

Updating Recommended Datastreams

Ric Cederwall¹, Scott Collis², Maggie Davis¹, Ranjeet Devarakonda¹, and Giri Prakash¹

¹ - Oak Ridge National Laboratory, Oak Ridge, TN ² - Argonne National Laboratory, Argonne, IL



Summary: “Recommended” datastreams address a need identified in the recent Triennial Review that is being met through Data Discovery: to assist users to find the “best” ARM datastreams for their needs. A ranked set of datastreams for core geophysical quantities were first identified in 2012 by Jim Mather and ARM Translators. An update is needed and is being led by the Metadata Team and Translators.

Objectives

- Update current list of recommended datastreams to reflect latest products, and geophysical quantities
- Better assist users in finding “best” ARM datastreams to meet their needs
- Develop process for continual updates

Criteria for recommendation

Geophysical quantities were selected based on the core mission of ARM: to improve the understanding and representation, in climate and earth system models, of clouds and aerosols as well as their interactions and coupling with the Earth’s surface.

Rationale of recommendation includes data quality, completeness of record, and representativeness for two user communities

1. point, column (soda straw) with high frequency and resolution for physical process studies
2. spatially / temporally averaged to model grid for model evaluation and comparison with data

Example for geophysical quantity, surface temperature:

1. Rank 1 for community #1 -- datastream **sgpmetE13.b1**, variable: **mean temperature**
2. Rank 1 for community #2 -- datastream **sgparmbeatmC1.c1**, variable: **surface temperature**

Best estimates are included but there are additional recommendations that may be useful.

There may be multiple recommendations at a site due to dimensionality (e.g., point, grid)

Current state of recommended datastreams

There are 46 **core geophysical quantities**. They cover these measurement categories:

16 - **atmospheric state**, 4 - **surface properties**, 10 - **cloud properties**, 6 – **aerosols**, 10 - **radiation**

There are currently **3474** recommended **geophysical quantity:datastream** pairs, distributed as follows:

	aerosols	atmos state	cloud prop	radiation	surface prop	TOTAL
Rank 1	120	58	163	1421	160	2449
Rank 2	264	257	105	28	136	790
Rank 3	0	12	33	190	0	235
Subtotal	384	854	301	1639	296	3474

Information about core geophysical quantities, recommended datastreams, and associated metadata are in arm_int2 database tables, and accessed by Data Discovery for display to users

Example:

ARM Best Estimate product for atmospheric measurements is the Rank 1 recommended datastream for latent heat flux and for atmospheric temperature – spatially averaged for modelers.

The screenshot shows the Data Discovery search interface. It includes filters for Start Date, End Date, and Search (Temperature, Sites: Southern Great Plains). Categories include Surface Properties, Atmospheric State, Radiometric, Cloud Properties, Atmospheric Carbon, and Aerosols. A list of datastreams is shown, with 'latent heat flux // Surface Latent Heat Flux, upward positive' and 'Atmospheric temperature // Temperature, at 2m' highlighted as recommended datastreams.

Preview of new Data Discovery:

New data discovery display will elevate display of recommended datastreams by allowing the user to toggle “off” a default display of only recommended datastreams. Recommendation level will be retained in the new display

The screenshot shows the new Data Discovery interface. It features a search bar, filters, and a list of datastreams. The 'Recommended Data (47)' section is highlighted, showing 'Surface Radiation Measurement Quality Control Testing, including chronologically configurable limits' as a recommended datastream.

Progress

Recommended Datastreams Update Status (see PRJTASK0031663):

- 11 out of 46 **geophysical quantities** have updates proposed for **paired datastreams**:
 - 448 **Rank 1** paired datastreams
 - 136 **Rank 2** paired datastreams
 - 35 **Rank 3** paired datastreams
- Scott Collis is liaison with Metadata Team: coordinating Translator assignments
- Tracking in ServiceNow for each geophysical quantity by assigned Translator and Metadata Team member

New Tool for Routine Updates:

The screenshot shows the 'Recommended Datasource Graph' tool. It displays a network of datastreams and their relationships. The central node is 'Downwelling Broadband SW Flux'. Other nodes include 'gradflong (Self-Estimate_down_short_hemis) [158]', 'beffluxlong (down_short_hemis) [1]', 'srs (down_short_hemis) [17]', and 'srs (down_short_hemis) [22]'. The graph is ranked, with Rank 1, Rank 2, and Rank 3 nodes.

Capabilities:

- 1) Shows recommended datastreams for selected geophysical quantity
- 2) Accesses current database information for that quantity
- 3) Allows revision of assigned datastreams and ranks
- 4) Uses interactive graphics and advanced data type

Future Plans

- Enhance display of recommendations in Data Discovery
 - Use recommendations for choosing what datastreams are displayed in Data Discovery
 - Rationale for recommendation and characteristics of datastream displayed to users
 - Additional datastream characteristics included in new Data Discovery Design
- Include identified Data Epochs as a source of recommended datastreams
- Allow requests from infrastructure and user community for recommended datastreams and core geophysical quantities

Workflow for processing requests for new or updated recommended datastreams

- Metadata Team reviews incoming requests, interacts with requester
- Processed request sent to Translator for final review and approval
- Approved request implemented in ADC and made available through Data Discovery

