

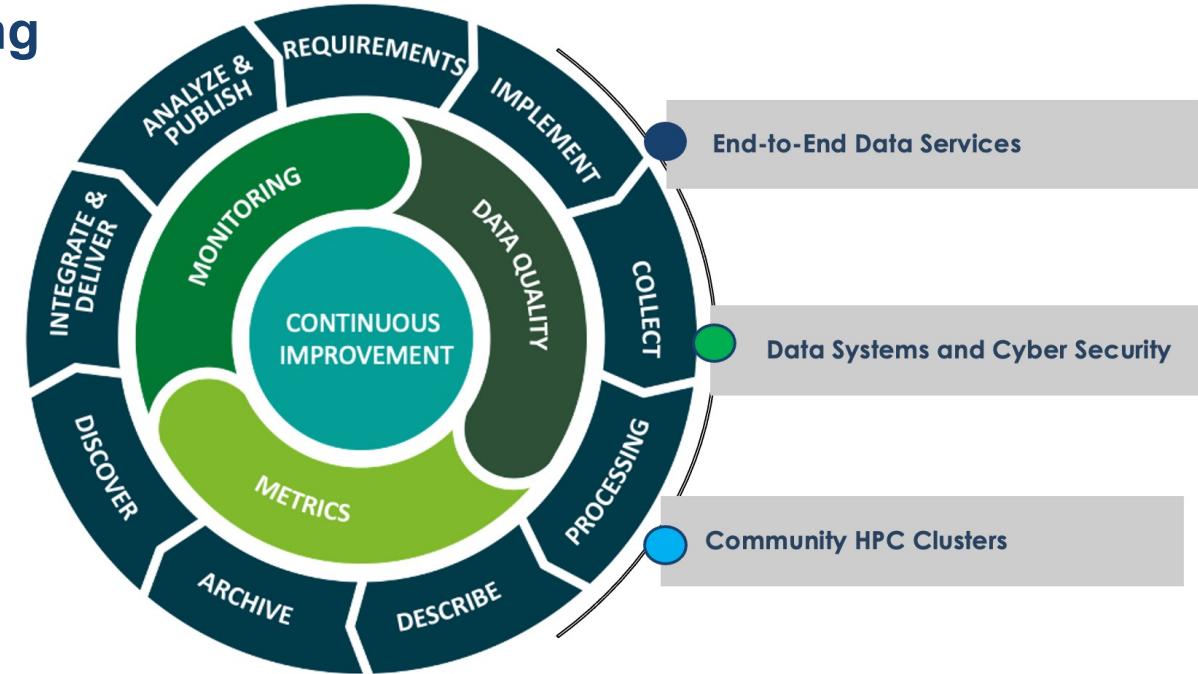
Data and Computing Services

Presented by Monica Ihli on behalf of ARM Data Services
Oak Ridge National Laboratory

ARM Data Center (ADC)

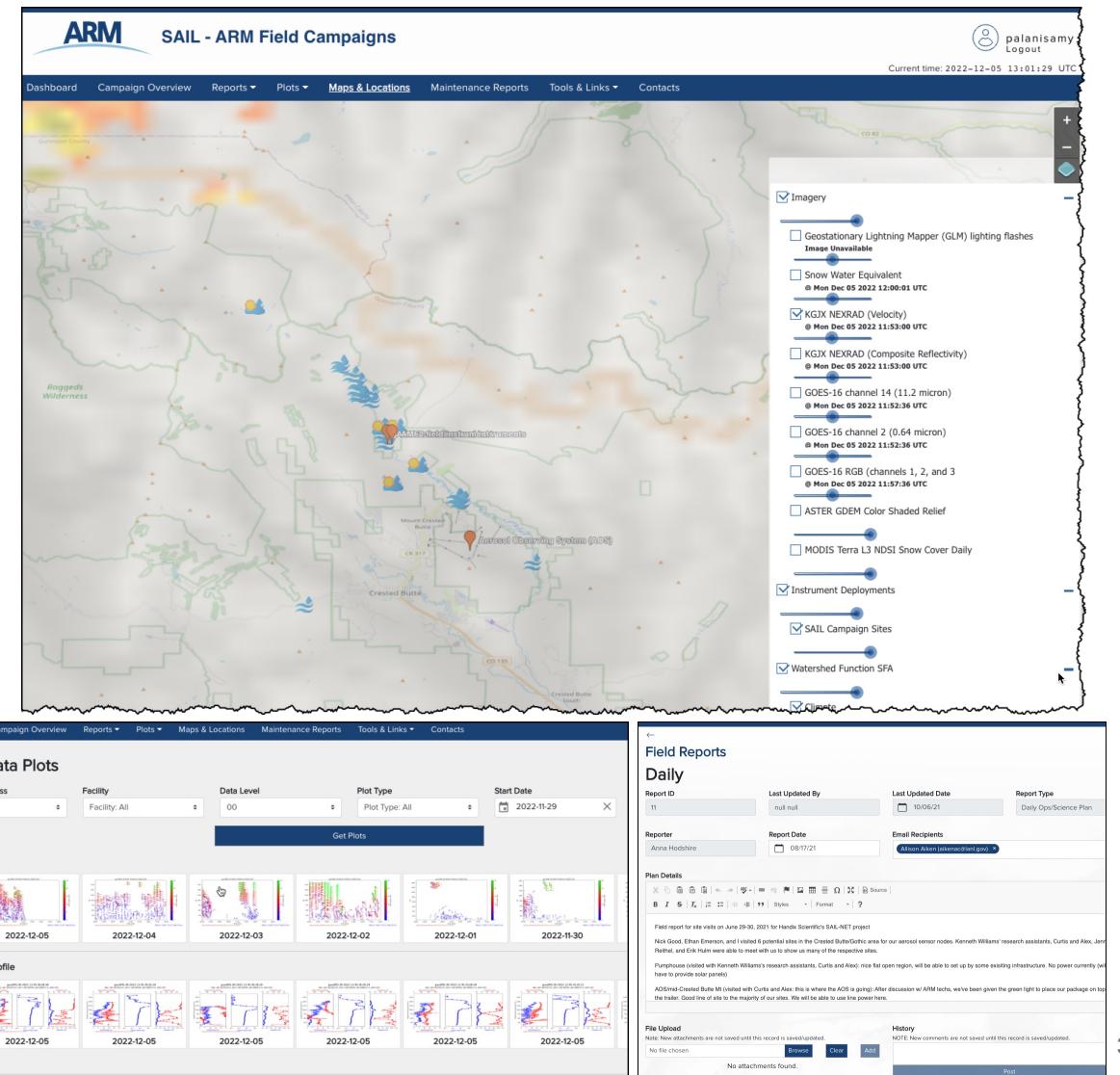
Provides a robust integrated data and computing ecosystem to advance understanding of atmospheric radiation

- ▶ Key components include:
 - Data management, operations, and monitoring
 - Data archive and distribution
 - Cyberinfrastructure
 - High-performance computing (HPC) environment
 - User metrics
 - Data analytics and visualization



ARM Field Campaign Notebook

- ▶ Central location to collect and manage field campaign notes, operations, and instrument status
- ▶ Incorporates Interactive maps, plots, and external charts
- ▶ Modular software allows expanding to future campaigns.
 - TRACER and SAIL: <https://adc.arm.gov/afcd/#/>
 - Requirements gathering for EPCAPE and TBS





ARM Data Submission Tool

- ▶ <https://www.arm.gov/guidance/datause/data-product-registration-and-submission>
- ▶ PI Data submission for archival using community followed standards
- ▶ New and modern user interface
 - Streamlined homepage and tabs
 - Templates
 - New data upload instructions
 - Readme viewer
- ▶ Share metadata drafts between team members

The screenshot displays two views of the ARM Online Metadata Editor. The top view shows the homepage with a button to 'Create a new metadata record' and a list of 691 records. The bottom view shows a detailed edit screen for a specific dataset.

Online Metadata Editor
Data Product Registration and Submission Tool

Create a new metadata record. Edit
Your current record count is: 0

Create New Edit

My Records All Records Templates new View Search

Records found: 691

Type to Search Clear (C)

| Record ID | Title | Creator | Update Date | Status | Action |
|-----------|---|----------------|-------------|----------|---|
| ARM0691 | EASTERN NORTH ATLANTIC ENA CONDENSATION PARTICLE COUNTER CPC with the AEROSOL MASK ENA AM for the YEAR 2018 | FrancescaGallo | 9/9/2022 | approved | Edit View |
| ARM0690 | EASTERN NORTH ATLANTIC ENA CONDENSATION PARTICLE COUNTER CPC with the AEROSOL MASK ENA AM for the YEAR 2017 | FrancescaGallo | 9/9/2022 | approved | Edit View |
| ARM0689 | EASTERN NORTH ATLANTIC ENA CONDENSATION PARTICLE COUNTER CPC with the AEROSOL MASK ENA AM for the YEAR 2016 | collierhr | 9/9/2022 | approved | Edit View |
| ARM0688 | NSA Cloud Occurrence Frequency from the Active Remote Sensing of Clouds ARSCL Value Added Product | collierhr | 8/22/2022 | approved | Edit View |

This file is READ ONLY. Any changes made will not be saved.

Approved

Dataset Title*
What is the Title of the Dataset? Include what, where, and when in the title. (Prefer 200 characters or less)*

ARM0689
EASTERN NORTH ATLANTIC (ENA) CONDENSATION PARTICLE COUNTER (CPC) with the AEROSOL MASK (ENA-AM) for the

Save

Contact Info Dataset Description Dataset Status Time and Place Tools Data Quality Keywords new Data Upload (E)

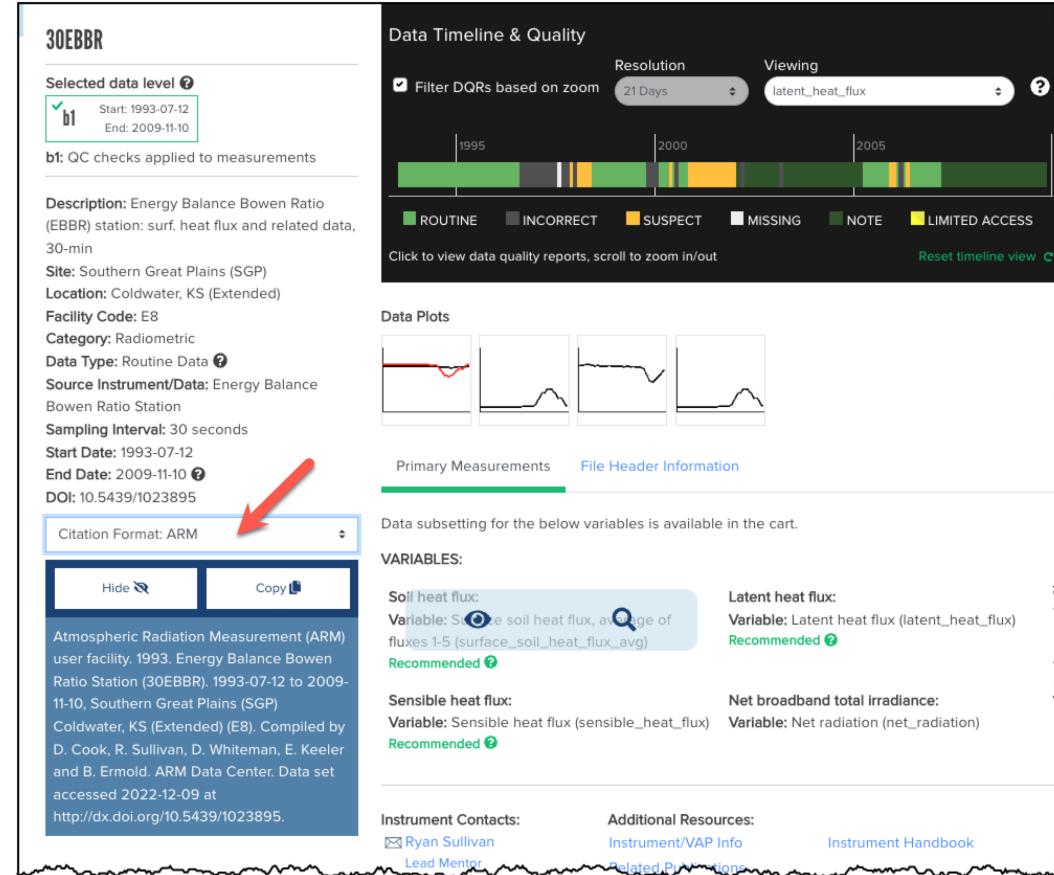
ARM Data Discovery Interface

- ▶ Intelligent search capabilities based on the data, guided search for user comfort
- ▶ New Capabilities via Continuous development and Integration
 - Location search
 - Timeline view
 - Unified details page for historical IOP data
 - DOI display improvements
- ▶ Recommendations, data tagging based on epochs or golden periods
- ▶ Upcoming Features
 - Linking with ARM Jupyter notebook

The screenshot displays the ARM Data Discovery Interface homepage. At the top, there's a navigation bar with links for HOME, DATA SEARCH, SUPPORT, ACCOUNT, and CART. Below the navigation is a large globe with location markers. A sidebar on the left contains sections for Data Search, Guided Search, and Location Search. The main content area features a "Data Search" section with a search bar and a "Search by category, measurement, datastream = more." link. It also includes a "Guided Search" section with a "For new users, search through guided questions." link. The "Location Search" section has a map and a "Use the map to search by site or facility." link. On the right, there's a "Data Search" section with a search bar and a "Select one of the categories below or type in the search text box to get started." link. Below this are three site-specific details panels: "ASCENSION ISLAND, SOUTH ATLANTIC OCEAN; MOBILE FACILITY (LASIC) (ASI)", "CORDOBA, ARGENTINA; MOBILE FACILITY (CACTI) (COR)", and "HOUSTON, TX; TRACKING AEROSOL CONVECTION INTERACTIONS EXPERIMENT (HOU)". Each panel shows a timeline of data availability from 2016 to 2017. To the right of these panels is a "FIELD CAMPAIGN DATA: ACME V TRACE GASES FROM THE AAF (G-1) (CD, N2O, H2O)" section with details like location, instruments, and data types. At the bottom right is a "FIELD CAMPAIGN DATA" section with a map and download links.

Data Citation

- ▶ Based on DOIs and citation guidance
- ▶ Enable metrics to quantify science impact and ensure data reproducibility
- ▶ Still evolving
 - Enabling time-based author credits
 - Additional citation formats
 - Nested citations
 - Data mashups
- ▶ <https://www.arm.gov/working-with-arm/acknowledging-arm/doi-guidance-for-datastreams>



<https://www.mdpi.com/154276>

Prakash, G, B Shrestha, K Younkin, R Jundt, M Martin, and J Elliott. 2016. "Data Always Getting Bigger—A Scalable DOI Architecture for Big and Expanding Scientific Data." *Data* 1:11.

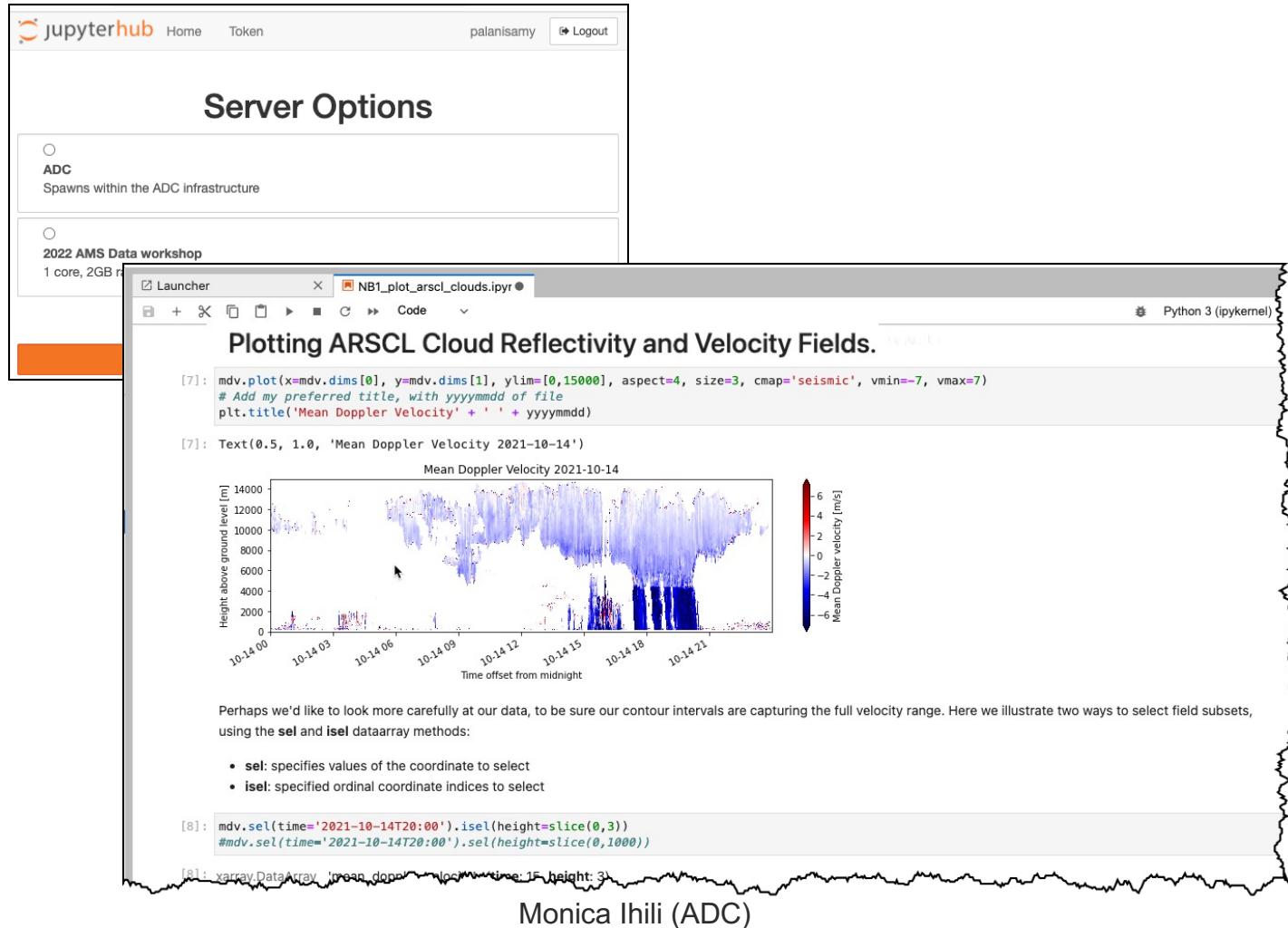
User Computing Capabilities - Cumulus

- ▶ Offers computing infrastructure to support next-generation ARM model simulations, Peta scale data storage, and big-data analytics for atmospheric and climate science research.
- ▶ New Cumulus cluster is now operational
 - 16,384 processing cores
 - Slurm cluster for
 - data processing, LASSO Operations, VAP generations
 - provides computing resources for Science projects (HPCR)
 - GPU nodes for ML/AI use cases (upcoming)
- ▶ <https://www.arm.gov/capabilities/computing-resources>



Enabling Data Analytics using Jupyter Notebook

- ▶ Integrated into ARM computing ecosystem.
- ▶ Use cases include scientific analysis, education, and instrument management.
- ▶ See [Knowledge Base](#) for more info.



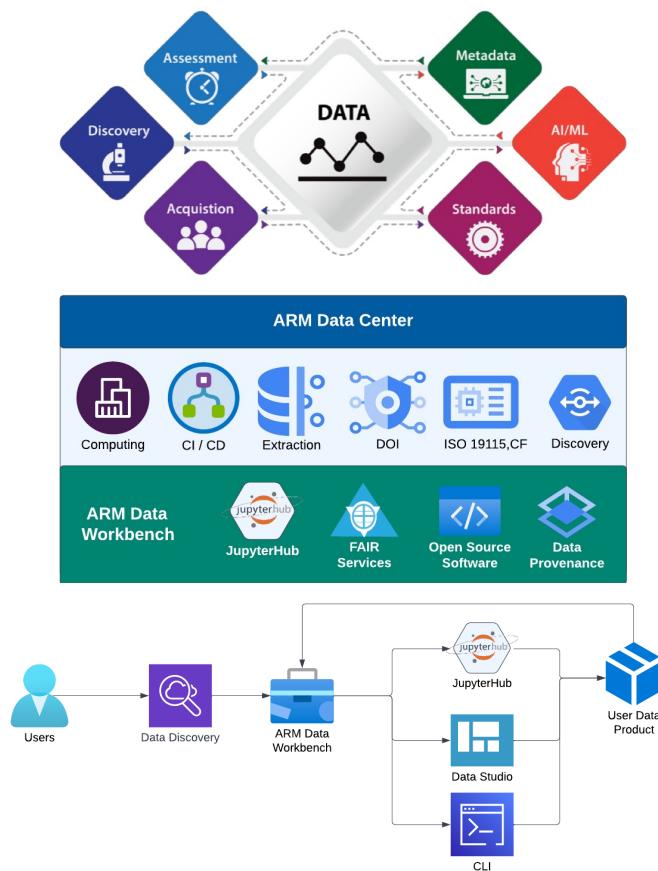
Fuy

Future Development & Product Roadmap

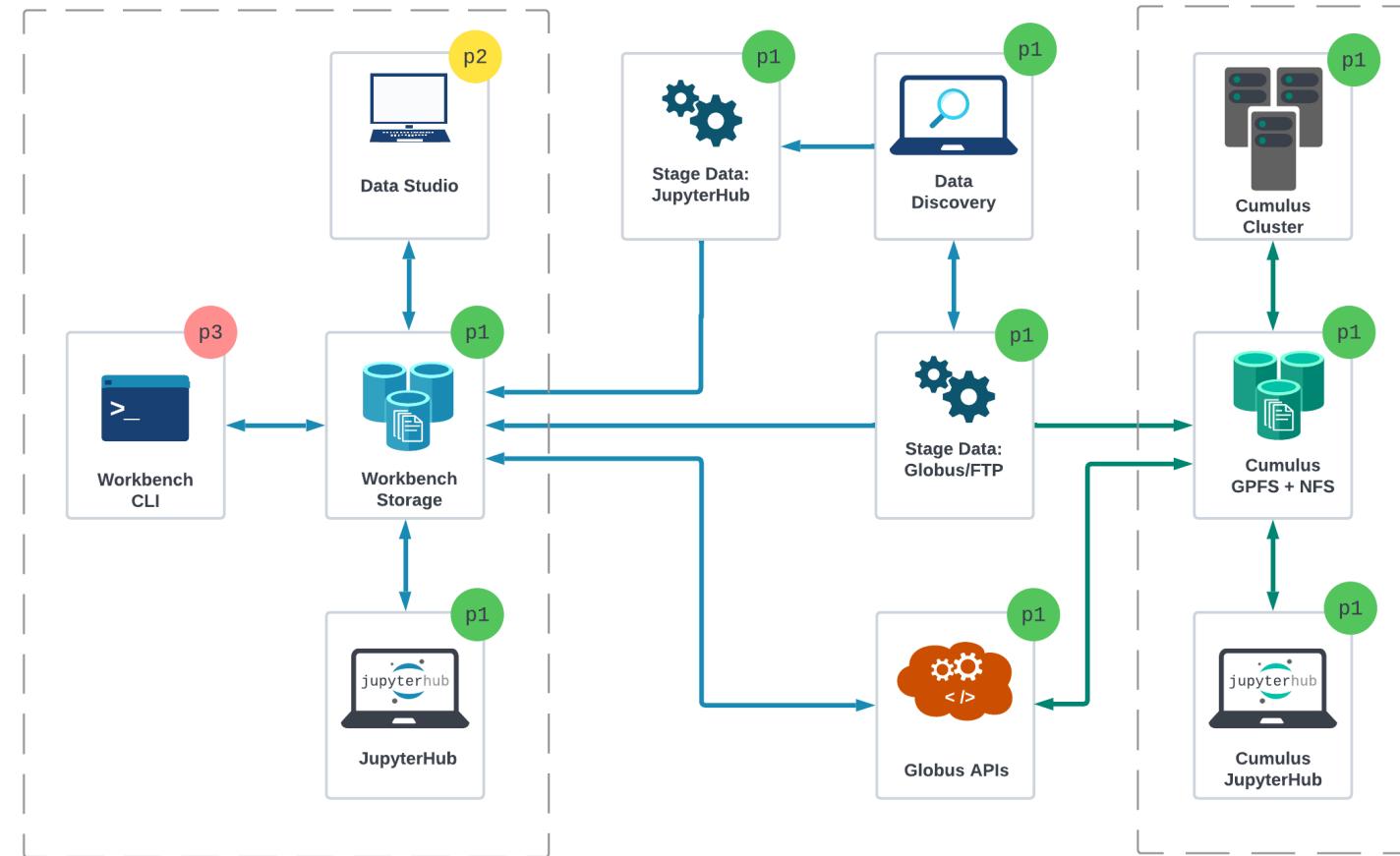


Data Workbench Ecosystem

Vision: Improve the user experience by providing unified ecosystem to access ARM computing, data, and software resources



ARM Data Workbench



K. Dumas and
M. Ihli et.al

ARM Cumulus

Open Source and Training Resources

- ▶ Enabling two-way contributions between ARM and user community
- ▶ Atmospheric data Community Toolkit (ACT)
 - Broader sharing of code
 - Reduction in duplicated efforts
- ▶ Py-ART
- ▶ ADI
- ▶ GitHub to support multiple requirements:
 - GitHub: ARM-DOE
 - GitHub: ARM-Synergy
 - GitHub: ARM-Development

Tutorials: <https://www.arm.gov/data/work-with-arm-data/webinars/>

The screenshot shows the 'Pinned' section of the ARM User Facility GitHub repository. It displays four projects:

- pyart** (Public): Python-based Radar Toolkit. A data model driven interactive toolkit for working with weather radar data. Python version, 405 stars, 241 forks.
- ACT** (Public): Atmospheric data Community Toolkit - A python based toolkit for exploring and analyzing time series atmospheric datasets. Python version, 89 stars, 28 forks.
- arm-gcm-diagnostics** (Public): A Python based ARM data-oriented diagnostics package for climate model evaluation. Python version, 13 stars, 4 forks.
- ADI** (Public): ARM Data Integrator. C version, 8 stars, 2 forks.

TUTORIALS

To help new users get started, ARM has developed a short [Introduction to Reading and Visualizing ARM Data](#) tutorial that steps through setting up some basic tools and illustrates how to read a simple ARM data file (using the Python programming language as an example).

An introductory video tutorial on how to [submit, discover, and acknowledge ARM data](#) is also available.

ARM is also conducting a series of data webinars and tutorials. On the [Data Webinars and Tutorials page](#), get information from past webinars, and see the upcoming webinar schedule.