

# ARM Evaluation Data User Metrics and Best Practices

Rachael Isphording<sup>1</sup>, Maggie Davis<sup>1</sup>, Ric Cederwall<sup>1</sup>, Alka Singh<sup>1</sup>  
<sup>1</sup>Arm Data Center (ADC), Oak Ridge National Laboratory



## Abstract

Evaluation data are an important component of scientific advancement featured at ARM Data Discovery. We currently store over 16 TB of archived evaluation data accessible to the ASR and Earth Science communities. Traditionally, this data has been stored separately from traditional ARM data and ordered differently, through the "IOP" portal along with Principal Investigator and Field Campaign data. ARM standard naming conventions or file formatting were not required. Data could therefore be more difficult for the user community to utilize and evaluate, as well as complicate any future transition to baseline ARM data.

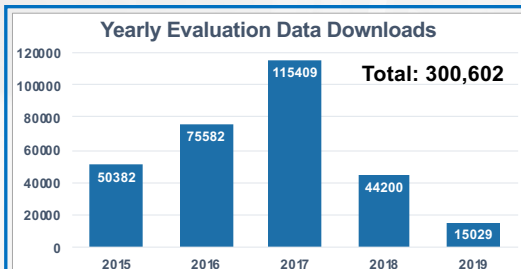
### Purpose:

- provide insight into the overall impact of ARM evaluation data to help inform scientific priorities.
- provide to ASR and ARM investigators best practices for using archived evaluation data including mechanisms for providing feedback on the utility, maturity, or quality of the data.

To help guide the community in developing new and effective evaluation data products, we provide:

- 1) basic user statistics (e.g. total number of downloads, frequency of downloads, etc.)
- 2) characteristics of "most used" evaluation datasets (data format, spatial dependencies, etc.)
- 3) suggestions for creating quality and impactful evaluation datasets.

## User Metrics



Annual data downloads from the evaluation area for the years 2015-2019 (excludes README.html documentation).

ARM naming convention:  
(sss)(instrument-code)(Fn).(dl).(yyyymmdd).(hhmmss).nc

43.3% of the evaluation data downloads from 2015-2019 met ARM Data File Naming Standards.

## Limitations

A lack of historical ARM records documenting evaluation data that have successfully transitioned into routine products limits the ability to provide informative suggestions for creating new and effective evaluation products.

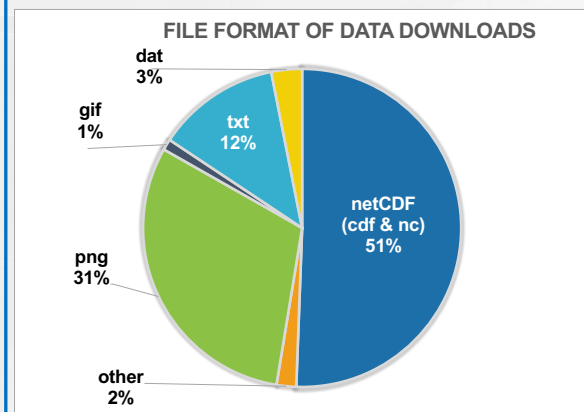
### Remaining questions:

- 1) Is there a correlation to how long data remains in an evaluation state and the impact/usage of the data in the ASR community once it has migrated to a routine/baseline data status?
- 2) Should there be a designated time period in which evaluation products are hosted and receive user feedback before transitioning into routine data or being revisited by the developer?
- 3) How best to solicit feedback on evaluation data to assist developers?

## Data Download Characteristics

Between the years of 2015-2019, 51% of the evaluation data downloads were in netCDF file format. This percentage is based on the 300,602 data downloads (does not include the README.html download records). The "other" category includes 38 different file formats such as: PDF, CSV, ASC, TXT, JPEG, and DOC. NetCDF files that were zipped together in the downloads as "GZ" files were considered netCDF format for this study.

Based on these results, we can infer that users are more inclined to utilize ARM evaluation data if it is submitted as netCDF or PNG data files.



Based on the total data downloads for 2015-2019 (300,602 downloads), the chart above summarizes the percentage of each data file format. The "other" category contains 38 different file formats.

Furthermore, the majority of evaluation data downloads were datasets geographically associated with the Southern Great Plains ARM Site, closely followed by the Eastern North Atlantic and North Slope Alaska sites.

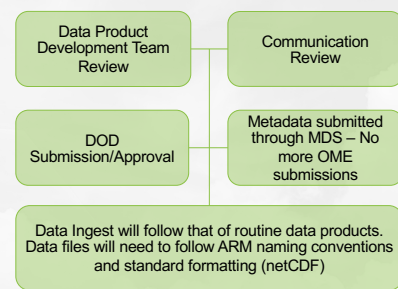
Based on the assessment of evaluation data downloads, a couple of points can be highlighted about the "most used" evaluation datasets:

- Evaluation products that are submitted in netCDF or PNG format
- File naming conventions follow that of ARM standards
- Evaluation data products that are based out of one of the 3 ARM fixed-location atmospheric observatories

As ARM transitions to an updated workflow to process and store evaluation data, 2/3 of these characteristics will become mandatory for all evaluation data products. This will hopefully increase the usability and popularity of all evaluation data products.

## Future Requirements

ADC is currently undergoing a large effort to transition the evaluation data out of the IOP area and into the data archive area. Future evaluation products will now follow a workflow similar to that of routine ARM data products. New requirements for processing evaluation data include:



## Contact & Links

<b>Metadata team:</b> adcmetadata@arm.gov	<b>ARM IOP Area:</b> <a href="https://iop.archive.arm.gov/arm-iop/">https://iop.archive.arm.gov/arm-iop/</a>
<b>Useful links</b> <b>ARM Data File Standards:</b> <a href="https://www.arm.gov/publications/programdocs/doe-sc-arm-15-004.pdf">https://www.arm.gov/publications/programdocs/doe-sc-arm-15-004.pdf</a>	<b>ARM Data Discovery Tool:</b> <a href="http://www.archive.arm.gov/discovery/">http://www.archive.arm.gov/discovery/</a>