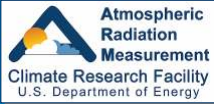


Investigation of new alternatives for data access and distribution from the ARM Archive



Raymond A. McCord, Giri Palanisamy, Richard C. Ward, Richard T. Cederwall, Dale P. Kaiser, and Sean T. Moore



Do you wish the Archive could:

- Give you only the measurements you want?
 - Provide extracted data in other formats?
- Read on...

Introduction: The ARM Program and the ARM Archive are seeking to give scientists an alternative mechanism to access and download ARM measurements that would be more flexible and efficient for both the data users and the Archive. This “new information sharing method” would allow the data requesters to dynamically select the measurements, locations, and time ranges required for their research, as compared to the current file-based method of ARM data delivery, which delivers data at the “whole file level” (netCDF files organized into data streams that typically contain numerous data fields/variables). The alternative method would give users the ability to request only specific measurements of interest, thereby streamlining data delivery and significantly reducing the volume of data sent to the user.

Similarities with Current User Interfaces

- Gathers selection criteria: contains menus for selecting site, facility, date range, and measurement
- “Extracts” the selected data from “Storage”
- Output files available via FTP after email notification
- Provides ancillary documentation: DQRs, path to web documentation, estimates of data volume, etc.

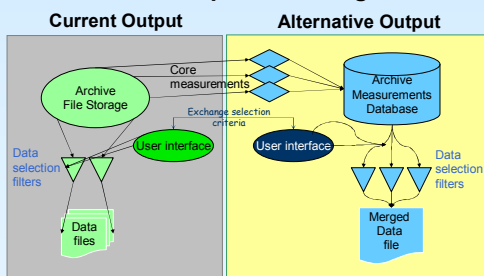
New Features of the Alternative User Interface

- Extraction of only specific measurements of interest rather than entire data stream(s)
 - initially limited to “core” measurements (determined via user feedback)
- Combines selected variables from different data streams with flexible options for output file format
 - netCDF, delimited ASCII, XML
- Provides various options for output file organization
 - one file per month containing n variables
- Construct conditional data selection queries
 - give me measurement M when direct radiation $< N$ w/m²
- Provides explicit cross-referencing between output measurements and associated source data streams

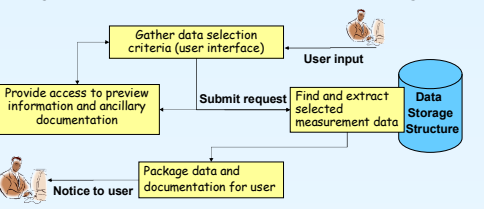
User Survey Questions

- **Core measurements:** What measurements (including both the variables and desired time interval) would you want available in an online database? For example, if measurement is near-surface temperature, would you want 1-minute or 30-minute data?
- **Conditional sampling:** Would you want to be able to select measurements based on the concurrent value of one or more other measurements? For example, obtaining the heights of cloud boundaries when the near-surface temperature is greater than 10° C and surface winds are between 135° and 225°
- **Continuous best estimate products:** If we constructed complete records for selected core variables (e.g., near-surface temperature, broadband radiation) using ARM’s best estimate, what variables and temporal resolution would you recommend?
- **Gap filling:** If there are holes in the data record, do you have specific requirements for “gap filling” approaches we should use? Would you prefer to do your own “gap filling” or do none at all?
- **Visualization options:** What kinds of online plots do you need for your measurements of interest? Are pre-existing thumbnails plots sufficient, or do you require interactive plots of some part of the data selected by your criteria?
- **Data quality:** What kinds of data quality information do you need included in the file with your selected data?

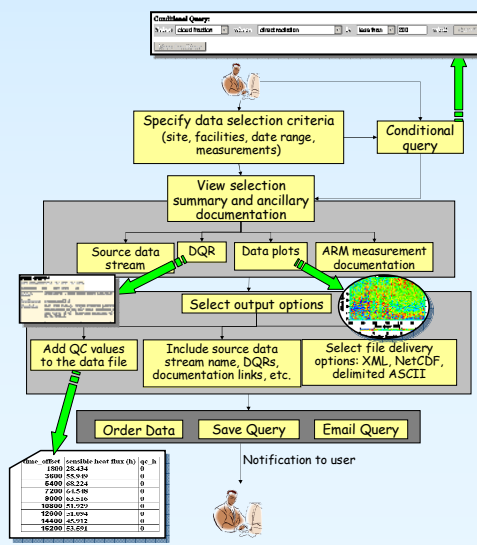
Archive Output Flow Diagram



Major Components of Alternative System



Alternative User Interface Flow Diagram



Possible Future Additions

- Include measurements from IOP data streams in the archive
- More detailed preview estimates of data volume per query
- Interactive display of information relating to chosen measurements
 - from a very large library of pre-computed plots and summary statistics
 - user-defined interactive graphics
- More options for user output:
 - additional file-type formats
 - user-selected sorting and splitting sequences, etc.
- Include ability to save/reuse queries, and share them via built-in email feature
- User-selectable options for temporal interpolation and integration
- Options for data gap filling