



ARM Data Quality Office Historical Data Quality Report Review



K. Kehoe¹, S. Moore², J. Monroe¹, A. Theisen¹, E. Kruse¹, H. Lorenzen¹,
J. Ragland¹, B. Sullivan¹, K. Western¹ and R. Pepler¹



1-ARM Data Quality Office, CIMMS/University of Oklahoma, Norman, OK
2-Alliant Techsystems Inc., San Francisco, CA



Background

The ARM Archive Data Discovery ordering interface enables users to filter data based on what is contained within Data Quality Reports (DQRs). By default, all data indicated as 'incorrect' in DQRs will be removed. This new feature requires DQRs to be more precise.

There are over 7300 DQRs currently in review by the ARM Data Quality (DQ) Office.

Phase 1 DQR Review

DQ Office Analysts are now reviewing DQRs to better understand the scope of the needed updates. During the process, analysts review each DQR to ensure correctness of problem reporting including but not limited to:

- Quality color
- Datastreams affected
- Variables affected
- Start/End dates and times
- Subject
- Suggestions to data users

The analysts review data via NCVweb or the new DQ Interactive Plotter (at left) to verify dates and times. All potential problems and time changes are then documented for future updating.

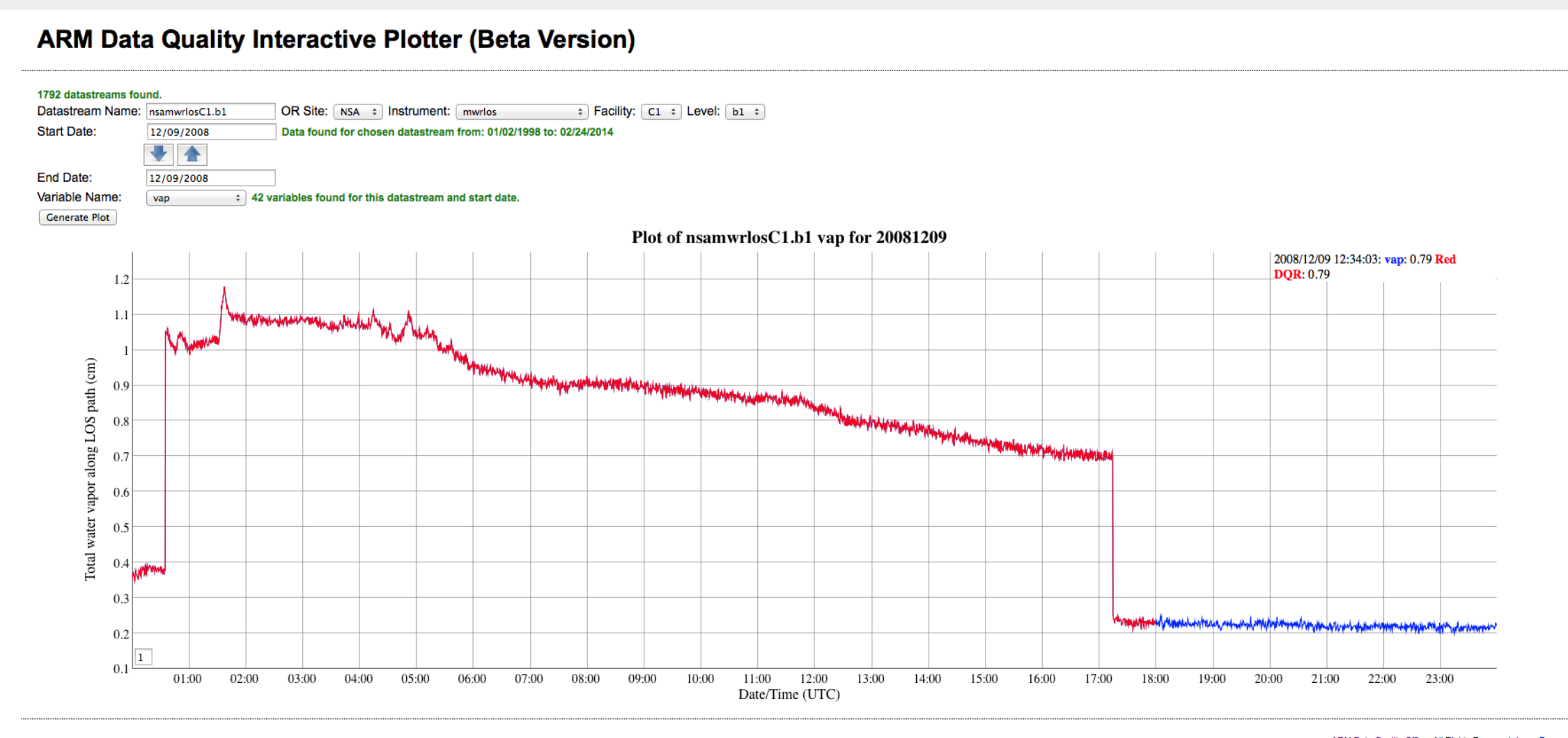
Phase 2 Define Problem Scope

Results - As of 3/5/2014

Total DQRs	7332
DQRs Reviewed	1009
Change/Further Review Needed	576
Start/End Date Problem	253

The number of DQRs needing some type of update (57%) suggests the development of a large scale modification process. This requires modifications to existing tools or development of new tools to efficiently update many of the individual DQRs.

Tools to Identify Problems



Phase 3 Proposed Modification Process Options

We are proposing two possible methods for updating the large number of DQRs

Method 1

1. Clone individual DQRs within production database, but with new PRB_reviewed status
2. Expand existing tools to add correct time ranges per DQR
3. Use existing or build new tools to verify updates
4. Enlist the help of Mentors and others in ARM to review modifications
5. Review modified DQR via the standard process currently used by the Problem Review Board
6. Delete or inactivate old DQR

Method 2

1. Clone entire DQR database to local copy
2. Fix problems using 3rd party database tools
3. Add correct time ranges to database copy
4. Expand or build new tools to verify updates
5. Enlist the help of mentors and others in ARM to review and approve modifications.
6. Synchronize corrections with the Production DQR Database.

Fix Reversed Time Fields

Edit: varname_metric

var_name	up_long_hemis
datastream	sgpqrnd1ongE6.c1
metric_value	3
entry_date	2012-05-10 21:00:00
start_date	2006-11-29 00:00:00
end_date	1999-03-18 15:00:00
id	D12051026
modify_comment	
modify_date	now

Buttons: Save, Save and continue edit, Delete

Update DQR Details as Needed

dqid	D0005222
submit_date	2009-09-04 18:24:00
subject	SGPMFRSR/E24 - Shadowband alignment
person_id	5327
description	Apparently cows broke through an electric fence and caused problems with the HFRSR after about 31 January. This caused the HFRSR to shade incorrectly. The shading problem was not corrected until 20080324.
suggestions	
PRB_reviewed	Y
warehouse	N
modify_date	now
modify_comments	

Buttons: Save, Save and continue edit, Delete