

Standards and Quality Control Processes for Earth Science Datasets

Josh Howie¹, Chitra Sivaraman¹

¹Pacific Northwest National Laboratory

Atmospheric Radiation Measurement User Facility

ARM

www.arm.gov

Input Data and Metadata

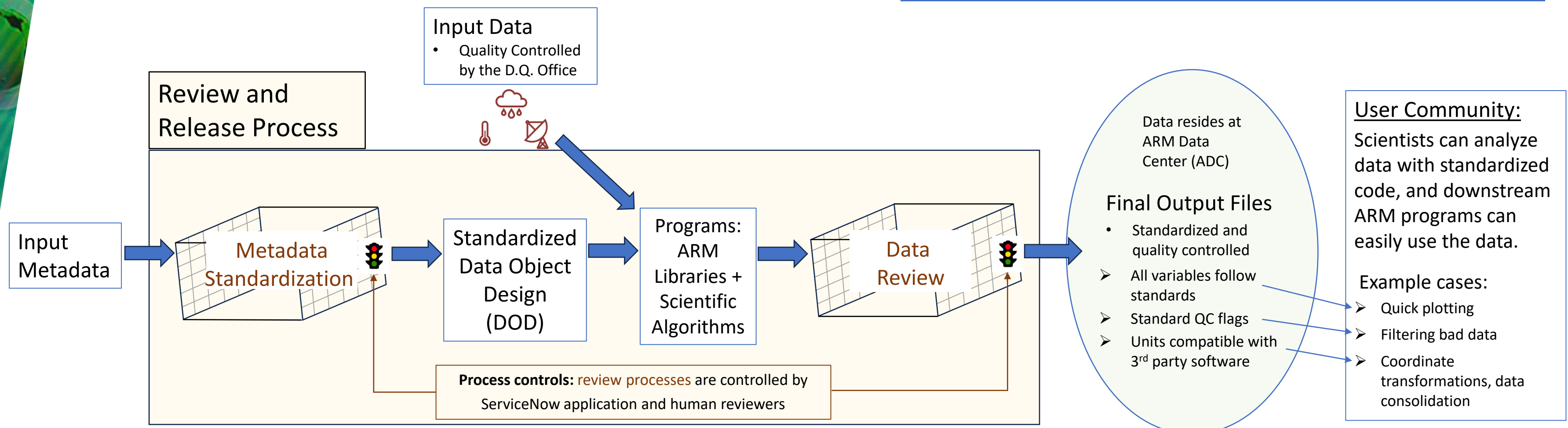
ARM has over 400 instruments measuring a wide range of atmospheric and geophysical properties. In addition, ARM has legacy data formats, some of which were implemented over 30 years ago.

- Inputs vary by spatial dimensions and time resolution.
- Types of data and metadata can vary widely by the type of instrument taking the measurement, such as radars, aerosol systems, and radiometers.
- **Summary:** heterogeneous data and metadata inputs.

Goal: Produce quality controlled and standardized output files from numerous disparate input data sources.

Approach: ARM uses multiple workflows and review processes to check the metadata and data. These processes are managed by software tools, but also include human reviewers at key steps.

Impact: Standardization of data ensures consistency, interoperability, and reliability, making it easier for users to integrate, analyze, and trust the data. It also simplifies data processing, enhances efficiency, and supports the creation of scalable and maintainable software tools.



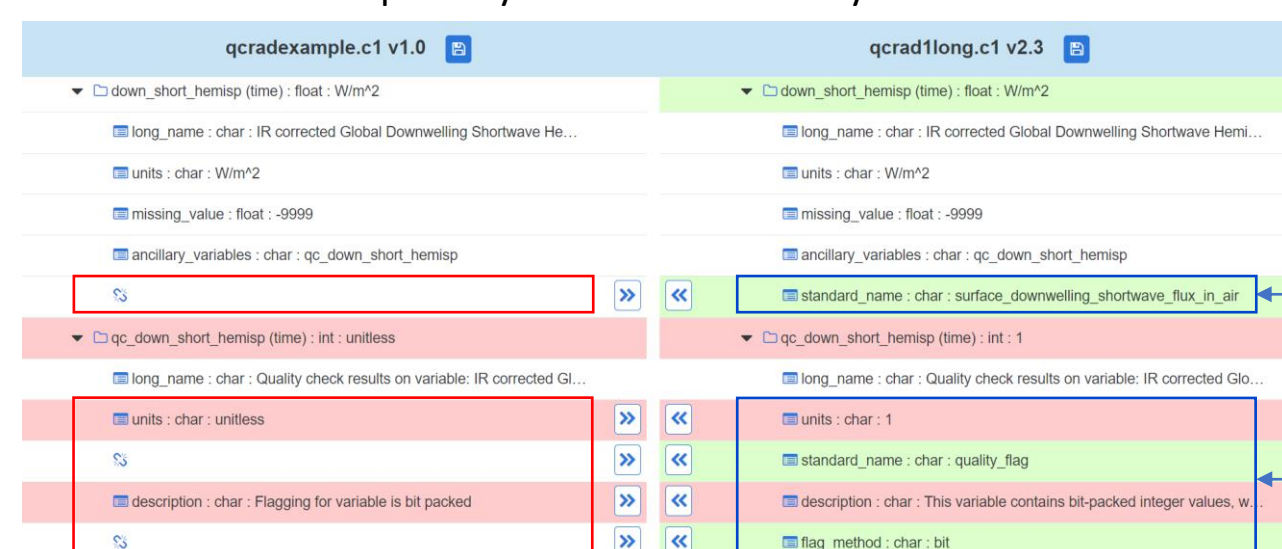
Metadata Standardization

Human review: Review committee looks at and approves or rejects the submitted Data Object Design (DOD).

- Review team compares DOD against ARM standards document.
- Standardization includes standard quality control flags, standard names for variables, and other variable attributes.
- DOD must be approved before data can be released to the ARM data center.

Software tools: ARM Process Configuration Manager (PCM) has manual and auto-fix options.

- Units and flagging method attributes are automatically fixed by PCM
- Standard name for primary variable is manually entered



DOD version comparisons in PCM

Data Review

Data reviewers inspect both data and metadata output after programs create the output files.

- Custom ARM software (ncreview) allows quick and thorough comparisons of a single datastream or multiple datastreams.
- Data reviewers ensure:
 - Metadata matches the data. Examples: are the units correct? Do the metadata descriptions of the variables make sense?
 - Basic data quality is achieved.
 - No large and unexplained gaps in the data exist.
- Production level data must be approved before release to the ARM data center.