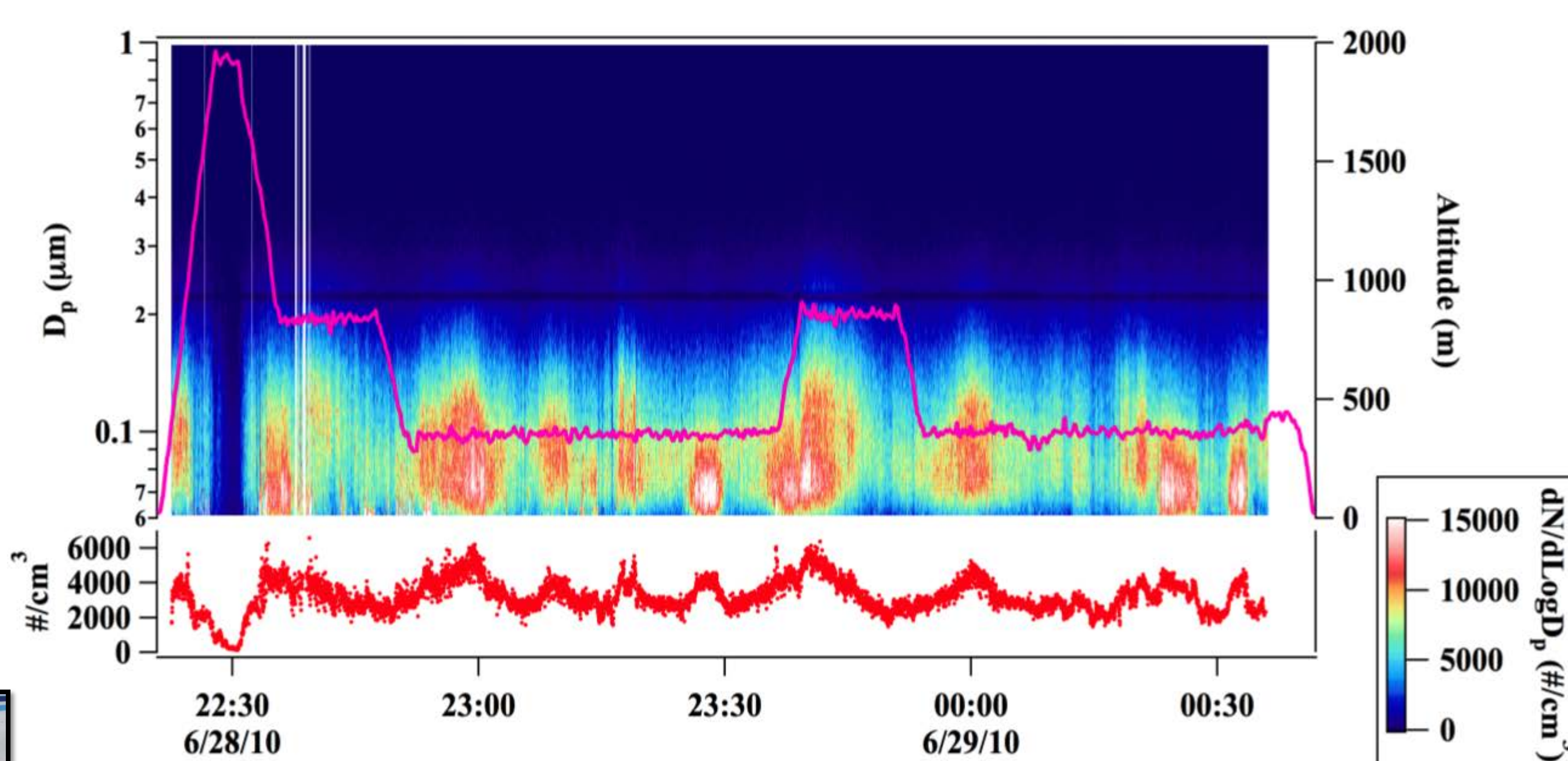


Figure 4. Time series of normalized ($dN/d\log D_p$) number concentration and total number concentration (bottom plot) from flight 20100628b.

3. Results

Field campaign metadata assignments increased to a record 95% by the end of the summer of 2011, with the primary effort put toward the CARES datasets. Figure 5 illustrates metadata assignments for a CARES field campaign dataset. The rest of the year was dedicated to entering and assigning all the metadata for the former ASP field campaigns.

Several new classes were created to accommodate the recent CARES datasets and upcoming aerosol and cloud physics measurements.

Additional insight from the process identified the need for a standardized procedure for PI documentation input.

Site	IOP Name	Start Date	PI Name	Data Source Name	Are we expecting data?	dataLink	Instrument Categories	Instrument Class	Source Classes	Measurement Categories	Primary Measurement Types
AAF (MCC)	Carbonaceous Aerosol and Radiative Effects Study (CARES) [IOP#-3820, Prop#-5793]	2010-06-02	Tomlinson [PIInst#-4043]	Ultra High Sensitivity Aerosol Spectrometer-G1 Aircraft	Yes	Yes	aerosol airborne	uhsas	iop	aeros(microchem)	aeropartszedistr aerosconc partsizedistrib

Figure 5. FC database tool for assigning metadata. Example of a completed metadata assignment of Tomlinson UHSAS dataset.

5. Future Work

The web-base documentation form will be beta-tested as a new means for entering data documentation by PIs after the completion of the upcoming AMF1 field campaign, TCAP on Cape Cod, MA.

2. Assignment Process

The metadata assignments are made in the ARM Field Campaign Database using an existing web-based database application. The flow chart in Figure 1 illustrates the process.

If an instrument class describing the instrument does not exist, a new class is created and instrument categories associated with it. Similarly, new primary measurement types (PMT) are defined and associated with measurements.

The ARM metadata review team discuss the assignments, via a database tool, and once in agreement the metadata are implemented in the database.

A dataset from the CARES field campaign (Figure 2) is used here as an example. Using a Ultra High Sensitivity Aerosol Spectrometer (UHSAS) instrument aboard the G-1 (Figure 3) the principle investigator (PI) collected aerosol number concentration data. Figure 4 is a sample of the output data.

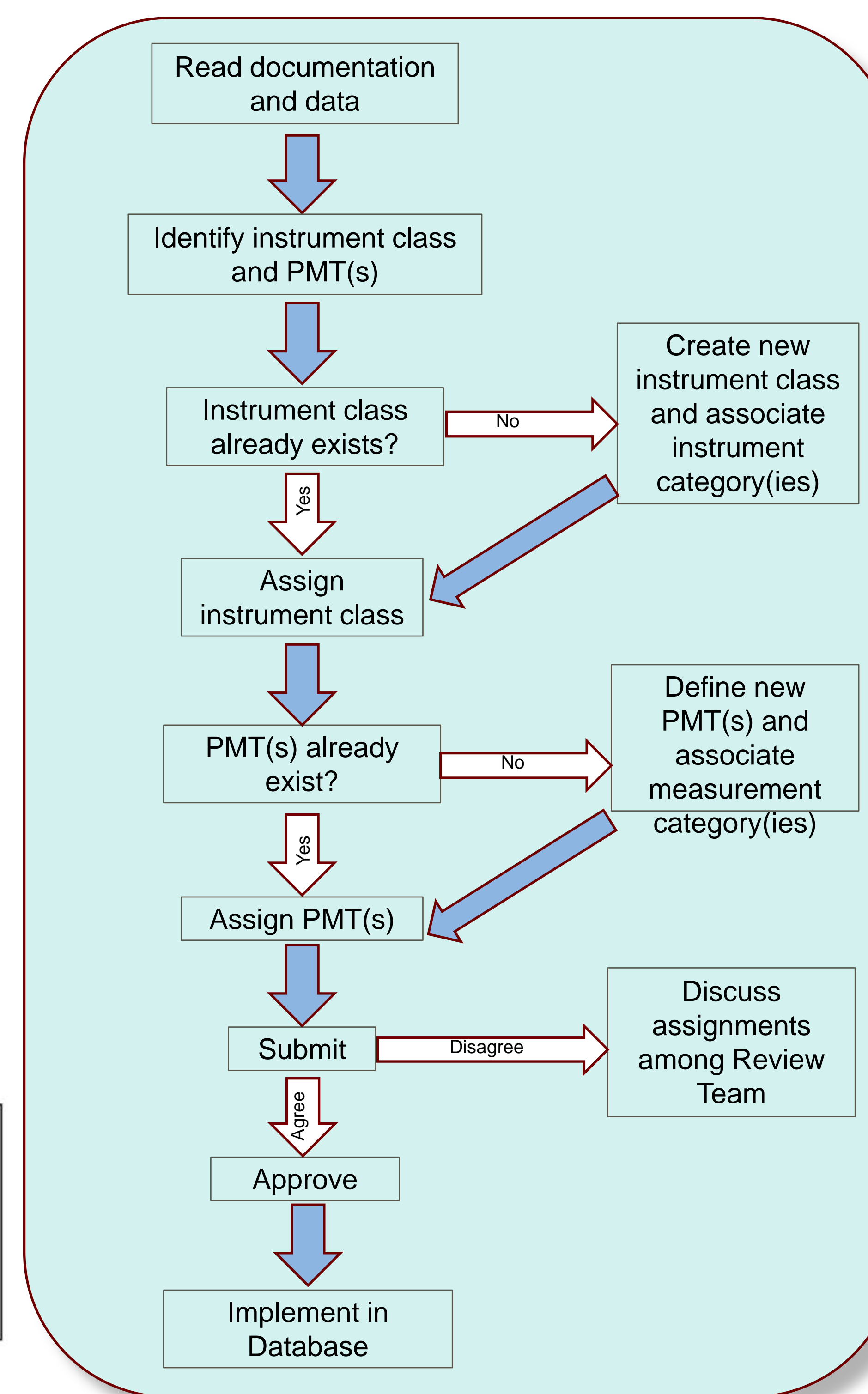


Figure 1. Flow chart of the metadata assignment methodology.

4. Discussion

The ARM program is dedicated to providing access and discovery of its scientific data to investigators in a timely manner. However by 2010, a large quantity of unassigned datasets in the FC archive had accumulated. As of December 2011, this project had successfully updated the FC database to an unsurpassed level of completeness.

Additionally, this effort highlighted a primary difficulty in the metadata assignment process. Often the metadata team must process complex datasets accompanied by incomplete documentation. Thus the need for a web-based documentation template for PIs to use is critical to prevent the emergence of a similar backlog. Therefore, a general field campaign documentation template that would incorporate the following input sections and allow input via a web form is in the early stages of development.

The documentation sections would include:

- Instrument Source (or algorithms, derived products, etc.)
- Primary Measurement(s)
- Time and Location
- Data Structure
- Flight Data Log (if applicable)
- File Format (including temporal frequency)
- Sample Figures
- Contact information
- Additional References
- Data Quality Information

Advantages are:

- PI ease of use
- Documentation completeness (omissions are reduced)
- Documentation uniformity