



# *Data management and discoverability challenges, opportunities, and Best Practices*

## *– Case study from the Atmospheric Radiation Measurement (ARM) Data Services*

GIRI PRAKASH

**Oak Ridge National Laboratory**

2022 Research Infrastructure Workshop, September 16, 2022

palanisamyg@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



Background atmospheric state



Surface energy balance



Aerosol and hydrometeor profiles



Near-surface aerosol properties



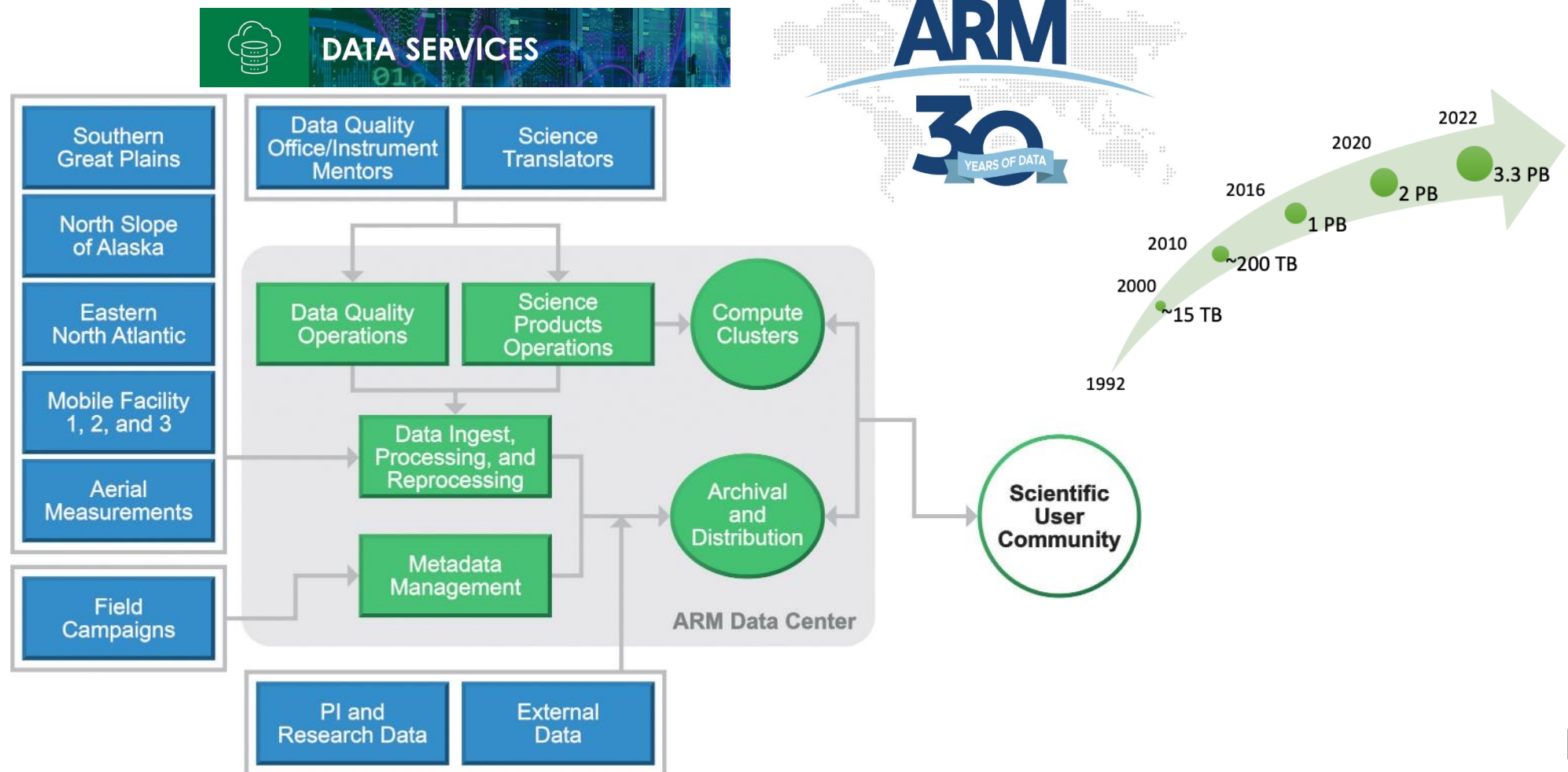
Aerial measurements



# End-to-End Data Services to Address the Data Growth Challenge

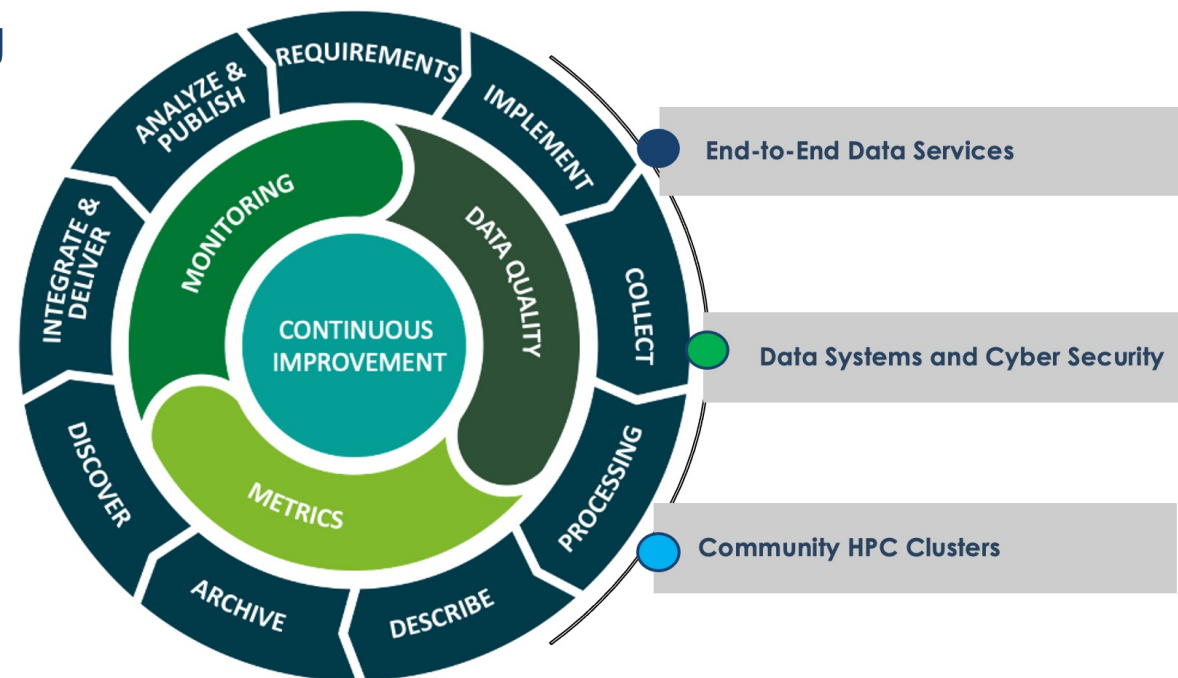


Providing powerful and adaptable computing resources to meet data analysis challenges



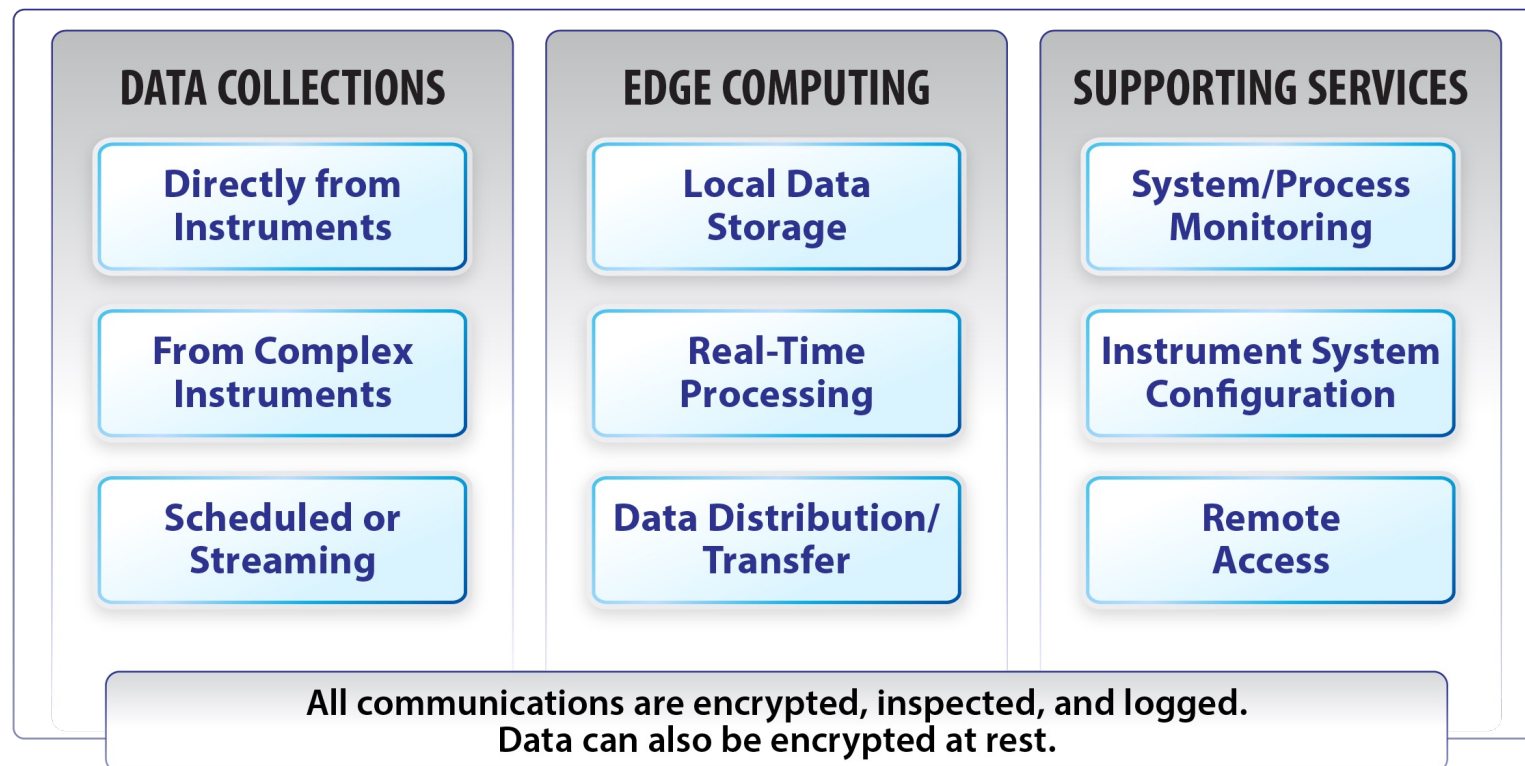
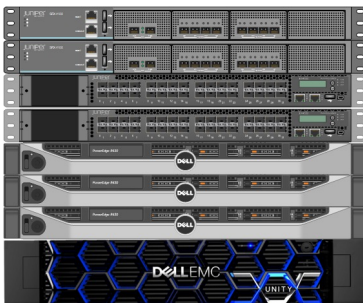
# ARM Data Center (ADC)

- ▶ Provides a robust integrated data and computing ecosystem to advance understanding of atmospheric radiation
- ▶ Always improving:
  - Data management, operations, and monitoring
  - Data archive and distribution
  - Cyberinfrastructure
  - High-performance computing (HPC) environment
  - User metrics
  - Data analytics and visualization



# Data Collection Systems – Challenges and Best Practices

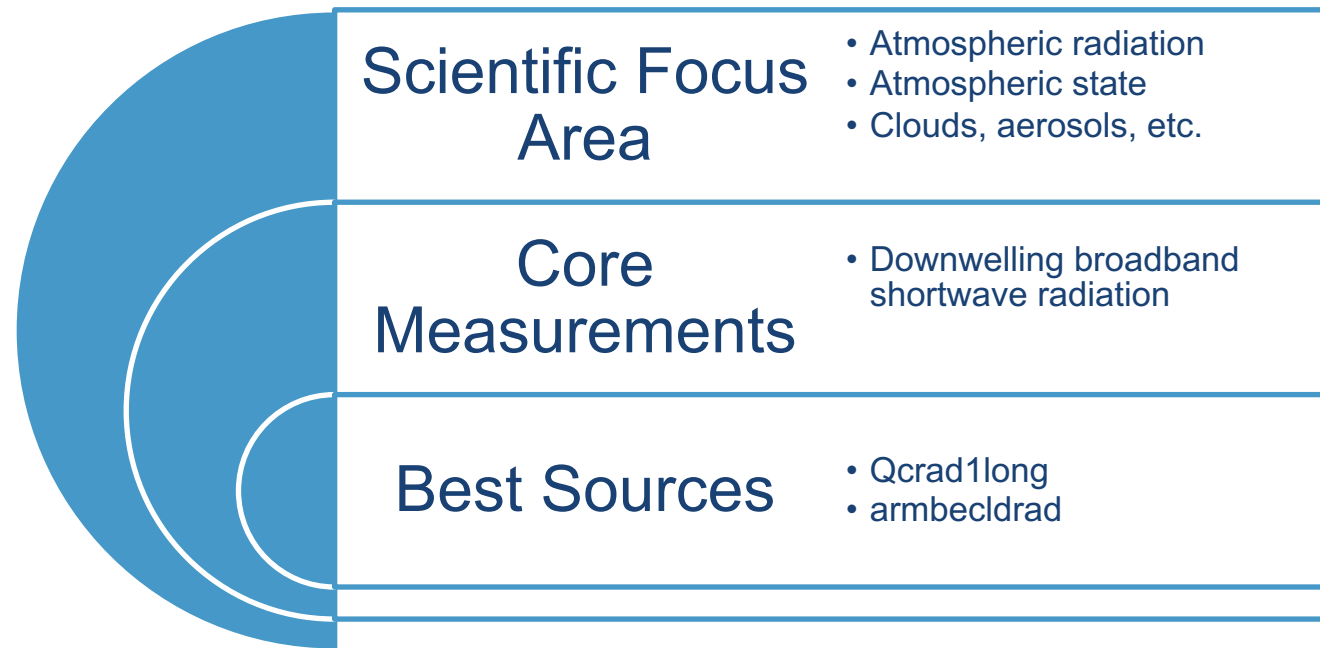
- ▶ Scalable data systems, hardware and software solutions proven effective over multiple generations/deployments
- ▶ Real-time data access to enable data reduction and edge computing (e.g., Supervised Learning)
- ▶ Operating next generation sensors with ML (e.g., Reinforcement Learning)



# Data Discovery Challenge - Data Recommender System Based Solution



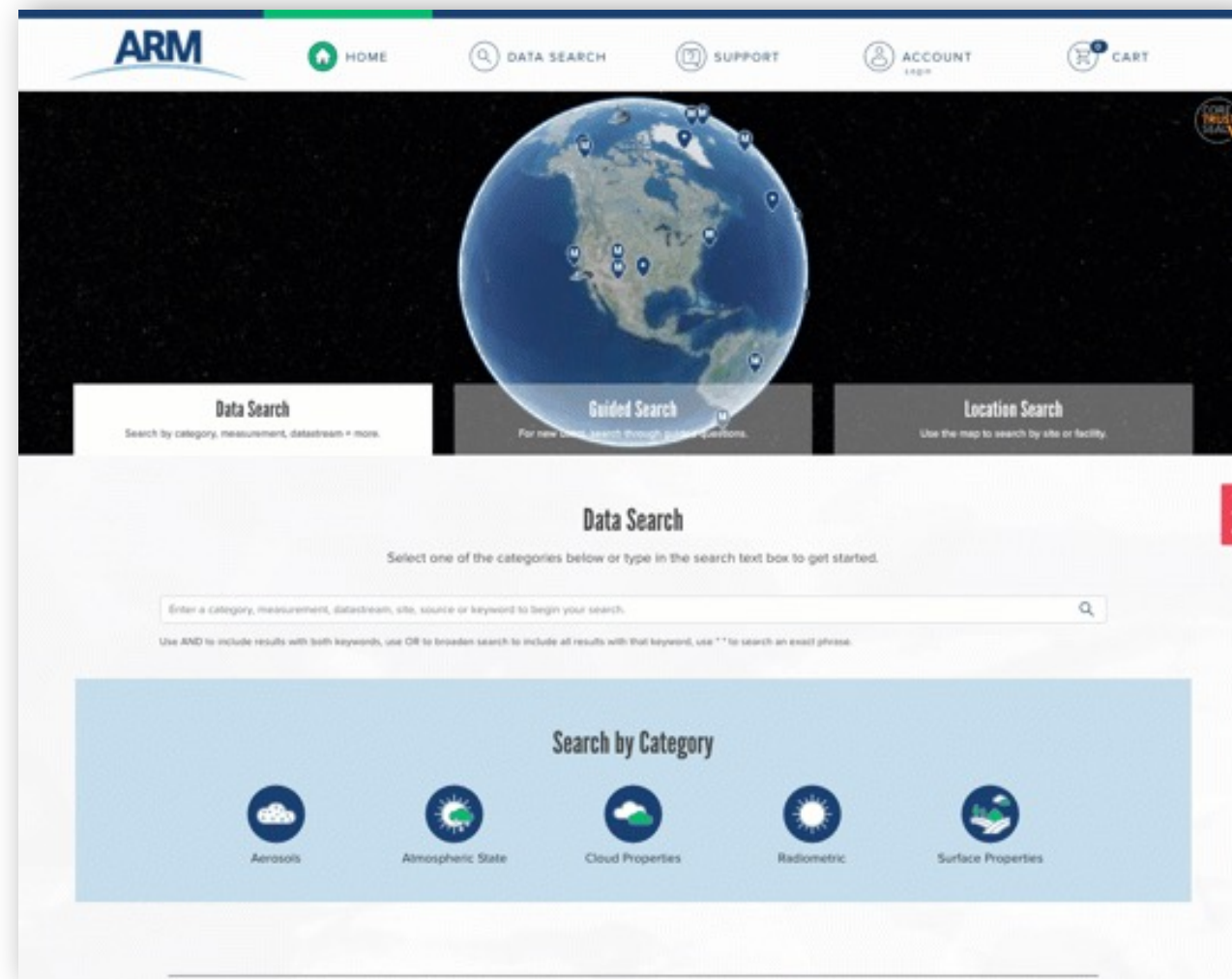
- ▶ Recommends best data sources for the core measurements
- ▶ Criteria for recommendations include:
  - Quality
  - Temporal and spatial coverage and resolution
  - Applicability for the research needs
- ▶ Process include input from subject matter experts



*PIs: Maggie Davis (ORNL) & Scott Collis (ANL)*

# Modern Data Discovery – Seamless Utilization of AI/ML

- Modern big-data software architecture with Continuous Integration (CI)
- Intelligent search capabilities based on the actual data, guided search for user comfort
- Recommendations, data tagging based on epochs or golden periods
- Near real-time access via secured webservices (API access)



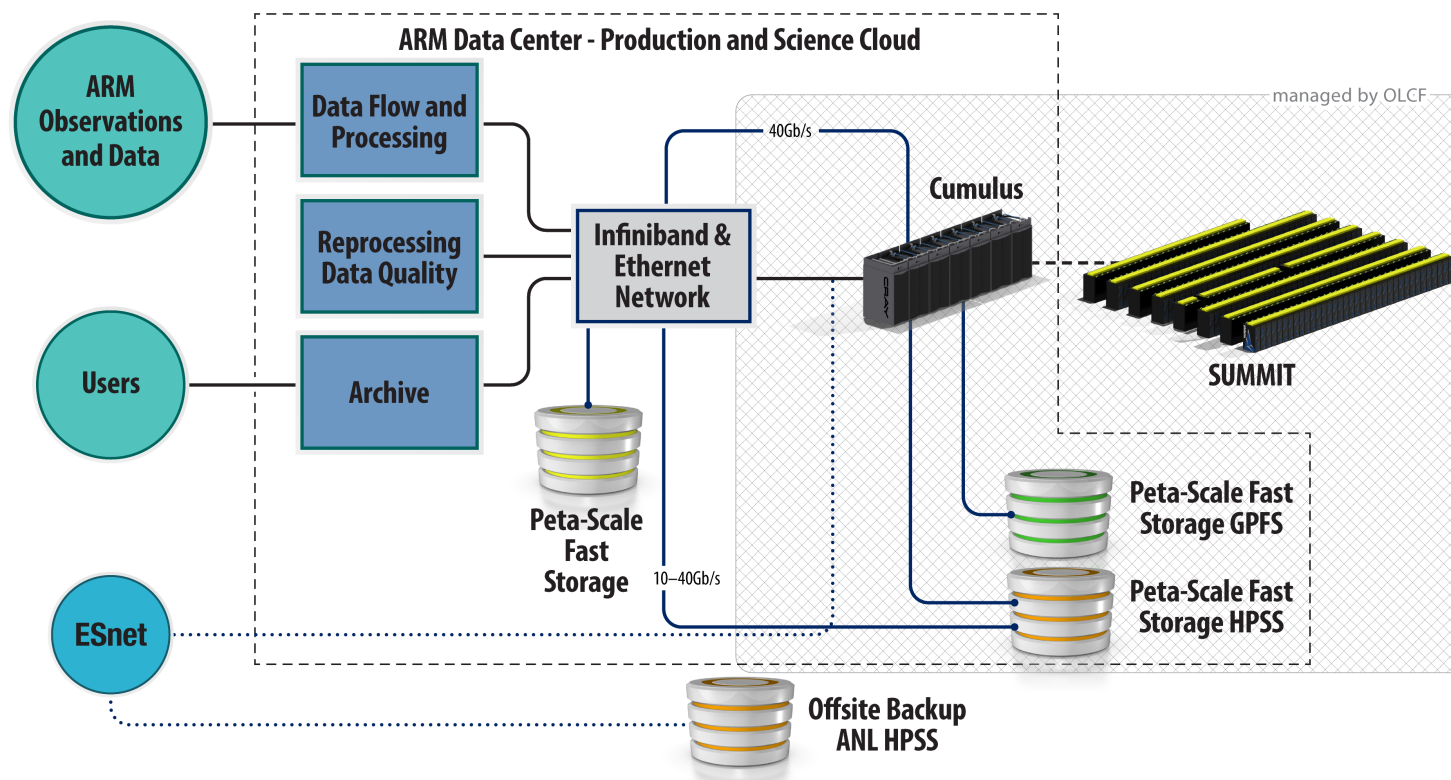


# Next-Gen Data and Computing Infrastructure

- ▶ Leveraging DOE Leadership Computing and commercial Cloud Capabilities

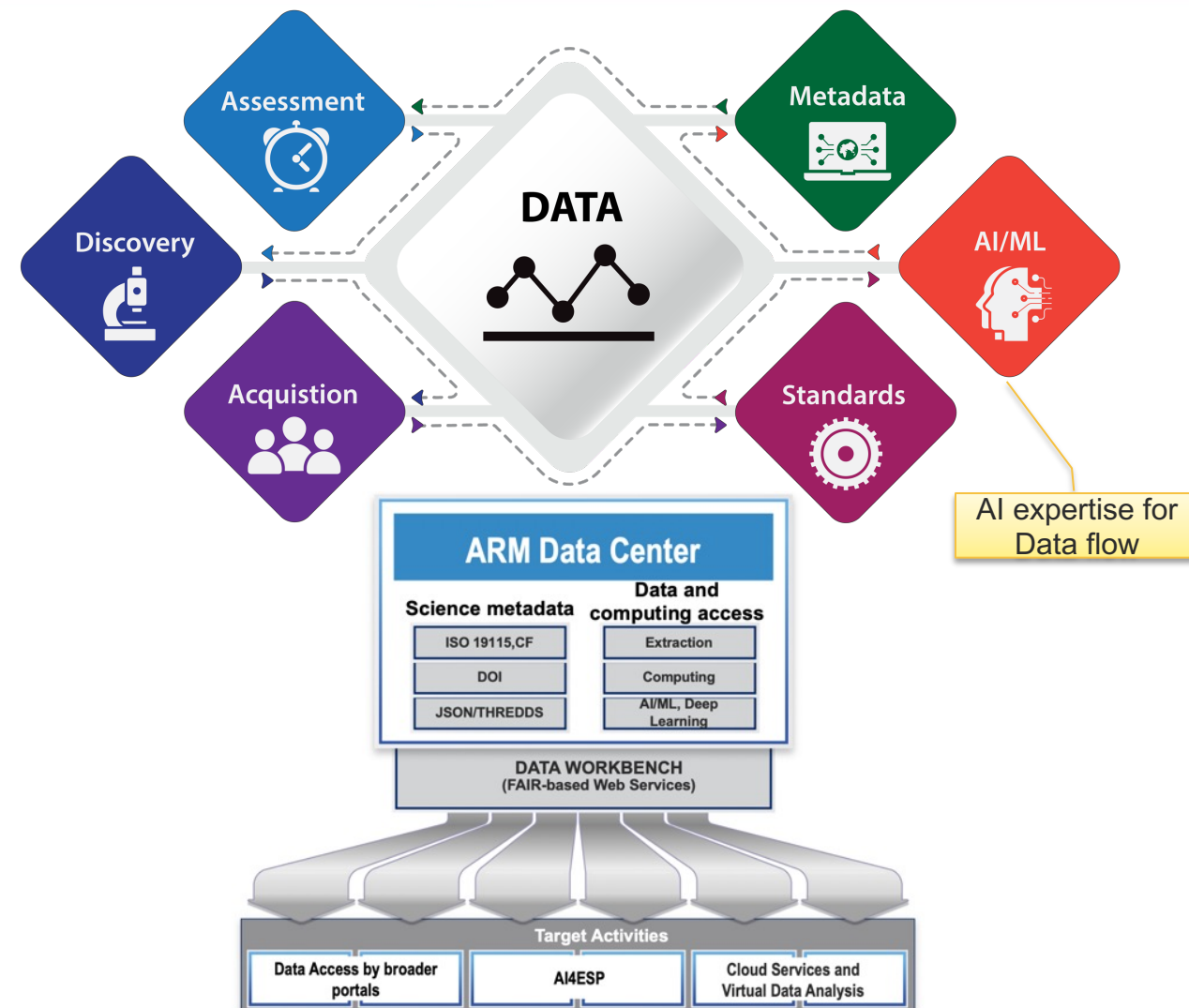
<https://www.arm.gov/capabilities/computing-resources>

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

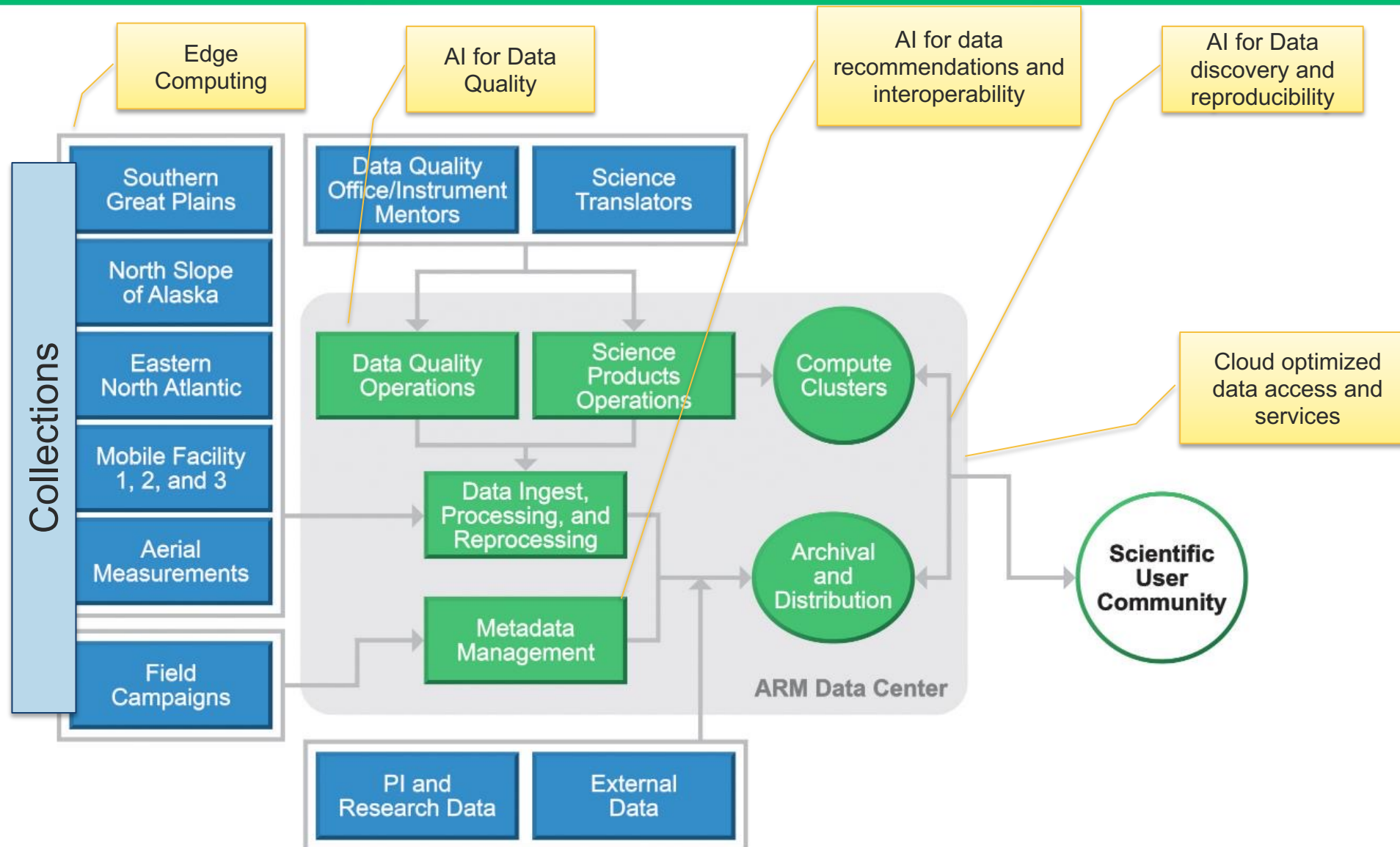


# Data Interoperability for AI and Beyond

- ▶ Aims to achieve transformative knowledge discovery by providing modular capabilities
- ▶ Enable FAIR-based access to ARM data and computing for initiatives such as AI4ESP
  - JupyterHub enabling data access and analysis
- ▶ Integrates future cyberinfrastructure for data analytics
  - E.g., Enable machine-learning framework to support data interoperability from diverse sources (ARM, NEXRAD, MODIS, ECMWF, etc.)








# Challenges and Opportunities



## Further Information & Contact

- ▶ <https://www.arm.gov>
- ▶ Decadal Vision: <https://arm.gov/publications/programdocs/doe-sc-arm-20-014.pdf>
- ▶ "Ask Us"
- ▶ ARM Data Center: [adc@arm.gov](mailto:adc@arm.gov)
- ▶ My contact: [palanisamyg@ornl.gov](mailto:palanisamyg@ornl.gov)

ATMOSPHERIC RADIATION MEASUREMENT USER FACILITY				
<p><b>CONNECT WITH ARM</b></p> <p>CREATE ACCOUNT</p> <p>ORGANIZATION</p> <p>    </p> <p>Reviewed September 2021</p>	<p><b>POLICIES</b></p> <p>DATA POLICIES</p> <p>CAMPAIGN GUIDELINES</p> <p>LINKING POLICIES</p> <p>PRIVACY &amp; SECURITY NOTICE</p> <p>DIVERSITY, EQUITY, &amp; INCLUSION</p> <p>VULNERABILITY DISCLOSURE PROGRAM</p>	<p><b>HELP</b></p> <p><b>ASK US</b></p> <p>ASK A UEC MEMBER</p> <p>DATA QUESTIONS</p> <p>FAQS</p> <p>ACCOUNT MANAGEMENT</p>	<p><b>RESOURCES</b></p> <p>MEDIA</p> <p>OUTREACH</p> <p>ACRONYMS</p> <p>GLOSSARY</p>	<p><b>WORKING WITH ARM</b></p> <p>USE ARM FACILITIES</p> <p>ACKNOWLEDGE ARM</p> <p>SUBMIT A PROPOSAL</p> <p>FIND EMPLOYMENT</p> <p>VIEW ARM PRIORITIES</p>