Data as an Ecosystem at ARM User Facilities

Monica Ihli (Presenter)
Kyle Dumas
Jitendra Kumar
Giri Prakash
Zach Price

Oak Ridge National Laboratory
DoE Data Days (D3) June 1st, 2022
ihlimi@ornl.gov

https://www.arm.gov
Atmospheric Radiation Measurement (ARM) Facility

MISSION:
Provide the climate research community with strategically located atmospheric observatories to improve the understanding and representation in earth system models of clouds and aerosols and their interactions with the Earth’s surface.
The World’s Foremost Ground-Based Atmospheric Observing Facility

Since 1992, the world-leading facility for measurements of cloud & aerosol properties, & their impacts on Earth’s energy balance

Network of 3 fixed-location & 3 mobile observatories

Piloted & unmanned aerial measurement platforms

Extensive data management infrastructure

Freely available data products to support atmospheric research & model development

Large-eddy simulation (LES) model simulations and analysis tools

Serves the international climate research community and has close collaboration with Atmospheric System Research (ASR)

Comprehensive measurements across diverse climate regimes

Source: Jim Mather
### Background
- Atmospheric state

### Surface energy balance

### Aerosol and hydrometeor profiles

### Near-surface aerosol properties

### Aerial measurements

---

#### Map
- Fixed Site
- Past Fixed Site
- Deployments
- Aerial Deployment
- Ship Deployment
End-to-End Data Services

Providing powerful and adaptable computing resources to meet data analysis challenges
Data Recommender System – A solution for Data Discovery

https://adc.arm.gov/discovery
Data Recommender System – A solution for Data Discovery Challenge

- Recommends best data sources for the core measurements
- Criteria for recommendations include:
  - Quality
  - Temporal and spatial coverage and resolution
  - Applicability for the research needs
- Process include input from subject matter experts

<table>
<thead>
<tr>
<th>Scientific Focus Area</th>
<th>Best Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Core Measurements</td>
<td>Qcrad1long, armbeclrad</td>
</tr>
<tr>
<td></td>
<td>• Atmospheric radiation</td>
</tr>
<tr>
<td></td>
<td>• Atmospheric state</td>
</tr>
<tr>
<td></td>
<td>• Clouds, aerosols, etc.</td>
</tr>
<tr>
<td></td>
<td>• Downwelling broadband shortwave radiation</td>
</tr>
</tbody>
</table>

PIs: Maggie Davis (ORNL) & Scott Collis (ANL)
Data Recommender System – A solution for Data Discovery

New Guided Search

Searching For: Measurement

1. WHAT CATEGORY OF MEASUREMENT?

- Aerosols
- Atmospheric Carbon
- Atmospheric State
- Cloud Properties

Active and passive remote sensing instruments are used to measure the macroscopic properties (horizontal and vertical distributions) of clouds, and the microphysical properties (sizes, shapes, and phases [water or ice]) of the particles that comprise the clouds.

Measurement Category

- Measurement Sub-Category
- Primary Measurement
- Data Level
- Location
- Dates
ARM & the Data Lifecycle

▸ 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
Leveraging DOE Leadership Computing and commercial Cloud Capabilities

https://www.arm.gov/capabilities/computing-resources

Offers computing infrastructure to support next-generation ARM model simulations, petascale data storage, and big-data analytics for atmospheric and climate science research.

Cumulus HPC: 16,384 cores, 4 PB GPFS storage
Enabling Data Analytics

- Highly scalable Jupyterhub deployment to Kubernetes cluster.
- Access to All ARM Data
- Spawn data analytics and processing to ARM HPC
- Trainings and tutorials

Zach Price (ADC)
Data & Compute Resources

- Direct mount of data archives to Jupyterhub in ADC Infrastructure.
- Globus transfers from Discovery Interface.
- API's for data retrieval to HPC and Jupyterhub.
Looking Ahead...

Collections
- Southern Great Plains
- North Slope of Alaska
- Eastern North Atlantic
- Mobile Facility 1, 2, and 3
- Aerial Measurements
- Field Campaigns

Edge Computing

AI for Data Quality
- Data Quality Office/Instrument Mentors
- Science Translators

AI for data recommendations and interoperability

AI for Data discovery

Cloud optimized data access and services

Collections

Data Ingest, Processing, and Reprocessing

Metadata Management

Data Quality Operations

Science Products Operations

Compute Clusters

Archival and Distribution

Scientific User Community

ARM Data Center

PI and Research Data

External Data
"Enable advanced data analytics and community use of complex ARM data sets through the advancement of computing infrastructure and data analysis."

Atomspheric Radiation Measurement User Facility: Decadal Vision
DOE/SC-ARM-20-014
Lower the barrier to entrance for using ARM data.

- Library of off-the-shelf data analytics, algorithms, and visualizations
- Easy access to tutorials and guides.
Widen the Impact of ARM Data

• Less restrictive authorization requirements for using Data Workbench opens up using data to a wider audience.

• Training for the next generation of atmospheric scientists.
New Features and Capabilities

- Build custom datasets based on metadata facets.
- Merge datasets ("custom VAPs").
- Integration with archival options for derivative datasets that are created.
- Seamless integration with Jupyterhub and HPC.
Questions?

► https://www.arm.gov
► ”Ask Us”
► ARM Data Center: adc@arm.gov
► My contact: ihlimi@ornl.gov