



U.S. DEPARTMENT OF
ENERGY

Office of
Science

DOE/SC-ARM/12-013

ARM User Survey Report: Data Access, Quality, and Delivery

JH Mather
LR Roeder
C Sivaraman

June 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 User Survey Report: Data Access, Quality, and Discovery

JH Mather
LR Roeder
C Sivaraman

June 2012

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

Summary

The objective of this survey was to obtain user feedback to determine how users of the Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive interact with the more than 2000 available types of datastreams. The survey also gathered information about data discovery and data quality.

Although 71 percent of respondents spent 1 month to 6 months or more time with ARM data, 29 percent estimated they spent less than a week or one week to one month using ARM data, suggesting that these users have very little experience with ARM data. Forty-three percent of respondents who answered the survey said they were involved in undergraduate or graduate studies/research, suggesting that a noteworthy number of ARM users are students.

Data and Data Archive interfaces are the most common pathways used by respondents when looking for data on the ARM website. All five pathways are used at least sometimes by more than 70 percent of respondents.

Eighty-one percent of respondents said they always or usually find the data they are looking for and none said they never find it. Some of the comments at the end of the survey expressed frustration locating data suggesting that though respondents find it, they may struggle with locating the information they want.

Respondents said most common reasons for not finding what they are looking for is the data does not exist for a particular site or time period. Only 6 percent of respondents said that data quality was always or usually the reason they do not find the data they are looking for. Data quality issues do not seem to prevent most respondents from finding data they are looking for while data not existing for sites or time periods appear to be more significant factors.

Eighty-one percent of respondents said they are completely or mostly satisfied with the information and tools provided to find data. Users who estimated that they had spent more than six months using ARM data were more likely to be completely or mostly satisfied. Those who spent less than six months are more likely to say they are somewhat, not really, or not at all satisfied. This would suggest experience may improve user satisfaction.

Seventy-three percent of respondents who said they use ARM services available for screening data as part of placing an order said they always, usually or sometimes view data quality reports. No respondents said they were not useful. Twenty-four percent of the total respondents said they were either aware of the reports but had never used them for order screening or were not aware of them. In addition, 77 percent of respondents said they always, usually or sometimes use Data Quality Reports as part of the quality information provided with the data. Data quality including data quality reports were a leading theme within the general comments. These reports appear to be an important and beneficial service provided to users though some users may not be aware of them or may not know how to use them.

All of the suggested additional options were judged very or somewhat useful by more than 80 percent of respondents. Sixty-one percent of respondents said that recommendations for the best source of a particular measurement would be *very useful*, suggesting that this service would be popular with ARM users.

More than half (55 percent) of respondents said they would definitely and 90 percent said they would definitely or probably use data quality information if provided as an integer flag included in a data file. This appears to be the most popular method of delivering data quality information to respondents.

Fifty-eight percent of respondents said that data of poor quality should be withheld from the ARM data file, and 74 percent said it should be provided only upon request. Those who commented on the question of poor quality data said the data may include useful information for some applications, that it should be clearly marked and made available, that they run their own quality control checks and prefer to judge the quality for themselves. Though most respondents said the poor quality data should be withheld from the ARM data file, there may be good reasons to maintain it.

Contents

Summary	iii
1.0 Background.....	2
2.0 Methodology.....	2
3.0 Findings	2
3.1 In what capacity do you use ARM data?.....	2
3.2 What areas of climate science interest you?.....	3
3.3 In the last 12 months, estimate how much time you spent using ARM data.....	3
3.4 What language/software do you use to read and manipulate ARM data?.....	3
3.5 When looking for data on the ARM website, what pathways do you use?.....	4
3.6 How often are you able to find the data you are looking for on the ARM website?.....	5
3.7 Are you satisfied with the information and tools provided to find the data you are looking for?	6
3.8 ARM currently provides several services to help you screen data as part of placing your order. How often do you make use of these services?	6
3.9 What additional options would you find useful before ordering data?	6
3.10 ARM provides data quality information in a variety of ways. These include quality fields in the data files and companion reports. Please indicate which of the following types of quality information you use.....	7
3.11 Would you use data quality information if provided?.....	8
3.12 Should data known to be of poor quality be:.....	9
3.13 If you found a problem with data you ordered would you:.....	9
3.14 Is there something we should have asked that concerns you about data discovery and data quality that was not covered? Would you like to elaborate on any of your answers? Please comment here.....	10
Appendix A.....	A.1

1.0 Background

The objective of this survey was to obtain user feedback to determine how users of the Atmospheric Radiation Measurement (ARM) Climate Research Facility Data Archive interact with the more than 2000 available types of datastreams. The survey also gathered information about data discovery and data quality.

The Market and Competitive Analysis group at Pacific Northwest National Laboratory worked with web administrators to develop a landing page from which users could access the survey. A survey invitation was sent by ARM via email to about 6100 users on February 22, 2012. The invitation was also posted on the ARM website and Facebook page. Reminders were sent via e-mail and posted on Facebook while the survey was open, February 22–March 23, 2012.

2.0 Methodology

The survey size was approximately 6100, and 177 responses were received, for a response rate of approximately 3 percent.

For questions that included a field for users to type their own response, full answers are provided in the appendixes.

3.0 Findings

3.1 In what capacity do you use ARM data?

Fifty-five percent of respondents said they use ARM data for atmospheric observation analysis, followed by 37 percent who said they use it for graduate studies or research, and 35 percent for atmospheric modeling.

Most of the 17 respondents who answered “Other” to this question specified that they use ARM data for ARM infrastructure purposes, including software and data development and testing. For a full list of “Other” responses, see Appendix A.

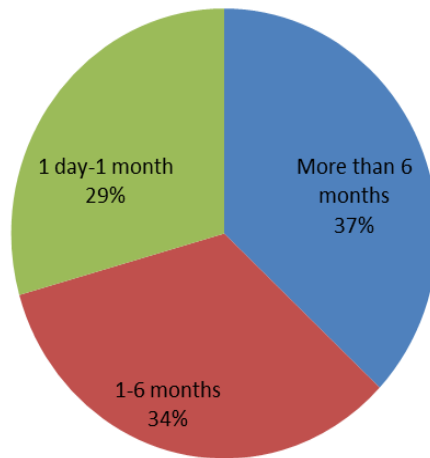
Response	Frequency	Count
Atmospheric observation analysis	55%	98
Graduate studies/research	37%	65
Atmospheric modeling	35%	61
Other climate related research	29%	35
Atmospheric instrument development	12%	22
Other (please specify)	10%	17
Undergraduate studies	5%	9
Instruction	2%	4
Renewable energy systems	2%	4

3.2 What areas of climate science interest you?

Sixty-five percent of respondents said that cloud science interested them, 36 percent said aerosol, and 27 percent said precipitation. Those who answered “Other” provided 16 responses that included the areas of solar radiation, all areas of climate science, and a variety of other topics. For a full list of “Other” responses, see Appendix A.

3.3 In the last 12 months, estimate how much time you spent using ARM data.

Thirty-seven percent of respondents said they spent more than 6 months using ARM data, 34 percent said they spent 1–6 months using it, and 29 percent said they spent 1 day–1 month using the data in the last 12 months.



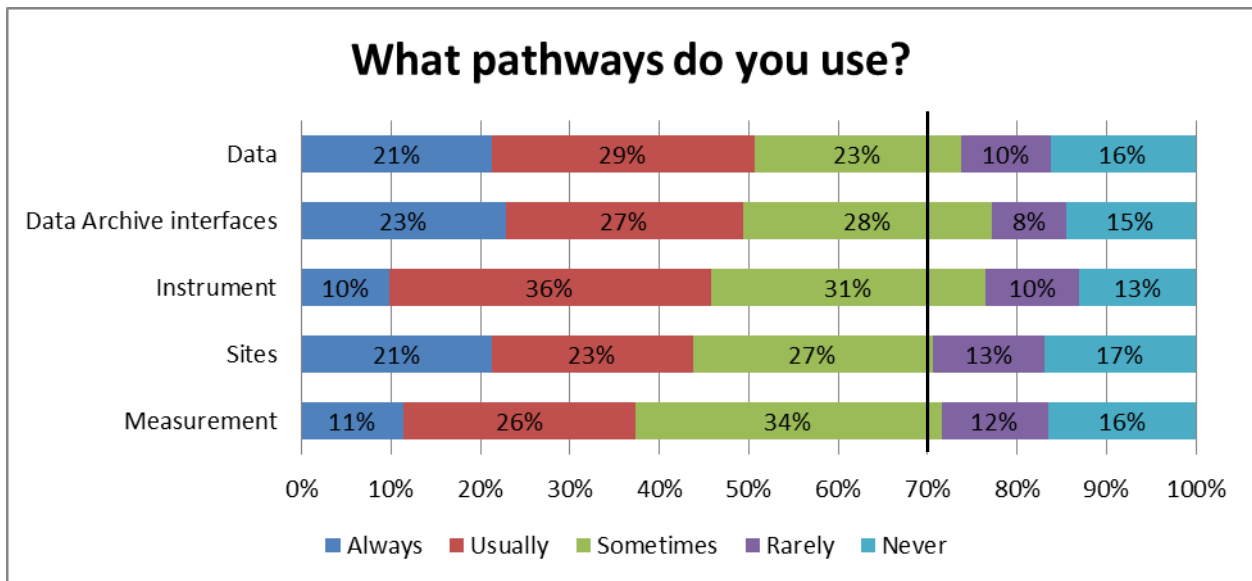
3.4 What language/software do you use to read and manipulate ARM data?

Forty-five percent of respondents said they use IDL to read and manipulate ARM data, 38 percent said they use Matlab, and 35 percent said they use FORTRAN. 29 respondents answered “Other” and most frequently specified NCL as the language or software they use. For a full list of “Other” responses, see Appendix A.

Response	Frequency	Count
IDL	45%	79
Matlab	38%	67
Fortran	35%	61
C	16%	29
Spreadsheet (Excel, etc.)	16%	29
Other (please specify)	16%	29
Python	10%	18
Not Applicable	3%	5

3.5 When looking for data on the ARM website, what pathways do you use?

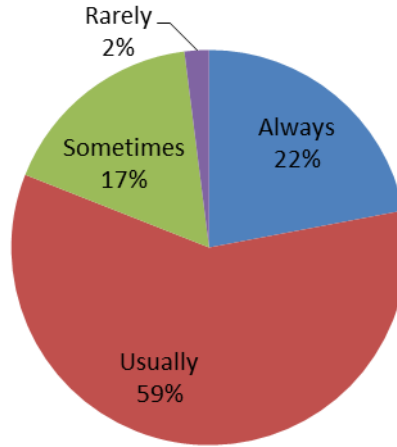
- Fifty-one percent said they always or usually use the Data pathway, 23 percent sometimes use it, and 26 percent rarely or never use it.
- Forty-nine percent said they always or usually use the Data Archive interfaces pathway, 28 percent sometimes use it, and 23 percent rarely or never use it.
- Forty-six percent said they always or usually use the Instrument pathway, 31 percent sometimes use it, and 24 percent rarely or never use it.
- Forty-four percent said they always or usually use the Sites pathway, 27 percent said they sometimes use it, and 29 percent said they rarely or never use it.
- Thirty-seven percent said they always or usually use the Measurement pathway, 34 percent sometimes use it, and 28 percent said they rarely or never use it.



More than 70 percent said they always, usually or sometimes use all five of the pathways.

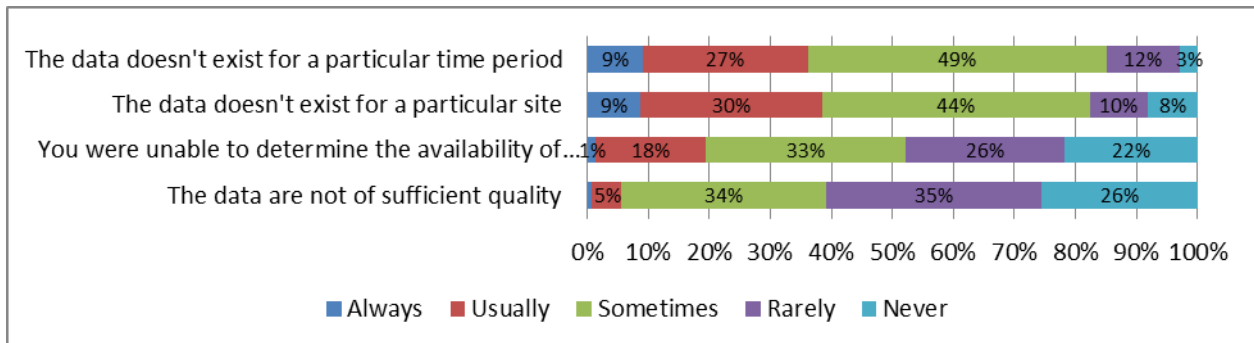
3.6 How often are you able to find the data you are looking for on the ARM website?

Eighty-one percent of respondents said they always or usually find the data they are looking for on the ARM website, 17 percent said sometimes, and 2 percent said rarely. No respondents said they never find the data they are looking for on the website.



When you do not find the data you are looking for is it because:

- Thirty-six percent said they always or usually are unable to find the data they are looking for because it does not exist for a particular time period, 49 percent said sometimes, and 15 percent said rarely or never.
- Thirty-nine percent said they are always or usually unable to find the data they are looking for because it does not exist for a particular site, 44 percent said sometimes, and 18 percent said always or usually.
- Twenty percent said they always or usually do not find the data they are looking for because they were unable to determine its availability, 33 percent said sometimes, and 48 percent said rarely or never.
- Six percent said they always or usually are they unable to find data they are looking for because it is not of sufficient quality, 34 percent said sometimes, 35 percent said rarely, and 26 percent said never.

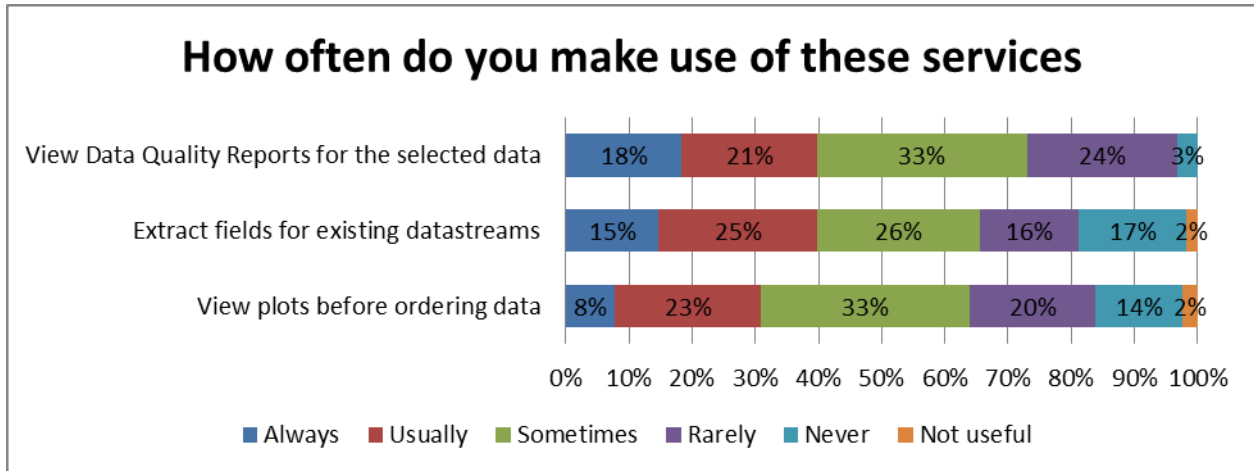


3.7 Are you satisfied with the information and tools provided to find the data you are looking for?

Eighty-two percent said they were completely or mostly satisfied with the information and tools provided to find data, 14 percent said they are somewhat satisfied, and 4 percent said they were not really or not at all satisfied.

3.8 ARM currently provides several services to help you screen data as part of placing your order. How often do you make use of these services?

- Forty percent of respondent said they always or usually view data quality reports for selected data, 33 percent said sometimes and 27 percent rarely or never. No respondents said they were not useful.
- Forty percent of respondents said they always or usually extract fields for existing datastreams, 26 percent said sometimes, and 34 percent said rarely, never, or not useful.
- Thirty-one percent said they always or usually view plots before ordering data, 33 percent said sometimes, and 36 percent said rarely, never, or not useful.



More than 60 percent said they always, usually, or sometimes make use of all three services.

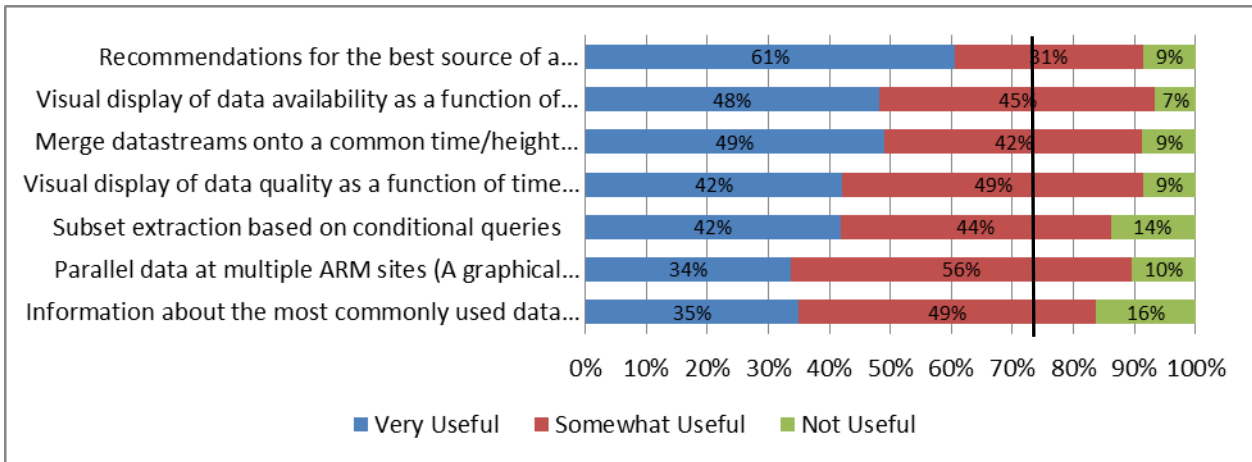
*A significant percentage of all respondents said they either were aware of a service but had not tried it or that they weren't aware it existed.

3.9 What additional options would you find useful before ordering data?

- Ninety-one percent of respondents said recommendations for the best source of a particular measurement would be very or somewhat useful, 9 percent said it would not be useful
- Ninety-three percent said visual display of data availability as a function of time for multiple datastreams would be very or somewhat useful, 7 percent said it would not be useful
- Ninety-one percent said merge datastreams onto a common time/height grid would be very or somewhat useful, 9 percent said it would not be useful

- Ninety-two percent said visual display of data quality as a function of time for multiple datastreams would be very or somewhat useful, 9 percent said it would not be useful
- Eighty-six percent said subset extraction based on conditional queries would be very or somewhat useful, 14 percent said it would not be useful
- Ninety percent said a graphical display that shows data availability at multiple ARM sites would be very or somewhat useful and 10 percent said it would not be useful
- Eighty-four percent said information about the most commonly used data sets would be very or somewhat useful, 16 percent said it would not be useful

All of the options would be very or somewhat useful to more than 80 percent of respondents

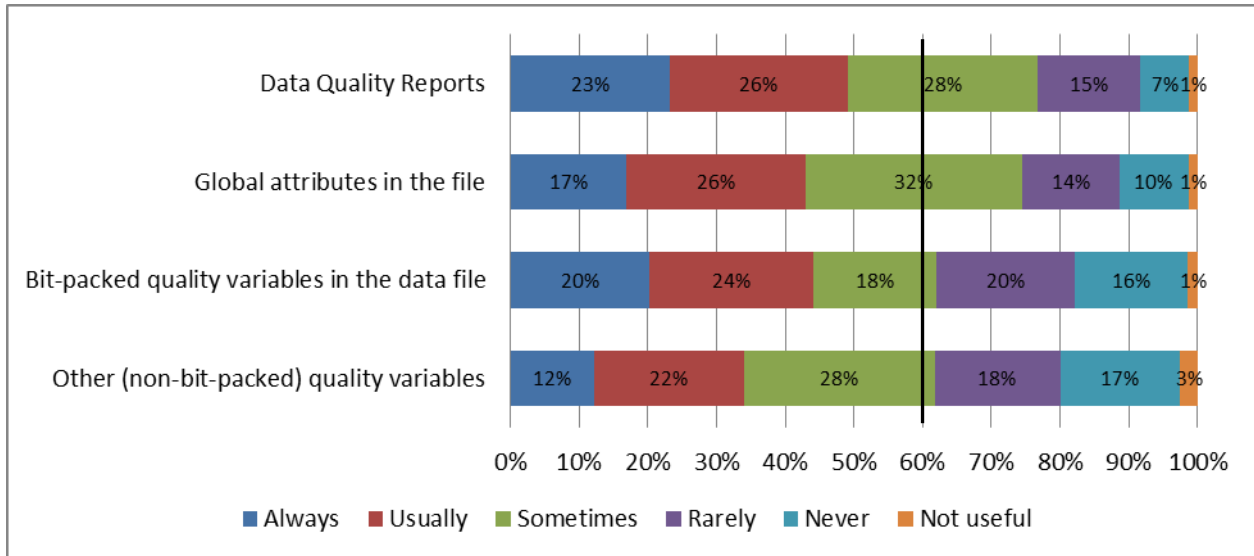


Respondents provided 14 comments to the question “Are there other pre-order options for data not listed in question 10 that you would find useful? For a full list of responses, see “Appendix A.”

3.10 ARM provides data quality information in a variety of ways. These include quality fields in the data files and companion reports. Please indicate which of the following types of quality information you use.

- Forty-nine percent of respondents said they always or usually use data quality reports, 28 percent said they sometimes use them, and 23 percent said they use data quality reports rarely or never, or that they are not useful.
- Forty-three percent said they always or usually use global attributed in the file, 32 percent said they sometimes use them, and 26 percent said they rarely or never use them or that they are not useful.
- Forty-four percent said they always or usually use bit-packed quality variables in the data file, 18 percent said they sometimes use them, and 38 percent said they rarely or never use them or that they are not useful.
- Thirty-four percent said they always or usually use other non-bit-packed quality variables, 28 percent said they sometimes use them, and 38 percent said they rarely or never use them or they are not useful.

****The chart below and the results reported above were calculated omitting responses for “Don’t know what this is” answers.****

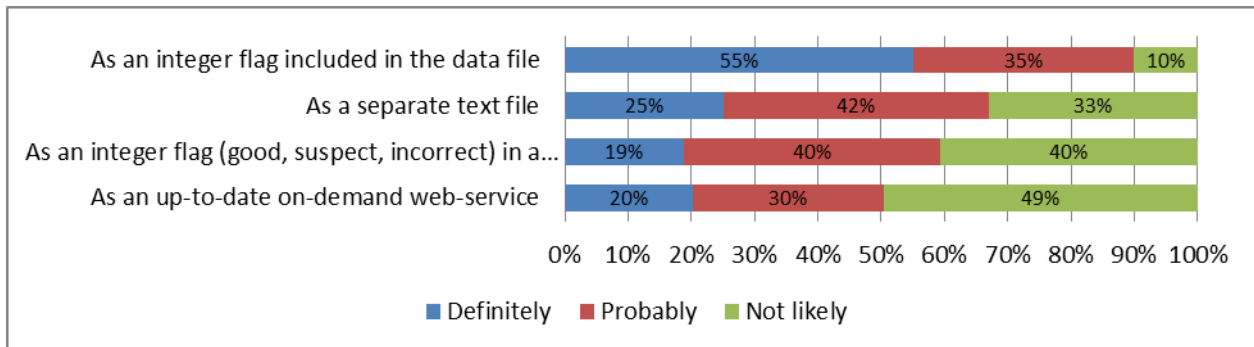


More than 60 always, usually or sometimes use all four types of quality information.

Twenty-seven percent of respondents said they don’t know what other non-bit-packed quality variables are.

3.11 Would you use data quality information if provided?

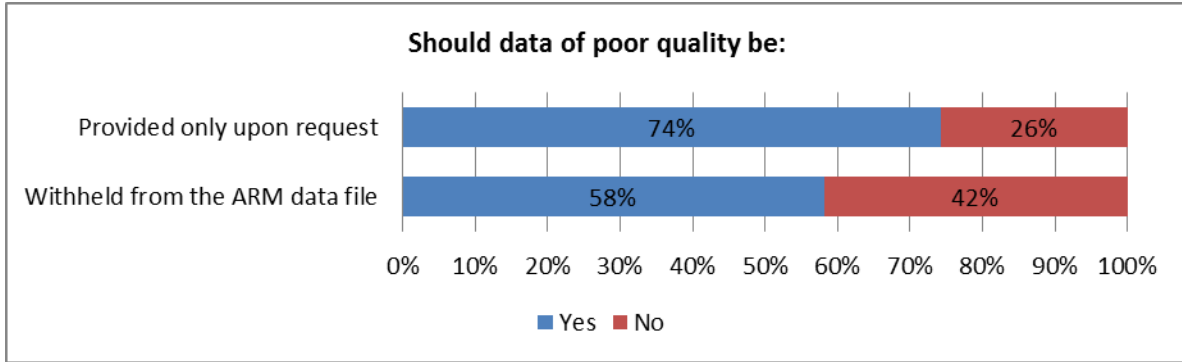
- Ninety percent of respondents said they would definitely or probably use data quality information if provided as an integer flag included in the data file and 10 percent said not likely.
- Sixty-seven percent said they would definitely or probably use data quality information if provided as a separate text file, and 33 percent said not likely.
- Fifty-nine percent said they would definitely or probably use data quality information if provided as an integer flag (good, suspect, incorrect) in a companion file, 40 percent said not likely
- Fifty percent said they would definitely or probably use data quality information if provided as an up-to-date on-demand web-service, and 49 percent said not likely.



Some respondents said they “Don’t Know” if they would use data quality information if provided.

3.12 Should data known to be of poor quality be:

- Seventy-four percent said data known to be of poor quality should be provided only upon request.
- Fifty-eight percent said data known to be of poor quality should be withheld from the ARM data file.



3.12.1 If you answered no to either part of question 13, please explain

Respondents provided 56 explanations for their answers to question 13. Several themes emerged from the comments.

- Many respondents said that poor quality data may contain useful information or said they want to have access to the data and judge its usefulness for themselves.
- Some respondents said data should be available as long as poor quality data is clearly marked or improved data quality information is included with it.
- Some respondents write their own filters and code to assess data quality or perform their own quality control checks. For this reason they would like poor quality data to be available.
- Some wanted a definition for “data of poor quality.”

For a full list of responses, see Appendix A.

3.13 If you found a problem with data you ordered would you:

- Fifty-four percent said they would contact an instrument mentor if they found a problem with data they ordered as did one respondent who marked “Other.”
- Forty-nine percent said they would contact the archive for data problems.

For a full list of “Other” responses, see Appendix A.

Response	Frequency	Count
Contact an instrument mentor (technical contact)	54%	95
Contact the Data Archive	49%	87
Contact the Data Quality Office	27%	48
Pass the information on to a colleague	21%	37
Don't know what I would do	12%	22
Other (please specify)	6%	10

3.14 Is there something we should have asked that concerns you about data discovery and data quality that was not covered? Would you like to elaborate on any of your answers? Please comment here.

Forty-two respondents provided additional comments and several themes emerged from them.

- Access to data is a problem either because the website is difficult to navigate, information is hard to find, or because it isn't available in a timely manner
- Data quality and quality control improvement suggestions were provided
- Kudos for the ARM Facility, website, and staff
- Instrument handbooks lack pertinent data or are hard to find
- Ordering data is a complicated or slow process.

For a full list of responses, see Appendix A.

Appendix A

A.1 In what capacity do you use ARM data? (“Other” Comments)

Response: In what capacity do you use ARM data? (Other, please specify.)		
ARM Infrastructure (7)		
ARM Communications/communications strategies (2)		
ARM Infrastructure development and testing		
Making quicklook plots at ARM Data Archive		
Metadata information to interagency and international groups		
Satellite algorithm validation		
VAP and infrastructure software development		
Data development		
Infrastructure data system		
Soil Moisture Database		
	Responses	17

A.2 What areas of climate science interest you? (“Other” Comments)

Response: What areas of climate science interest you? (Other, please specify)		
All (2)		
Geoengineering-Solar Radiation Management		
Attribution		
Data quality of ARM data products		
Turbulence, planetary boundary layer		
Solar radiation		
Statistician, developing statistical spatial-temporal models, using ARM SGP temperature data to motivate the models.		
Radiation		
Energy balance of climate system		
Evaporation		
Diurnal cycle of near surface variables		
Plant ecology		

Response: What areas of climate science interest you? (Other, please specify)		
Radiation, radiative closure studies		
Boundary-layer dynamics		
Impacts on mountain climates		
	Responses	16

A.3 What language/software do you use to read and manipulate ARM data? (“Other” Comments)

Response: What language/software do you use to read and manipulate ARM data? (Other, please specify)		
NCL (6)		
Perl (4)		
R (3)		
C++ (2)		
Java (2)		
Igor Pro (2)		
GrADS (2)		
Cygwin, C#		
NC tools		
NetCDF and Unix utilities		
Unix Shell Scripts		
Perl, NCL		
ncbrowse		
Converted data using NetCDF DUMP		
GRIB utilities (GrADS), NCO, NCL		
	Responses	29

A.4 Are these other pre-order options for data not listed in question 10 that you would find useful?

Response: Are there other pre-order options for data not listed in question 10 that you would find useful?		
No, not at this time.		
A visual (or other display) of data quality as a function of time would be the most helpful. TO BE USEFUL, however this MUST entail some information about what level of data quality assessment has occurred. I expand on this point, below.		
Crowd-sourced tags, comments, related data sets, think "Amazon": Users that looked for this measurement ultimately downloaded this "datastream". Don't just show "most commonly used data sets" but give percentage distribution. Show publications' use of datastreams.		
Date that data were stored at Data Archive.		
The option to download data in a compressed file format.		
Option to get data in ASCII format with readme.		
The RUC analysis grids in grib format require an email to be sent to acquire. While the email method works great and is timely, a web interface would be even better.		
A list of whether the measurements are not *yet* available for a time period or whether they have been discontinued. If they will be available at some point, an estimate for when the data will be available would be helpful.		
For the "Subset extraction based on conditional queries" question, it would be quite useful for several instruments to allow for selection of data on 'cloudy' days. I currently look at TSI or VCEIL to determine when downloading radar data would be useful.		
The biggest problem I have right now is figuring out which radar or lidar is installed where (there is conflicting information on different ARM website pages) and which ones are working or not working. Going to the archive and not finding data for a particular instrument is not sufficient - one always wonders whether one is just looking in the wrong place. What I'd like is a page, not in the archive but on the ARM website, that lists not only all the instruments at each site but whether they are operating. Two examples: Various AMIE-Manus documents say that there will be a C-POL radar at Manus. But at the Manus instrument site http://www.arm.gov/sites/twp/C1/instruments there is no mention of a C-POL. Is there one there, or will there be one there in the future? When will we have data from it? Example 2: A potential highlight of AMIE-Gan was going to be the deployment of X-band radars there. But when I go to the AMIE website and link to instruments on that page, I can find the X-SACR, and then of I click on place a data order I get to another page that indicates that more or less data are not available. Is that because the data are not in the archive yet or because the instrument did not work? I had to go to the AMIE session at the Science Team meeting to find out it was the latter.		
Not that I can think of.		
None that come to mind.		
1) temporal resolution of the data set. 2) data object designs - are mostly easily found difficult to find in rare cases - a very discreet and standard link to these near information for the datastream would be useful.		
If you can provide raw data also, that will be very useful. For example, when I am looking for Cloud Radar Raw data (spectral and I & Q data) are not available in any site. I am requesting you to provide those data also on requested by users. This will helps to me to analyze those data files by myself.		
	Responses	14

A.5 If you answered no to either part of question 13, please explain.

Response: If you answered no to either part of question 13, please explain
I write my own code to filter data. (4)
Would like to have the option to write own filters.
Users should be made aware of potentially poor quality data, but still have opportunity to use data since many users such as myself write our own code to check for bad data. This is valid as long as there is some data quality information with the file.
If other expert research groups who may not directly associated with the ARM data processing may find it useful and extract something if the data is very crucial unless storing redundant data is an issue.
I sometimes find that I can use data that has DQR issues. I would argue strongly for keeping the data available, but building/creating much improved data quality information.
I can't evaluate data for my specific use if I am not allowed to see it at all. Requiring a special request is, however, a VERY GOOD warning about quality issues.
There are no simple answers to these questions. What is 'poor' quality?
If it's flagged it's easy enough to exclude in any analysis, but having the unaltered data even if bad, might still contain some useful information, e.g. timing of events, when calibration is completely off or signal very noisy.
Because, I would like to know what those specific data problems are and expect ARM to provide the quality control information.
Gaps in the data record can make it that much harder to match e.g. model output to observations. I much prefer a continuous data record (without gaps), but with missing values/quality flags to screen out poor quality observations.
You do not define 'poor quality'. Quantify this term.
Having all data is valuable for some projects. However, the file containing the data needs to have the quality flags in the same file to easily identify which data is believable at a given threshold of usability.
As long as the source of the poor quality is documented I can deal with it or choose to reject it...
Data may be of poor quality for a variety of reasons. If you knowingly request poor quality data it is likely that the issue does not involve the quantity you are interested in.
I would like to have the option to examine the bad data to be sure it is 'bad' for my purposes.
The data should be flagged as poor quality in the file, but provided so the user can decide whether or not they use the value. If users want it withheld that should be a user decision made and implemented at the time the data is ordered.
I have enough knowledge to know what level of filter the data that can be used.
The best possible path forward would be to provide the data and then provide easily read data quality flags from within the same file to serve as filters of bad data. Sometimes the quality is not a binary good/bad, in which case how would you decide what is not to be shown? Somehow the program needs to deal with flagging data that gradually transition from good to bad. This is no easy thing to do.
I think the user should always have the option to retrieve the original unfiltered data.
As long as there is an indication of the data quality, poor quality data should still be available, and users can filter bad data.
I filter on the qc_ fields, but it would be more convenient if the bad data weren't in the file to begin with.

Response: If you answered no to either part of question 13, please explain
I write my own code to do quality checking/filtering using the quality flags/variables in the files. If there are composite variables in the file (variables derived from other variables), it is helpful to trace back through the underlying variables/qc flags to see why certain data were rejected.
This would prevent people from using poor quality data without knowing the difficulties of going into that data beforehand.
I write my own code to filter data. Depending on the application, some data may be useful even though it is of poor quality.
I always do my own data quality checking.
Better to provide more information (like an integer flag and an explanation field) than to have users wondering why data are missing. Placeholders for bad data would be acceptable.
I have sufficient expertise to identify bad data in many datastreams. Sometimes it is important to know that the instrument collected bad data during certain conditions.
Poor quality data can sometimes provide very important information especially when there is no other data to provide information. When used carefully and wisely poor data is much better than no data.
Mostly ideological! I would prefer to have as much data as possible, then choose how/if I filter it.
Poor quality data are still better than no data at all.
Usually you can tell if the data are missing. I probably should be reading the quality reports to see if there was a problem with the data I'm using.
In the spirit of "public" data, even known poor data should be available on request, with all the appropriate warnings. I do use my own filters, and sometimes what you consider poor is "good enough" for my purposes.
Data quality flags should be good enough for the user to filter bad data. In reality they do not seem to be. For example, filtering by the quality flag and wet window flag for MWR, I find that there are still plenty of bad data there. If there is another flag I should be checking, I missed it. Having a list of all the quality flags and an explanation of them all together in one place in the README would help.
I regard neglecting any form of data, whether it be of good or bad quality, as a loss of information, information that can be used to build/tune algorithms etc.
I do most of my own QC, as I believe that for good science I truly understand the data I use. However, I will check the results of my QC against ARM provided QC and DQRs to see if I find the same issues they did (usually I find more).
Noting that the data are of poor quality or using data flags should be sufficient.
Even the data with poor quality can sometimes give valuable information.
I write my own code to analyze. Poor quality data should be flagged.
Sometimes, poor quality of data may be due to a signal that may be useful for a different application. For example, insects may make the KAZR data of poor quality for cloud estimation, but may be good for estimating boundary layer depth (assuming the insects stay within the boundary layer).
Flag it; caveat emptor; flag as column in data file.
Sometimes poor data in one sensor yield helpful information about the state of other sensors taking data simultaneously, so don't zap it, but do warn people not to use it.
Just make sure user is aware of data quality.
Data should always be kept visible, because it is important to know that they exist and may be used only for qualitative purpose.

Response: If you answered no to either part of question 13, please explain		
All data should be kept because we never know in advance for what they could be used in the future.		
Because a few quality checks might be too restrictive for certain type of investigation and it might be better to let the users free to select their data also considering the quality check implemented by ARM.		
There may still be useful information in data which some considered of poor quality.		
Poor quality is not defined!!!		
Generally filtering of data should be an option. Filtering out the bad data can be the default option, but users should know that the data have been filtered and can be provided by request or by choosing a different option. Data labeled as "poor" data may still be useful, for example, to diagnose problems with an instrument.		
Sometimes it's good to have bad data so you can make sure your algorithms for detecting bad are robust.		
No, there are uses for poor quality data if the individual can perform adequate QC.		
Data showed should always be available to be ordered. Yes, warn me about problems, but I want to judge for myself.		
So far I have not filtered data. Whatever data I have downloaded from ARM, that itself good quality data based upon my experience.		
	Responses	56

A.6 If you found a problem with data you ordered would you: ("Other" Comments)

Response: If you found a problem with data you ordered would you: (Other, please specify)		
File a DQPR.		
Depends on situation.		
Contact the DMF.		
Contact the individual/contact listed in the file.		
I'd contact the instrument mentor if I knew who they were, of course!		
Probably just deal with it myself for my specific purpose.		
Move on.		
Look for a DQR & if there isn't one, file a PIF (or is that a PIR?)		
Send an email to the address that is easiest to find.		
Contact a translator or developer.		
	Responses	10

A.7 Is there something we should have asked that concerns you about data discovery and data quality that was not covered? Would you like to elaborate on any of your answers? Please comment here.

Response: Please comment here.
No, not at this time.
Sometimes you aren't very quick to provide data. The data are available in your FTP for a very short period. Some data should be available directly from the web page. You should have more actualized data of some measurements or VAPs (I mean, for some data you lost a lot of time to put it available in the archive).
I don't use the data for scientific purposes, only for infrastructure purposes. My answers should be downweighted accordingly.
The largest issues with ARM data are all about data quality. Broadly I see three separate issues (1) assessing data quality, (2) providing data quality info to users, and (3) ensure ongoing measurements are of high quality. There are problems in all three of these areas!! I often find problems in the data that are without any associated DQR. Often this is because there has been little quality control attempted (beyond min/max checks, which are a pretty sad level of effort). In my option, every data product should have identified "key" variables, and these variables should be carefully AND manually checked by the instrument mentors. These key variables should have IN FILE data quality information that (A) gives some general rating, (B) information pointing to document describing what QA was done and what "nominal" data quality means (including uncertainty values!), (C) if rated less than "nominal" points to a DQR/text description of the problem, (D) indicates WHO "approved" the data. Some files do have "IN FILE" QA flags. But I often find problems not identified by these flags -- they are currently not very good. Variables that are not "key" could be subject to simpler automated checking ... more of a buyer beware condition. Broadly speaking the goal of the QA effort should be to identify when KEY data are good -- not track "known problem" -- but find problems!! Vigorous efforts to do the above would (I hope) also be coupled with better efforts to ensure the measurements are of good quality (with problems more quickly identified and fixed).
One thing that comes to mind right away is that I have had some trouble deciphering mentor reports. Specifically, I had difficulty determining which site the comments were for. Sometimes, more than one site may be referenced in the same report, and I found this confusing. The other item that comes to mind is that I have had difficulties due to missing, but pertinent information, in instrument handbooks. Those are the two most frustrating issues I've encountered. Other than that, the ARM Data Archive works well! Thanks!!
I think it would be useful to provide a rough timeline for updated versions of data. For example, if a group is working on flagging bad data for a particular data set, is there a time frame when we can expect this to be available?
None.
What about field campaign data? There are so many campaigns each with a scientific goal in mind, but they may contain good cases for other studies as well? But I find it hard to find the right campaigns with the right combination of instruments? Could campaigns be categorized by study objectives, and maybe actual study results could be linked in through the publications.
I use mostly IOP data, which may not be well addressed by this survey. For instance, IOP data quality has been harder to assess in some situations. When I did encounter a major data quality problem, the PI was no longer available and I was somehow discouraged from reporting the problem (I can't recall the details, but it brought to my attention that data quality in IOP data is a special issue that may need special attention).
ARM is an impressive program as exemplified by the request for this type of response from past users. To be noted, I do not use any data myself. But, this information is important for me to understand ARM as a program and/or institutional entity.

<p>Response: Please comment here.</p>
<p>I've spent quite some time on occasion trying to find out what the measurement/retrieval uncertainty is for an observable. Particularly for some of the higher-order products (e.g. CMBE), I've had to dig down into the documentation of the lower-level products incorporated into the higher-level product to find answers. I would much appreciate it if measurement/retrieval uncertainty would be included in an obvious, easy-to-find place. For example, as an additional attribute in the netCDF file, or in a single text file associated with the product. I know this is already the case for some products/observables, but not for all.</p>
<p>ARM data files, and the website, are still designed mostly for the expert user and not for the novice. Two quick examples: (1) MWR: For this instrument there is at least an online handbook. But when one goes to it, one struggles to figure out how best to screen the data for bad data. The novice user might guess that the data quality flag would take care of that. But it does not. One has to at least also look at the wet window flag. And in the document, the wet window flag is mentioned several pages earlier than the discussion of why the wet window flag is important and why you might want to look out for it. This type of thing drives the user crazy. (2) ARM wants people to use the new ARRA instruments. If so, then they need to give people clues about what they are measuring, why you might want to use one over the other, etc. So for example when one goes to the KAZR page one finds no handbook like that for MWR. One also finds a lot of data products (kazrbl, kazrci, kazrge, which at least maybe the user can guess about), but also the cross-polarized and co-polarized measurements, which the typical user will not have a clue about. Why are these different things provided? How do they differ? Why might I want to use one vs. the other? One might expect that when one clicks the hypertext "kazrge" there would be an explanation of what kazrge means, but there is not. And so on for the other products. Worse yet, there is little or no discussion about how one might choose to use a Ka-band scanning radar over an X-band scanning radar at the same site. One would think that by now, there would be a simple English discussion of the different instruments, a comparison of the strengths and weaknesses of each, and some guidance on which instrument might be better for which purpose.</p>
<p>Several of the questions asked about including data quality in a separate file. This would add a lot of difficulty in scripting the processing of the data, particularly if one had to go to a website for an on-demand service. Also, the data quality flags would be more useful if they were better described and consistently applied across time. I recently used the full time series of radiosondes from SGP and it appears the way data is quality controlled changes over time. This made processing the data difficult. If a methodology is changed, it should be retroactively applied to the older data as much as possible.</p>
<p>Thank you for the survey.</p>
<p>I sometimes have to login to the 'ui1.ornl.arm.gov' site in order to order raw data or to order data for several isolated dates. The latter reason, especially, is something the Data Archive might consider making easier to do. If one uses the Data Archive Interface, it requires separate orders to acquire, for example, individual days of data scattered over several years or months without having to scroll through large numbers of files to order just a few files of interest. I find bit-packed data quality variables to be very off-putting. I far prefer any other form of data quality fields, as long as they are self-describing.</p>
<p>I use the publications database extensively, and two items are frustrating: Multiple entries for the same record, and inconsistent publication names.</p>
<p>I would like an interface to order data simply by using the full datastream name and date ranges all in one interface on one web page.</p>
<p>I found it difficult sometimes to navigate and find the data I was looking for, but I have to admit I did not try this extensively so it might just be that I need to get some more experience with the system.</p>
<p>I think the emphasis above in providing high quality data flags that are easily accessed and interpreted by a data user is a good idea. The problem becomes implementing such a scheme for data with gradual quality transitions. How does one flag such data? I am not sure there is an easy answer to this question.</p>

Response: Please comment here.		
It seems some data are available but do not show in the website, and there is no way to know beforehand unless we contact the mentor, which may not be convenient always.		
I think that the ease with which one may obtain ARM data is rather a miracle.		
Data are not always in obvious places. IAP data for example and sometimes data for field campaigns.		
Yes.		
Your service is great, and all of your staff are extremely responsive and helpful! Thank you.		
I really appreciate that these data are easily and publicly available. However, I find the interface difficult to navigate and find what I want. I typically look for airborne or tower measurements of CO2, CO and CH4 over the SGP. I tend to give up after a while and email the mentor directly.		
This isn't a data quality issue, per se, but it would be useful for the netCDF files to all have some common global attribute fields, and for these fields to be documented.		
Please replace the ARM ftp data server. It is horribly slow and sometimes doesn't work properly.		
SMAPEX.		
No. But thank you!		
Do not use -9999. Use NaN for missing data. This is IEEE standard. I hate it when I get a data file and make a graph and then the scale goes all the way to -9999. That is so 1970's. Or even better, let this be a user option: missing data [radio button] -9999 [radio button] NaN. I applaud you going to allow user to specify file type (ascii, netcdf). Easier for you to write script once than 1000 users each writing their own scripts. Igor now reads netCDF, but for a mac person like me it was a royal pain to deal with netcdf. Now with igor converter it's fun.		
Can you get the NEON people to do as good a job as ARM? Can you get the Europeans to let us use their data? Can you get the NWS to archive all their soundings with you? :@)		
People new to ARM usually find it challenging to discover the data they are looking for. I have bookmarked the relevant links, which works for me.		
No.		
Not yet.		
I use the full model output fields generated by the RUC atmospheric model (on the native RUC grid) and always have to contact someone at ARM to get access to this data.		
Organizing the data in terms of temporal and geographical availability would be very useful—the current method requires going to the 'datastream' link and finding links to sites there, then you have to go to the Data Archive to find out the temporal availability—this is a process—would also be helpful to see what measurements are made consistently across sites in the program.		
Surveys provides rather poor measures, I don't think you should be using them as a principal or even secondary way to decide on what you are going to do. Rather, you should only use them to get additional ideas ... which doesn't seem to be the focus of this survey.		
Radar calibration is a major concern of mine. It is known that the calibration for multiple ARM radars has drifted in time, and this is a major issue when attempting to analyze long-term data sets. It would be good if the calibration for all radar measurement was reviewed and all data files corrected according to the best knowledge of the calibration history.		
I have already mentioned in Question #10.		
	Responses	42



U.S. DEPARTMENT OF
ENERGY

Office of Science