



U.S. DEPARTMENT OF
ENERGY

Office of
Science

DOE/SC-ARM-TR-116

ARM Standards Policy Committee Report

A Cialella
A Koontz
R McCoy
G Palanisamy
C Sivaraman

M Jensen
S McFarlane
J Monroe
R Perez

September 2012



DISCLAIMER

This report was prepared as an account of work sponsored by the U.S. Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the U.S. Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the U.S. Government or any agency thereof.

ARM Standards Policy Committee Report

A Cialella	M Jensen
A Koontz	S McFarlane
R McCoy	J Monroe
G Palanisamy	R Perez
C Sivaraman	

September 2012

Work supported by the U.S. Department of Energy,
Office of Science, Office of Biological and Environmental Research

Acronyms and Abbreviations

ARM	Atmospheric Radiation Measurement (Climate Research Facility)
CF	Climate and Forecast
DOD	Data Object Design
DOI	Digital Object Identifier
DQ Explorer	Data Quality Explorer
DQ Office	Data Quality Office
DQR	Data Quality Report
EWO	Engineering Work Order
IDL	Interactive Data Language
ISDE	Integrated Software Development Environment
PCM	Process Configuration Manager
PI	principal investigator
QC	quality control
VAP	value-added product

Contents

Acronyms and Abbreviations	ii
1.0 ARM Standards Policy Committee	1
1.1 Purpose and Value of Committee.....	1
1.2 Issues Around Imposing a Universal Approach to Standards Compliance.....	1
1.3 Viewing Standards as a “Black Box”	2
2.0 Documentation and Communication of Current ARM Standards.....	2
2.1 Current Standards Locations	2
2.2 Recommendations for Current Standards.....	3
2.2.1 Absolute Versus Recommended Standards.....	3
2.2.2 Committee-Noted Omissions	3
3.0 Data Products Compliance	4
3.1 Cost of Total Compliance	4
3.2 Alternatives to Total Compliance	5
4.0 Exceptions to Standards.....	5
4.1 Description of Current Exceptions	5
4.1.1 Exceptions to Documented Standards:.....	6
4.1.2 Exceptions for Undocumented Standards:	8
4.2 Implications of Meeting/Not Meeting Standards	9
5.0 Process for Evaluating Exceptions	10
5.1 Current Exception Process	10
5.2 Identifying Exceptions	10
5.3 Proposed Exception Process.....	11
5.3.1 Size and Members of Review Committee	12
5.3.2 Review and Approval of Exceptions.....	12
5.3.3 Documentation of Exceptions	12
6.0 Summary and Recommendations	12
Appendix A ARM Standards Policy Charge Letter.....	A.1
Appendix B List of Current Standards Documentation	B.1
Appendix C ARM Standards Policy Committee Members	C.1

1.0 ARM Standards Policy Committee

1.1 Purpose and Value of Committee

Data and metadata standards promote the consistent recording of information and are necessary to ensure the stability and high quality of Atmospheric Radiation Measurement (ARM) Climate Research Facility data products for scientific users. Standards also enable automated routines to be developed to examine data, which leads to more efficient operations and assessment of data quality. Although ARM Infrastructure agrees on the utility of data and metadata standards, there is significant confusion over the existing standards and the process for allowing the release of new data products with exceptions to the standards. The ARM Standards Policy Committee was initiated in March 2012 to develop a set of policies and best practices for ARM data and metadata standards.

The goals of the committee were to:

1. assess current documentation and communication of ARM standards and eliminate the confusion about what the standards are
2. understand what standards apply to different levels of data (a0–c1), what exceptions currently exist, and the cost of bringing all data into compliance
3. understand the reasons for allowing exceptions to the standards and create a process for evaluating exception requests.

The outcomes of the committee should lead to less confusion about ARM standards within the infrastructure, consistent application of policies across the ARM Facility, and a more formal process for evaluating requested exceptions to standards.

1.2 Issues Around Imposing a Universal Approach to Standards Compliance

The diversity of ARM data products, from raw instrument-level data (00) to highly processed data intended for climate model evaluation (showcase data sets), and the range of instruments, from standard surface meteorological instrumentation to one-of-a-kind scanning radars, make it difficult to impose a universal approach to standards compliance. For instance, most ARM data files are daily, but some products are yearly, and standard variables or attributes required in daily files may not be meaningful for yearly files. Quality information may be well-defined for instrument-level data but much less clear for higher-level derived products. Some non-ARM communities (radar community, climate modeling community, etc.) who are users of ARM data products may have their own standards or conventions, and it may not be possible to simultaneously satisfy both communities.

Additionally, the long history of the ARM Facility means that data products developed at different times may have different standards. Although efforts were made to update all products to the current standards, some historical datastreams were “grandfathered” in as standards evolved due to the large effort required to reprocess them. This exception for historical products has led to further propagation of non-standard datastreams. For example, when one of these grandfathered datastreams is used in a downstream value-

added product (VAP), the VAP output has retained the historical non-standard variable names from the original datastream to reduce confusion for data users, which makes the downstream VAP data non-standard as well.

1.3 Viewing Standards as a “Black Box”

In order to focus on defining policy, rather than risk getting sidetracked with discussions of what the standards are or should be, the committee decided to treat the standards as a “black box”. Therefore the guidelines and processes proposed in this document assume that:

1. All standards have been reviewed and updated, and it has been determined which standards are required versus recommended.
2. The standards are located in one openly accessible web location or linked from a single web page.
3. Exceptions will be allowed but should be kept to a minimum, and must be reviewed and approved by a designated group.

2.0 Documentation and Communication of Current ARM Standards

The current state of documentation of ARM standards is scattered, duplicative, and incomplete. It is strongly recommended that the ARM standards documentation be reviewed, revised, and consolidated in a single location.

Additionally, many ARM standards are “de facto” standards. That is, they are standards that have been applied to nearly all ARM data since the beginning of the program but never explicitly documented. When the existing standards documentation was prepared, the audiences for the documentation were the experienced ARM ingest and VAP developer and other infrastructure members. As a result, some standards were assumed because “we have always done it this way”. As the cadre of developers expanded, these assumed rules were sometimes unclear or completely unknown. It is reasonable to expect new data to meet these standards, but only if they are known and documented. The committee recommends that all standards be documented.

ARM standards are constantly evolving. New ARM standards were adopted in 1995, 2001, 2004, and 2007, and there is currently a proposal to more closely meet Climate and Forecast (CF) compliance. It is reasonable to expect that this evolution will continue as the Facility, user community, and computing technology continue to evolve. See Section 3 for more discussion on CF standards compliance.

2.1 Current Standards Locations

A list of websites where existing ARM standards documentation is located is included in Appendix B. These documents include filename formation rules, rules for assignment of data level, variable naming standards, variable attribute standards, preferred units, global attribute standards, and quality control (QC) standards and documentation requirements.

2.2 Recommendations for Current Standards

Given the current confusing picture of ARM data and metadata standards and the lack of clear and comprehensive documentation, this committee recommends that another working group be formed to revisit and carefully review, update, and/or define new ARM standards.

2.2.1 Absolute Versus Recommended Standards

Defining required standards that are too strict would likely result in ARM data never becoming fully compliant with its own standards. As a result, we strongly recommend that the standards be divided into two categories:

- *core (absolute)* standards to which all data be strictly held
- *recommended* standards that define preferred metadata formats, positions, etc., but which if not followed would not raise an exception.

We also strongly believe that the core standards should only be the smallest set of rules that are vital for the existing infrastructure to work properly. The Process Configuration Manager (PCM) tool should be modified to strictly enforce the core standards and raise a warning flag if the recommended standards are not met.

We also recognize that the same standards cannot be met for ingest and VAP or evaluation/principal investigator or showcase data (for example, QC standards, or standards for new datastreams such as scanning radar data). Therefore we recommend that some standards should be defined flexibly enough to encompass future data needs as well. In particular, while the base-level QC standards for ingests and VAPs should be essentially the same (formatting, attribute names, etc.) to provide less confusion for developers and those charged with reviewing these products, some flexibility should be added to the core VAP QC standards to account for those products for which strict adherence to the core ingest requirements is prohibitive.

2.2.2 Committee-Noted Omissions

As noted previously, this committee treated ARM data standards as a “black box” for the purposes of this assessment. However, during this process the committee identified several undocumented standards (rules that are rules because they have been followed for many years). This information would be useful to a future working group committee reviewing ARM data standards, so the undocumented standards and recommendations for clarification are included here.

- No dashes and underscores in formation of filenames (filename formation is documented at <http://www.arm.gov/data/docs/plan#naming>. Explicit forbidding of dashes and underscores should be added.)
- Filenames must be less than or equal to 61 chars, per Data Archive database storage requirements.
- Set of previously used variable names (recommended for re-use) should be available.
- Recommended units for all the recommended variable names (‘m’ versus ‘meters’, ‘m’ versus ‘km’) should be available.

- Recommended long names for those defined variables.
- Preference for descriptive variable names, avoiding use of single letters (e.g., e, h, or q) should be a topic for *core* vs. *recommended* standards.
- Recommended names and position of the variable subscript qualifier (names of statistically derived variables, as in mean, climatology, diurnal cycle averaged data etc.), and the position of this qualifier (e.g., variable_std, or std_variable) should be available. (The guideline for the subset of these rules is at <https://engineering.arm.gov/task/standards-dod.html>).
- Recommended position of base_time, time_offset, time, lat, lon, and alt in the netCDF file should be available.
- Rules for recommended global attributes (input_datastreams_*, etc.) should be different for ingest versus VAP/evaluation/yearly products. (See documentation at <https://engineering.arm.gov/task/standards-dod.html>.)
- CF standards should be adopted when available.
- Global attributes for DOI numbers and recommended citation for the data that have the DOI number should be available. (The general guidelines about DOI numbers are at <http://www.arm.gov/data/docs/doi-guidance>.)

3.0 Data Products Compliance

The ARM Facility has been collecting and storing data for nearly twenty years. Many of the early netCDF datastreams do not meet the current standards. Many VAPs that rely on those historical datastreams also do not follow current standards, as mentioned in section 1.2. As a way to break this cycle, a future date, such as January 1, 2013, should be chosen after which all new datastreams under development would adhere to the ARM standards (once the standards have been reviewed and updated).

3.1 Cost of Total Compliance

The committee discussed the question of what it would take to bring all ARM netCDF data into compliance with the currently identified standards. It was agreed that the cost would be prohibitive and outweigh the benefits when considering existing ARM resources and the number of data files that would need to be reprocessed.

An equally challenging prospect is converting all the ARM data to comply with a broader community standard such as CF. For example, total CF compliance could be achieved by following a two-step process:

1. Begin generating all the new data files in CF format.
2. Reprocess all the existing-format data files to convert them to CF format.

A major problem with this approach is that users would need to modify their scripts to process data that will be delivered in two different formats until ARM reprocesses all the existing data. To avoid this issue,

the ARM Data Archive would need to obtain the conversion algorithm from the developers and use it to convert the existing-format data before delivering the files to the users.

3.2 Alternatives to Total Compliance

As an alternative to converting all its data to a standard, ARM could enable an option in the ordering interface to allow users to choose the output file format. The ARM Data Archive would then convert the data from the existing ARM format to the standard format (e.g., ARM standard or CF) before delivering the files to the users. To do this, scripts would need to be developed to convert the ARM data to the proper format. The order could be reproduced in the future based on the version of the data set ordered, the date ordered, and the version of the parsing algorithm used for the order.

An advantage to this approach is the flexibility in delivering the data in one or more user-requested formats including one that complies with standards. In addition, this approach should be less costly than reworking ingests and VAP codes and reprocessing all existing data. A disadvantage is that extra processing time would be required before delivering the data to the users.

Whether total compliance or an alternative such as conversion on delivery is chosen, we recommend using a staged approach, focusing first on key VAPs and other datastreams with recommended measurements (about 100 datastreams, around 5700 GB), then focusing on all VAPs back to the year 2000. It may not be necessary to extend compliance or conversion beyond VAPs and recommended measurements to the source datastreams. This could be decided after ARM has gained experience both with the conversion technology and with user responses to the option of output in a standard form.

4.0 Exceptions to Standards

The committee necessarily departed from the “black box” approach to discuss exceptions to the ARM standards that exist in the ARM Archive. The current exceptions primarily fall into the following categories: datastream naming, data dimensioning, header information (variable naming, ordering, and units), QC information, and global attributes. The reasons for the exceptions vary from historical precedence, standards too narrowly defined, and inappropriateness for a particular user community. Some detailed examples follow.

4.1 Description of Current Exceptions

Several types of exceptions to standards currently exist. Some of the exceptions are for existing well-documented standards, and others are for currently unwritten “de facto” standards. The exceptions discussed below are not necessarily comprehensive.

4.1.1 Exceptions to Documented Standards:

Datastream names:

(Documented at <http://www.arm.gov/data/docs/plan#naming>)

1. Data integration period in minutes: As ARM standards have evolved, the position of temporal information in the datastream names has changed. For example, there is a datastream named “twpskyrad60sC1.b1”, in which the temporal information (“60s”) is near the end of the datastream name in seconds. The temporal integration time should be placed just after the site identifier, as in *sgp30ecorE10.b1*.
2. Time stamp in the file name: The standard is YYYYMMDD, but some historical data have the date set as MMDDYY as in, *sgp30smosE13.a1.941231.000000.cdf*.
3. Base names versus optional qualifiers: The current ARM standards call for the base name or “class name” (e.g., mwr, qcrad, aod) of a product to come before any optional qualifiers in the datastream name, but this is not always followed. For example, the datastream name for the “qcrad” VAP “sgpqcrad1longC1.c1” correctly places the “qcrad” base name before the “1long” optional qualifier, while the datastream name for the “aod” VAP “sgpmfrsraod1michC1.c1” incorrectly places the base name “aod” after “mfrsr”.

Variable names:

(Documented on the engineering web site that is not linked from the main ARM website <https://engineering.arm.gov/task/standards-dod.html>)

There are a variety of issues with variable names. These include:

1. Variable names with a mixture of uppercase and lowercase letters, names that begin with a numerical digit, etc.: One of the guidelines for construction of field names is:

Names will consist of lower-letter, digits and underscores and begins with a letter. Upper case is not to be used. (site/facility designation is the exception).

An example of a non-standard fieldname is *Bs_B_Wet_10um_Neph3W_2*, from the *1aosnephwet.a1* DOD. In this case, the *Bs_B_Wet_10um_Neph3W_2* fieldname was allowed because data were matched from a historical product with a similar measurement. The developer and translator also thought that the field names were much more readable with upper case letters, and since historical data fields were being matched, the exception was allowed.
2. One-character variable names: There are some one-character variable names, which are currently considered exceptions to an unwritten rule. An example of this is the *h* field for sensible heat flux in the *30ecor.b1* DOD.
3. Missing “coordinate” variables: If a dimension other than “time” is used, it is a requirement that a coordinate field be created for that dimension. Example: dimensions:

time = UNLIMITED ; // (2878 currently)

```
range = 1999 ;
```

variables:

The “range” field should be created and filled with 1999 appropriate values. There are netCDF files that do not have this range field.

In the case of aipfith1ogren.c1, the “param2” dimension in the AIPFRH VAP does not follow standards because the input datastream does not have a field level variable named “param2”. While confusing (see example below), in this particular case it is not clear that it would result in a “better” netCDF file. Thus, an exception to the standard was granted.

```
netcdf sgpaipfith1ogrenC1.c1.20070501.000000 {
```

dimensions:

```
time = UNLIMITED ; // (24 currently)
```

```
param3 = 3 ;
```

```
param2 = 2 ;
```

variables:

```
float fRH_Bs_R_10um_2p(time, param2) ;
```

```
fRH_Bs_R_10um_2p:long_name = "Coefficients for 2 parameter fit of Bs_R_10um  
hygroscopic growth as a function of RH" ;
```

```
fRH_Bs_R_10um_2p:units = "unitless" ;
```

```
fRH_Bs_R_10um_2p:missing_value = -9999.f ;
```

The first parameter of “param2” is the “a” term, and the second parameter of “param2” is the “b” term. The 2 parameter fit is described in a global attribute:

```
:equation_2_param_fit = "bsp(RH%)/Bsp(~40%) = a*[1-(RH%/100)]^(-b)" ;
```

4. Improperly named “coordinate” variables: Some data products include coordinate variable names that do not match the dimension name they are referring to. An example of this can be seen in the MICROBASE VAP which uses the second dimension “nheights” and the coordinate variable “height”:

dimensions:

```
time = UNLIMITED ; // (8640 currently)
```

```
nheights = 512 ;
```

```
float height(nheights) ;
```

```
height:long_name = "Height of computed value" ;
```

height:units = "m AGL" ;

5. Inconsistent units: In some netCDF files “meters” is used, for example, while in others, “m” is used.

Bit-packed QC variables:

Documented in:

- Gaustad et al. (2010) at <http://dl.acm.org/citation.cfm?doid=1851476.1851556>
- https://engineering.arm.gov/~gaustad/qc_standards_need_to_fix/final_doc/update_c_datastream.htm
- https://engineering.arm.gov/~gaustad/qc_standards_need_to_fix/final_doc/ingest_requirements.htm
- https://engineering.arm.gov/~gaustad/qc_standards_need_to_fix/final_doc/high_level.htm

Though the QC documentation provides flexibility to the mentors/translators to add QC information, there is an expectation that all VAPs will follow QC standards. The process for allowing exceptions to the QC standards for VAPs has not been well-defined or known.

The high-level QC documentation states that *“It was recognized that it may not be possible to apply the proposed changes to all VAPs, particularly the more complex VAPs that process many datastreams, and those that use or create multi-dimensional fields. If a translator believes that applying any of these changes to their VAP will decrease the integrity of the VAP’s data or its usability they may submit a request to the QC Standard committee for an exception to a particular change or all three.”*

The current documentation also states *“A qc field can be created for all fields for which the translator feels it is necessary to provide access to qc information in a non-bit-packed format. These fields should also have the qc information they present also presented in a qc field that follows ARM qc standards using a bit packing representation.”* This standard is not always followed, as some VAPs are released with only non-bit-packed, integer-based aqc fields. Other VAPs have bit-packed fields along with various aqc fields that do not have a corresponding bit-packed QC field.

Global attributes:

(Documented on the engineering web site that is not linked from the main ARM website

<https://engineering.arm.gov/task/standards-dod.html>)

1. *history*: attribute value is not standard for all VAPs.
2. *input_datastreams*, *input_source*: attributes are not provided for all data sets and may not be appropriate for some VAPs, like yearly data sets.
3. *site_id*, *facility_id*, *command_line*, *process_version*, *dod_version*: attributes not always provided.
4. missing some ingest-specific global attributes.

4.1.2 Exceptions for Undocumented Standards:

Apart from the above, there are some unwritten rules that are enforced either by the PCM tool or by the person in charge of the reviewing the Data Object Designs (DODs). These exceptions were discussed.

Datastream names:

The creator of a VAP algorithm may or may not be included in the datastream name. An example of a datastream name with the creator included is *sgpqcradllongE13.c1*. There currently is no documented standard for including the VAP creator's name in the datastream name.

Order of the variables in the file:

The lat/lon/alt fields are most often the last fields in a netCDF file because most of ARM software products follow this order. However there are some files, such as those for the ARM Climate Modeling Best Estimate showcase product, that do not have these fields placed at the end.

4.2 Implications of Meeting/Not Meeting Standards

The implications of meeting ARM standards are:

1. Automated utilities that expect netCDF files to adhere to standards can be used. For example, utilities used by the Data Quality (DQ) Office such as DQ Explorer and DQ Inspector are built around standardized bit-packed QC variables and can be used to easily process the data when the files meet this standard. Another example is if the datastream name does not follow the standard. In this case an existing tool may be unable to parse the site and facility ID from the name. Overall, if data products meet a core set of standards, the software products used to assess and/or display them can be developed much more efficiently by the DQ Office, other members of ARM Infrastructure, and external data users.
2. Adherence to the standards would lead to better quality and more readily understandable netCDF files. For example, descriptive fieldnames are more meaningful to the users rather than one-character fieldnames.
3. It presents a consistent "look and feel" to data users who are familiar with ARM standards.
4. ARM Infrastructure relies on the specific file name structure and expects to find required variables (for example *base_time* and *time_offset*) and their dimensions in the netCDF files. If those core standards are not met, then special-case software needs to be developed to identify and properly handle the datastream.
5. For example, DQ Inspector and IDL libraries used by the DQ Office parse ARM file names based on the expected location of the site, class name, facility, data level, date, and the periods used for separators. These components of the file name are used for both reading and creating various directory structures on the file system, and many of these directory names show up as menu selections in the DQ Explorer web application. The current libraries do not know how to handle underscores or dashes in the filename, and specialized code would need to be developed to identify these cases.
6. In addition, the DQ Office IDL libraries always use the *base_time* and *time_offset* fields to construct a time array for plotting and the creation of QC metrics. Without the *base_time* field, the typical "time" variable provided in the ARM netCDF files is only an index from the starting point of the file and must be used in conjunction with the date and time in the file name or the "units" attribute of the time field to develop a time array that provides the actual date.

7. Adherence to the ARM data standards provides a much more efficient workflow for developers in the ARM infrastructure when creating new products. For example, through the use of standards, developers can use standardized libraries from ARM's Integrated Software Development Environment (ISDE) during product development. As more products adhere to the standards, fewer exceptions must be added to data product software such as VAPs when ingesting various input datastreams. For developers, encountering fewer exceptions means less of a chance to introduce errors in the code and quicker development time. This lowers the costs for instrument mentors and translators, and it also lowers the cost to the ARM Facility for what should often be unnecessary reprocessing tasks.

5.0 Process for Evaluating Exceptions

The process to identify and evaluate exceptions to ARM data standards currently relies on informal communications and decisions and is in need of a formalized description and proper documentation. This section of the report will attempt to formalize the process by which exceptions to existing ARM data standards are identified, evaluated and implemented.

5.1 Current Exception Process

The current process by which exceptions to ARM data standards are considered is often informally documented and relies on email and phone communication between various parties in the ARM Infrastructure. While components of the ARM reporting system such as Extraview (Engineering Change Request, Baseline Change Request, Engineering Change Order, Engineering Work Order (EWO)) and Data Quality Reports (DQRs) are sometimes used to document these issues, there is no designated official reporting mechanism where exceptions to ARM data standards are documented. This makes the information highly fragmented and difficult to find.

In addition, there is no official process for a person or governing body in the ARM Infrastructure to make binding decisions on disagreements that occur between various parties when an exception to the ARM data standards is proposed. There is a need to define a specific process whereby exceptions to the ARM data standards are identified as simply as possible; evaluated, taking into account several factors including implications for users, other program priorities and costs; implementation if an exception is approved; and communication to the end users.

5.2 Identifying Exceptions

There are currently two primary methods used to identify exceptions. One method involves the use of the ARM PCM web-based tool at <https://engineering.arm.gov/pcm/Main.html>. The PCM tool is used mainly by ingest and VAP developers and allows them to design the DODs for ARM data products. The DODs provide all of the metadata that are present in the header of a typical ARM-standard netCDF file. The PCM tool has been programmed to validate this metadata against many of the current ARM data standards and flag exceptions so they are visible to the user on screen. However, the PCM tool does not provide a feature that allows the user to export a list of these exceptions to an external document.

The second method for identifying exceptions is simply visual inspection of data products by members of the ARM Data Management Facility (DMF), the ARM DQ Office, ARM Data Archive, and ARM instrument mentors and VAP translators. This method relies on the expertise of the various parties, but the complexity of the data standards makes this process prone to human error. (Note: For the VAP products, and other non-ISDE produced data, there should be a procedure established to verify that the DOD of the submitted data matches the approved DOD. Currently one can correct the DOD in PCM tool and submit the data with different DOD by error.)

5.3 Proposed Exception Process

Some specific steps must be taken in order to provide clarity to the process of identifying and reporting exceptions.

First, a small review committee should be created to evaluate each case in which an ARM data product does not meet the standards prior to official release. This committee will have the authority to make decisions regarding the acceptance of these exceptions on a case-by-case basis (with approval by ARM management team). This review committee should be included on all data product EWOs.

Second, prior to the release of an official product (including evaluation products) exceptions must be documented in a clear and consistent manner, in the appropriate EWO, so they can be effectively communicated to the review committee. If an exception is approved, this information should be included in the corresponding EWO, but should also be documented in a global DQR upon release of the product and any official documentation supporting the data product (e.g., instrument handbook, technical report, web page) in order to communicate the exception to the end users.

Third, tools should be created and/or updated to help members of ARM Infrastructure and the review committee identify and document when exceptions occur. One of these updates should include adding a new feature to the PCM tool so it is able to generate a summary report for each data product listing the exceptions. Other tools should be created to (1) track the consistency of data products over time, such as tracking DOD changes or reprocessing of a datastream that occurred due to an updated calibration or other change in an input product, and (2) verify the DODs for products such as non-ISDE VAPs whose processing is not linked to the PCM tool.

Once these steps have been completed, the typical process for identifying and reviewing exceptions should be as follows:

1. The ingest/VAP developers work with the instrument mentors/VAP translators to create a DOD for their product using the PCM tool. Any exceptions that are identified by the PCM tool are exported to a summary report. These exceptions are then reviewed by the developer/mentor/translator and either changed to meet standards or submitted for consideration by the review committee with a description as to why the exceptions are necessary. In addition, relevant information obtained from data reviews performed by members of the ARM DMF and/or DQ Office will be added to the summary report.
2. The summary report is sent to members of the review committee. The review committee makes a decision regarding the allowance of the exception(s).

3. The comments and decisions of the committee members, along with the exceptions for each case, are documented within the appropriate data product EWO for easy searching and access.
4. The data product exceptions for each product are then documented within the relevant reports (e.g., instrument handbook, technical report, web page) such that they are available for external data users.

5.3.1 Size and Members of Review Committee

The review committee should consist of five permanent members that draw from various areas of expertise within the ARM Infrastructure. As such, the committee should include: the ARM VAP manager, the Metadata QC reviewer from the ARM DMF, a representative from the ARM Data Archive, a representative from the ARM DQ Office, and a translator/mentor to represent the scientific community.

5.3.2 Review and Approval of Exceptions

The review committee will review each data product exception on a case-by-case basis using the summary report supplied by the ingest/VAP developers and instrument mentors/VAP translators. The committee will make a decision based on the severity of each exception and a cost/benefit analysis that considers available resources, program priorities, and potential impacts of allowing the exception. At least four committee members must agree for a decision to be finalized.

5.3.3 Documentation of Exceptions

The summary report and decisions of the review committee should be documented in the appropriate EWO. Any approved exceptions to ARM data standards should be included in an initial global DQR (thus communicated upon ordering to all end users) and documented in the appropriate technical documentation (e.g., instrument handbook, technical report and/or web page).

6.0 Summary and Recommendations

The ARM Standards and Policy committee (Appendix C) met via well-attended conference calls over the course of several weeks. Many critical topics relating to policy toward following current ARM standards were respectfully debated. From those discussions a summary of the committee's proposals and recommendations follow:

1. The current set of standards should be reviewed, revised and consolidated in a single location, accessible via the ARM main website.
2. A working group should be formed to accomplish #1.
3. The standards should be divided into core and recommended standards.
4. The PCM tool should be updated to reflect new core and recommended standards.
5. Implications for meeting ARM standards should be documented and made available with the standards.
6. Exceptions to the standards should be kept to an absolute minimum.

7. Exceptions will be handled in the following manner:
 - a. Form a small review committee of five to review exceptions.
 - b. Document exceptions in a clear and consistent manner and communicate to review committee.
 - c. Tools should be created to assist in identifying and documenting exceptions.
 - d. Review team will work on case-by-case basis, using cost-benefit analysis, with four out of five in agreement to finalize decision.
 - e. Exceptions will be documented in general DQR(s) for each data product.
 - f. Exception process:
 - i. Exceptions will be identified via the PCM tool and exported to a summary report
 - ii. Summary report submitted to the review team with description of why the exception(s) were necessary
 - iii. Review team makes its decision and reports via a data product EWO
 - iv. Exceptions are documented so they are available to the end user
8. It is recommended that standards be more flexible, in particular QC standards.
9. It is recommended that efforts be focused on bringing VAPs and recommended datastreams under standards compliance back to the year 2000.
10. It is recommended that a date be set, e.g., January 1, 2013, from which point all new data products in development would adhere to newly revised standards.
11. It is recommended that a path forward be described for bringing ARM data under CF compliance.
12. It is recommended that an alternative to reprocessing be considered that would apply standards “on the fly” during data ordering, as an alternative to total compliance.

Appendix A

ARM Standards Policy Charge Letter

A.1 ARM Standards Policy Charge Letter



21 March 2012

Alice Cialella
Brookhaven National Laboratory
cialella@bnl.gov
631-344-3286

Dear Alice,

Thank you for agreeing to organize and chair a working group to develop a set of policies and best practices around the subject of ARM data and metadata standards that support ARM's mission objectives. Over the past nearly 20 years ARM has developed and evolved standards that apply to ARM data products. The ARM user community has come to depend on the stability and quality of ARM data products in support of their scientific activities. Standards also play a role in ensuring efficient operations and data quality.

The significant diversity of ARM instrumentation, Value Added Product, and Showcase data products makes it difficult to impose a universal approach that makes sense for all ARM data products. The working group that you will be chairing is expected to develop the details behind policy and best practices for applying standards to data products. **The working group is not charged with reviewing or recommending changes to the standards themselves** — if needed that will be a separate activity.

I ask that by March 28 you provide me with suggested names of members of your working group.

I would like an assessment of current state of each area listed below and a set of recommendations for changes or improvements.

- Documentation and communication of ARM data standards. What exists today? Where does it reside (e.g. web link)? What is missing? What activities are needed?
- What standards apply to instrument data (a0, a1, b1)? What is the current

Technical Director's Office
3200 Q Avenue, PO Box 999 MSIN K9-38
Richland, WA 99352

phone: 509-375-2111 | email: www@arm.gov | website: www.arm.gov

Appendix B

List of Current Standards Documentation

B.1 Current List of Standards Documentation

Documentation on the main ARM website:

Location	Contents	Notes
http://www.arm.gov/data/docs/policy	Main documentation page	Right-hand navigation area on this page has links to plans and miscellaneous documents
http://www.arm.gov/data/docs/plan#naming	ARM file naming conventions	
http://www.arm.gov/data/time	Time in netCDF files	
http://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-093.pdf	VAP development guide	
http://www.archive.arm.gov/internal/ARM-Sites/namestd/	Data Archive naming standards	Obsolete in time treatment as YYYYMMDD
http://www.arm.gov/data/docs/doi-guidance	DOI guidance for ARM Facility datastreams	

Documentation on the ARM Engineering website:

Location	Contents	Notes
https://engineering.arm.gov/task/guidelines.php	General ARM guidelines and references	
https://engineering.arm.gov/task/standards-dod.html	DOD standards guide	
https://engineering.arm.gov/~gaustad/qc_standards_need_to_fix/final_doc/update_c_datastream.htm	VAP QC standards	Not publicly accessible, needs to be updated
https://engineering.arm.gov/~gaustad/qc_standards_need_to_fix/final_doc/ingest_requirements.htm	Ingest QC requirements	Not publicly accessible, needs to be updated

The documentation on the wiki ARM page:

<https://wiki.arm.gov/bin/view/Engineering/StandardizingDODs>

Paper describing QC standards for ARM data products:

Gaustad, KL, CJ Flynn, SJ Beus, and BD Ermold. 2010. “The development of QC standards for ARM data products.” In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing*, Chicago, Illinois, USA. <http://dx.doi.org/10.1145/1851476.1851556>.

Appendix C

ARM Standards Policy Committee Members

C.1 ARM Standards Policy Committee Members

Alice Cialella (Brookhaven National Laboratory), chair

Mike Jensen (Brookhaven National Laboratory)

Annette Koontz (Pacific Northwest National Laboratory)

Sally McFarlane (Pacific Northwest National Laboratory)

Renata McCoy (Lawrence Livermore National Laboratory)

Justin Monroe (ARM Data Quality Office, Oklahoma University)

Giri Palanisamy (Oak Ridge National Laboratory)

Robin Perez (Pacific Northwest National Laboratory)

Chitra Sivaraman (Pacific Northwest National Laboratory)



U.S. DEPARTMENT OF
ENERGY

Office of Science