

ARM

CLIMATE RESEARCH FACILITY

ARM Data Integration Group's Roles and Responsibilities

Chitra Sivaraman, Sherman Beus, Brian Ermold, Todd Hull, Krista Gaustad, Nicole Keck, Annette Koontz, Matt Macduff, Tonya Martin, Robin Perez, Yan Shi, Tim Shippert, and Katarina Younkin

PACIFIC NORTHWEST NATIONAL LABORATORY



CONTRIBUTIONS

DATA PROCESSING

- **Collections and Ingests:** Raw data are aggregated into daily files and processed into netCDF files.
- **Reviews:** Datastream reviews ensure consistency and compliance with ARM data standards, including naming conventions and metadata.
- **Reprocessing:** Reprocessing requests are performed due to data re-calibration, time offset corrections, ingest and value-added product updates and quality check limit adjustments. Reprocessing tasks include documentation and tracking, raw data clean-up, staging and processing, data and metadata review and notification.

BY THE NUMBERS

Data Integration Group 2012 accomplishments:

65
ingests ported to Linux

Collected and processed, on average, **130GB** of data per day from nearly **300 instruments**, amounting to **32,000 files** daily

6
new VAPs related to production

Transferred, on average, **292GB** of data per day to the ARM Data Archive, amounting to **4,200 files** daily from more than **1,000** unique datastreams

3
VAPs running in production using ADI

Processed over **8.2 BILLION** data samples

VALUE-ADDED PRODUCTS

Stage 0:
Initiation

Value-added products (VAPs) help fill unmet measurement needs or improve the quality of existing measurements. In addition, when more than one measurement is available, ARM also produces best-estimate VAPs.

Stage 1:
Development

In 2012, the Data Integration Group developed several VAPs:

Stage 2:
Evaluation

- Six VAPs are in Stage 1.
- Eight VAPs were developed and moved from Stage 2 to Stage 3.
- Six VAPs were released and they were moved from Stage 3 to Stage 4.

Stage 3:
Release

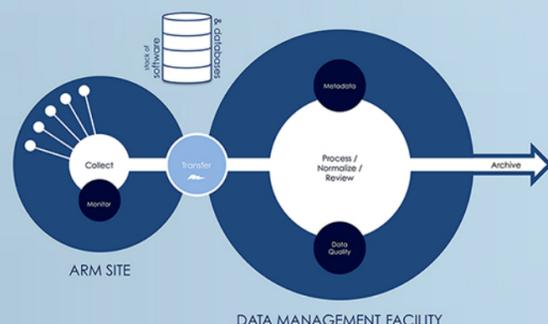
INFRASTRUCTURE SUPPORT

- **ARM Data Integrator Software (ADI):** Provides a development environment. Fully functional ADI libraries were released, and Python and IDL ADI library bindings were completed. Developers adopted ADI and developed eight VAPs—three in Production and five in Evaluation.
- **Data Discovery Browser:** Allows users to use pre-selected sorts and search logic to help find atmospheric and climate data faster, as well as data quality reports and data plots.
- **DSView:** Data availability tool that allows users to view near real-time instrument status across the data system in addition to a rolling window of hourly resolution.
- **DDTrack:** Track and display data delivery status at the instrument level.
- **Relocation:** Hardware for the Data Management Facility (DMF) was relocated to Oak Ridge National Laboratory.
- **Amazon Web Services:** Infrastructure web services, such as wikis and mail, were moved to Amazon Web Services.

MISSION

The ARM Climate Research Facility's mission is to provide the climate research community with strategically located in situ and remote sensing observatories designed to improve the understanding and representation, in climate and earth system models, of clouds and aerosols as well as their interactions and coupling with the Earth's surface.

The **ARM Data Integration Group** at PNNL supports that mission by delivering high-quality products in a timely fashion that impact the science needs of the ARM user community. The group develops and supports a wide variety of software, web tools, and computation resources that contribute to the success of reliable data products delivered to the ARM Data Archive in a reasonable time frame. The group engages with instrument mentors, translators, external developers, the Data Quality Office, and Data Archive staff to ensure completeness in all steps required for ARM data production.



▲ Data flow managed by the ARM Data Integration Group

www.arm.gov