

Orchestrating Metadata, Ingest, and Data Quality Workflows for Scalable, High-Quality Data

Presenter: Adam Theisen

The Challenge

Atmospheric Radiation Measurement (ARM) User Facility



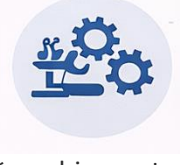
400+ Instruments deployed across fixed, mobile, and aerial platforms



100+ types or configurations of instruments



1+ PB of data produced each year



16+ Laboratories and other institutions



Automated and manual processes



Increasing AI demands

*graphic created using AI

Coordinating data flow at this scale is inherently complex

The Solution: INST

- ▶ Dependency-aware tasking framework
 - Connects metadata, ingest, site operations, and data quality
 - Assigns tasks to the right expert at the right time
 - Supports parallel and sequential activities
 - Enables automation
- ▶ Enhanced with AI for Data Quality and Metadata

The Impact

- ▶ Timely creation and release of **high-quality** datasets
- ▶ **Reduction** of operational burden
- ▶ **Transparent** and auditable workflows

Transforming operational complexity into coordinated, scalable data operations