

CONSENSUS METADATA STANDARD: Data Quality Flags

REVISION DATE: November 5, 2001

CHECK FOR LATEST VERSION ONLINE: Before using this document, and periodically thereafter, please check for an updated version at <http://cdiac.esd.ornl.gov/programs/NARSTO/metadatastandards>. The latest version will be called `consensus_flag_standard_yyyymmdd.pdf`. The referenced Supersites detailed data quality flag list is also located at this URL and named `SS_detailed_flags_yyyymmdd.xls`.

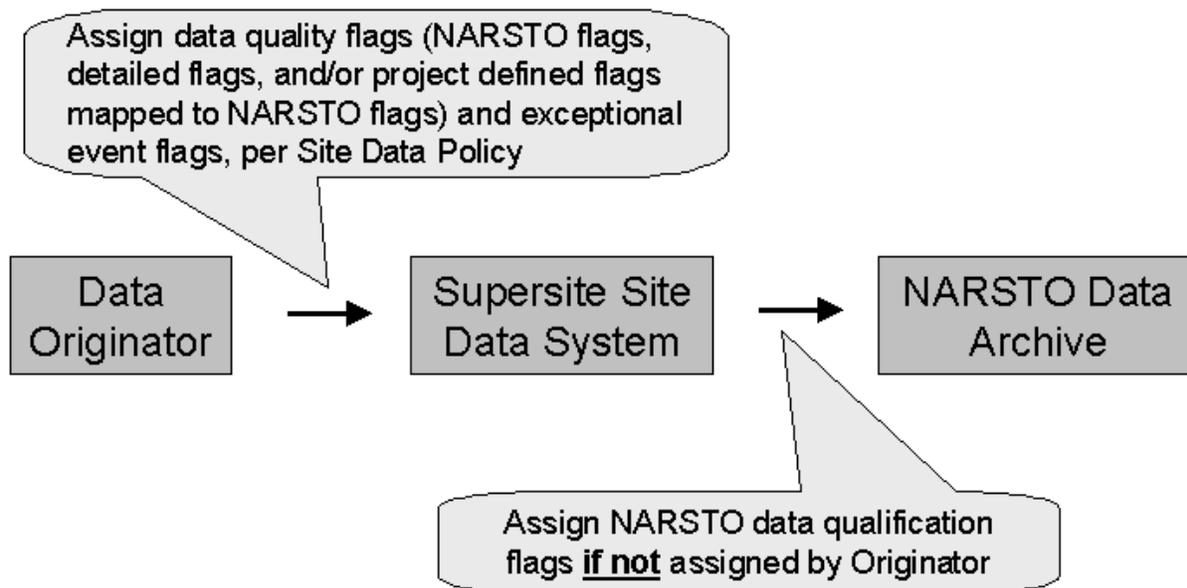
SCOPE

This standard will be used by the U.S. EPA Supersites Program as the framework for Projects to implement their site-specific data flagging processes. This standard is a resource document for Projects to use as they determine the most appropriate data flagging approach. The development of this standard has been guided by the requirement that Project data will ultimately be submitted to the NARSTO Permanent Data Archive (PDA) and must have assigned NARSTO data qualification flags (Attachment 1).

Through the implementation of a Project's flagging approach, reported data values must be assigned at least one data quality flag by the data originator that indicates to a data user whether the data are valid without qualification, valid but qualified/suspect, or invalid due to serious sampling or analysis problems. These flags may be the NARSTO data qualification flags or other flags as defined by a Project. Project-defined flags must be mapped to NARSTO flags (Attachment 1) before sending the data for archiving by NARSTO.

A Project **may** adopt this standard in its entirety, or use only parts of it, depending on the Project's needs. Described are (1) the minimum implementation, the NARSTO data qualification flags (Attachment 1), (2) a more comprehensive list of Supersite detailed data quality flags that **may** be used by a project to flag individual atmospheric monitoring and analytical laboratory measurement results (Attachment 2), and (3) a set of exceptional event flags (Attachment 3) that a project **may** use to flag a sample or set of samples. A reference list of the Supersite detailed data quality flags will be maintained on the referenced web site and a mechanism for requesting changes or additions to the list is described.

Although either the NARSTO data qualification flags or the Supersite detailed data quality flags might adequately qualify measurement values, the detailed quality flags provide much more information about the nature of the qualification and would be the preferred flag to be assigned by the Supersites data originator. Exceptional event flags may be assigned to a sample as needed to indicate the possible influence of local or larger scale events that may impact the representativeness of the sample.



SUMMARY

Each reported data value must be accompanied by one or more data quality flags assigned by the Supersites data originator that clearly indicates whether that value is valid or invalid. Detailed data quality flags must map to the NARSTO data qualification flags. Exceptional event flags may be assigned to a sample as needed to indicate the possible influence of local or larger scale events that may impact the representativeness of the sample.

SPECIFICATIONS

These are the specifications for data originators to follow when applying data quality flags. Recommended guidelines and optional procedures are in the next section.

1. A data quality flag assigned by the data originator must accompany each reported data value.
2. NARSTO data qualification flags (Attachment 1) **must** accompany data submitted to the NARSTO Permanent Data Archive (PDA). Only one NARSTO flag may be assigned to a result.
3. The data originator **may** assign either the NARSTO data qualification flags (Attachment 1), the Supersite detailed data quality flags (Attachment 2) or both, to an individual value or under certain circumstances to a sample. However, Site data management policy **may** require data originators to use a certain set of flags (e.g., Supersite detailed flags or others).
4. The data originator should also assign exceptional event flags (Attachment 3, AIRS Exceptional Event Flags), as needed, to a sample or set of samples based on their observations of local or larger scale events that may impact the representativeness of a sample and the validity of any analysis results for the present study.
5. If the data originator assigns NARSTO data qualification flags to their data, these codes and definitions cannot be modified.

6. The Supersites Data Management Working Group will compile and maintain a list of detailed data quality flags and their definitions for use by data originators. The 3-character flag codes and their definitions should not be modified. Investigators are encouraged to select flags as appropriate from this reference list. See Attachment 2.
7. Additions or other changes to the preferred Supersite detailed data quality flag list will be made by consensus of the Data Management Working Group. Refer to "Changes and Additions to the Detailed Flag List" later in this document.
8. The exceptional event flag codes and definitions (Attachment 3) are U.S. EPA AIRS flag codes and definitions and cannot be modified.
9. The flag fields will be identified per the "Variable Naming Consensus Metadata Standard". Each flag field will be maintained as a separate element in the database / data file, distinct from the measured data value.
10. If the data originator chooses to use quality flags other than (or in addition to) the Supersite detailed data quality or NARSTO qualification flags, the codes must be different, the meaning of the data originator's flags must be documented in site quality assurance documentation, and the mapping to NARSTO data qualification flags must be provided.
11. Reference this standard if data quality flags from the NARSTO list (Attachment 1), the Supersite detailed flag list (Attachment 2), and/or the exceptional events flags (Attachment 3) will be used. Indicate deviations, if any, from these sets of flags.

GUIDANCE

1. Either the NARSTO data qualification flags or the Supersite detailed data quality flags might adequately qualify measurement values. The detailed quality flags potentially provide much more information about the nature of the qualification and would be the preferred flag to be assigned by the Supersites data originator. However, the NARSTO flags may, if permitted by the Site's data management policy, serve as the data originator flag as well as the final NARSTO data archive flag, because they also clearly distinguish among valid, valid but qualified, and missing/invalid values.
2. If a measured value is flagged or coded as invalid or no result was reported, that data value will be replaced with a missing value code before being sent for NARSTO archiving. No invalid or null data values are permitted in the NARSTO archive.
3. Multiple Supersite detailed flags may be assigned to a given data value as needed.
4. If multiple detailed flags are used, the most important flag is to be placed first, followed by the second most important, etc. This order may be used later to determine which single NARSTO data qualification flag will be assigned.
5. NARSTO and detailed flags may be applied to either an individual measurement value or to an entire set of measurements on a sample (i.e., an individual spreadsheet cell or an entire spreadsheet row, respectively). In the latter case, a set of measurements on a sample may be flagged if the measurements are of the same parameter. Examples of sample measurements that might be flagged in this manner include results of single particle continuous measurements (e.g., ATOFMS, SMPS) and of particle scanning techniques (e.g., FTIR).

6. Exceptional event flags may also be applied to a set of measurements on a sample and would be separate from any detailed or NARSTO flags. The set of measurements may include any mix of measured parameters.

IMPLEMENTATION

This standard is the framework within which Supersite Projects will implement their specific data flagging processes. The principal requirement is that data values submitted to the NARSTO PDA have assigned NARSTO data qualification flags (Attachment 1). This standard can be directly implemented by Supersite Projects to enable them to meet this requirement, while also allowing use of comprehensive lists of pertinent flags. It does, however, have sufficient flexibility to allow projects to define various infrastructures and even different flag values within this structure to meet their individual site's data collection, processing, and management needs while meeting this requirement. Specific project activities relating to this implementation should be described in the site's data management policy contained in appropriate project documents.

Mapping Project-Defined or Supersite detailed Data Quality Flags to NARSTO Data Qualification Flags

If the data originator has not done so, the site data management coordinator must map project-defined or Supersite detailed data quality flags to NARSTO flags (Attachment 1) before sending the site data for archiving by NARSTO. Detailed data quality flags and corresponding NARSTO flags are paired in the detailed flag list (Attachment 2).

For data-originator-defined flags and for multiple flag assignments, the site data management coordinator must ensure that the flags map to an appropriate NARSTO data qualification flag.

The site data management coordinator may include all project-defined and Supersite detailed flags and the corresponding NARSTO data qualification flag in the Data Exchange Standard file to minimize the possibility of confusion.

Flagging Individual Measurement Values.

NARSTO flags **must** and detailed flags **may** be applied to individual measurement values and in either case, will be separate values (e.g., different spreadsheet columns). (An exception involving sample-level flagging is identified below.) At a minimum, two spreadsheet columns, the value and the NARSTO flag, are needed for each parameter. If the Supersite detailed flags are added, three columns will be needed.

Flagging Measurements Values for an Entire Sample

With certain types of measurement data, rather than flagging individual measurement values with the NARSTO and Supersite detailed flags, the set of measurements on a sample may be analogously flagged if the measurements are of the same parameter. Examples of sample measurements that might be flagged in this manner include results of single particle continuous measurements (e.g., ATOFMS, SMPS) and of particle scanning techniques (e.g., FTIR). The NARSTO flag, and optionally the detailed flag, would be entered in columns that precede the sample measurement values and that are appropriately identified as applying to the entire sample/row. No individual value flags would be necessary.

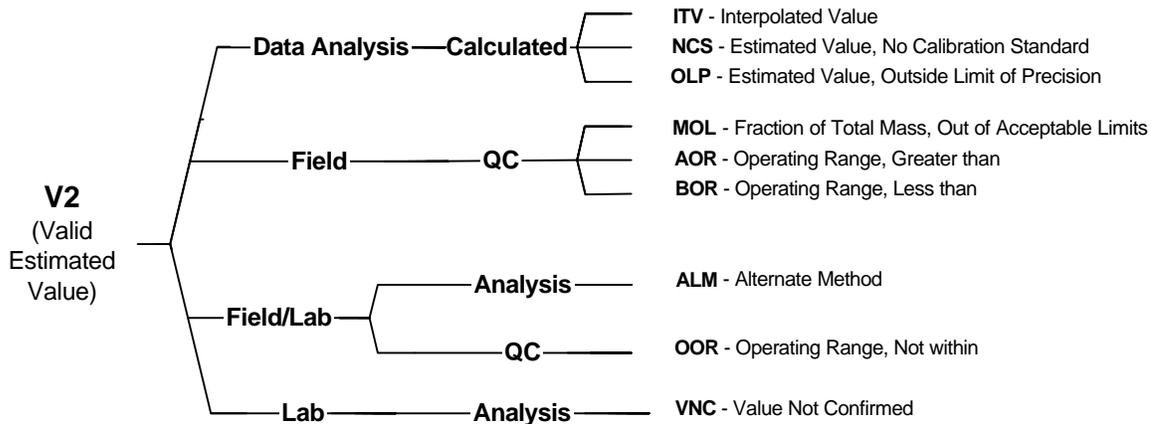
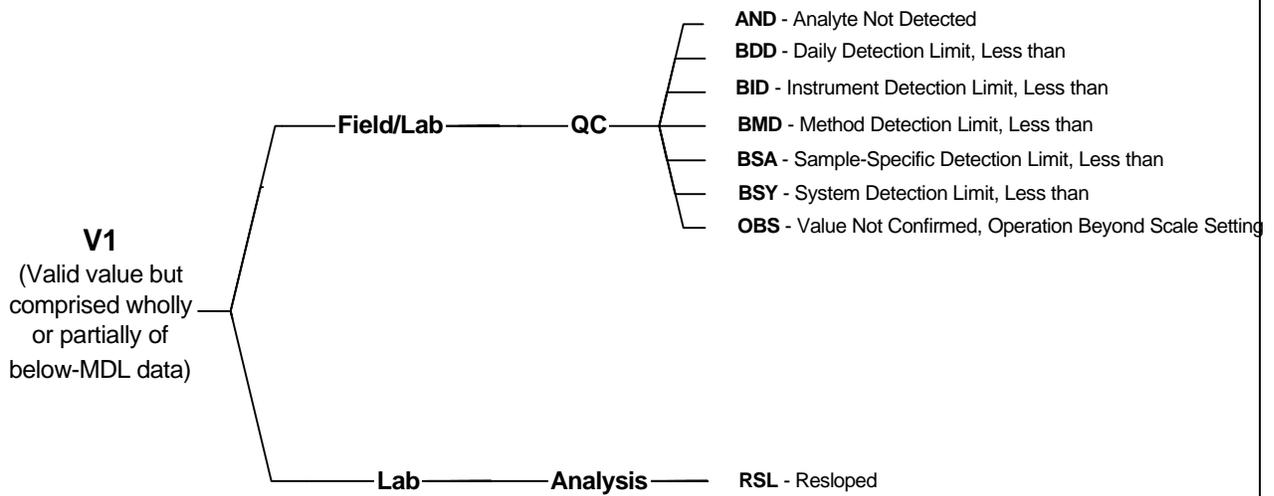
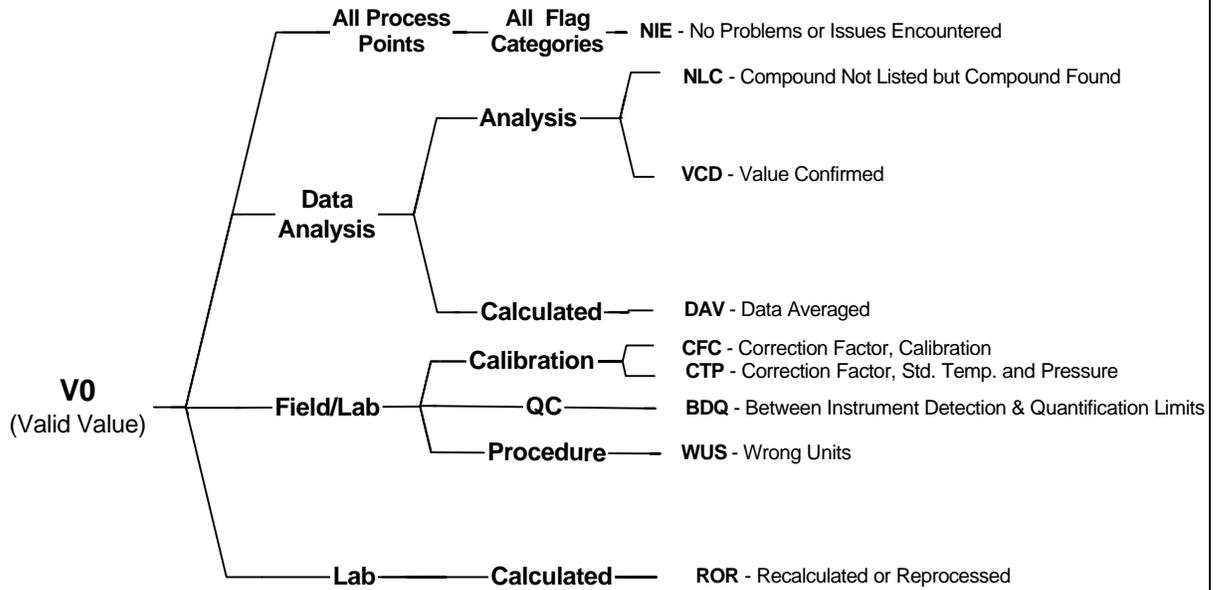
If the results for the set of measurements on a sample include both valid and missing values, the appropriate NARSTO valid flag should be assigned to the sample/row. Missing results must have the defined missing value code for that column. Contact your Site Data Coordinator for assistance with a specific implementation.

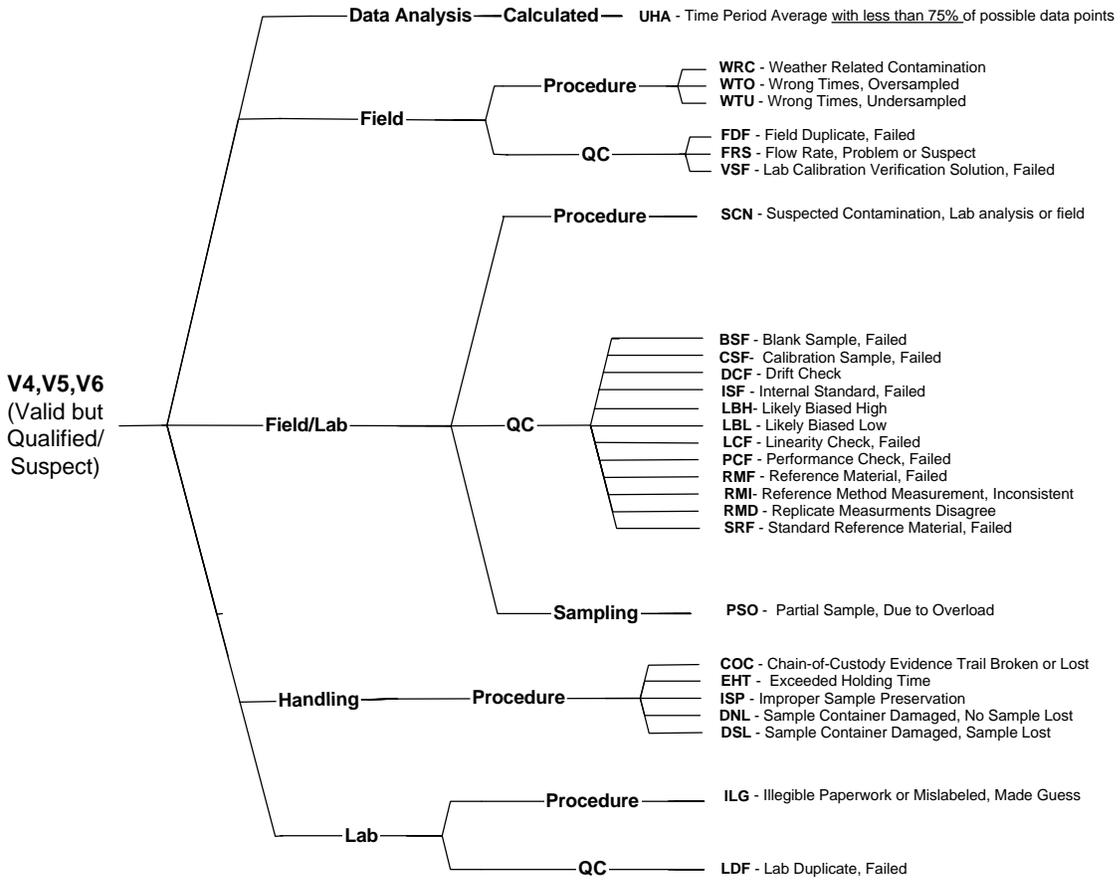
Assigning Exceptional Event Flags

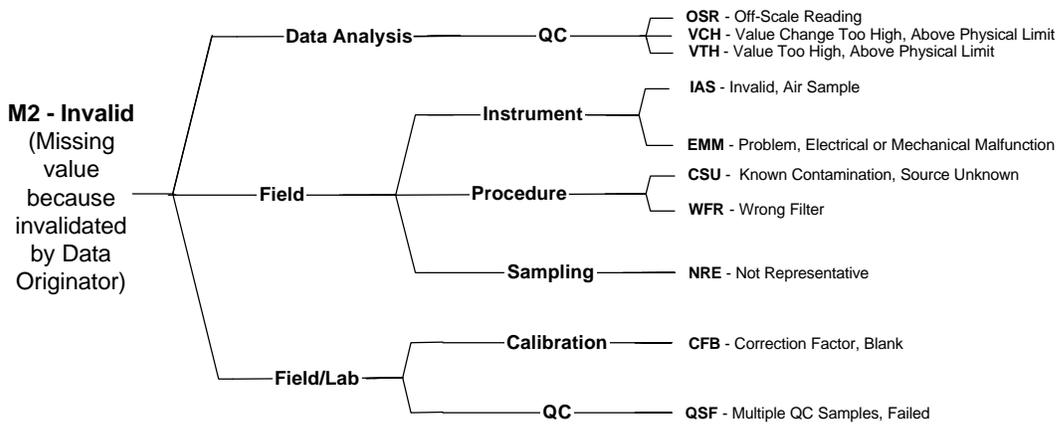
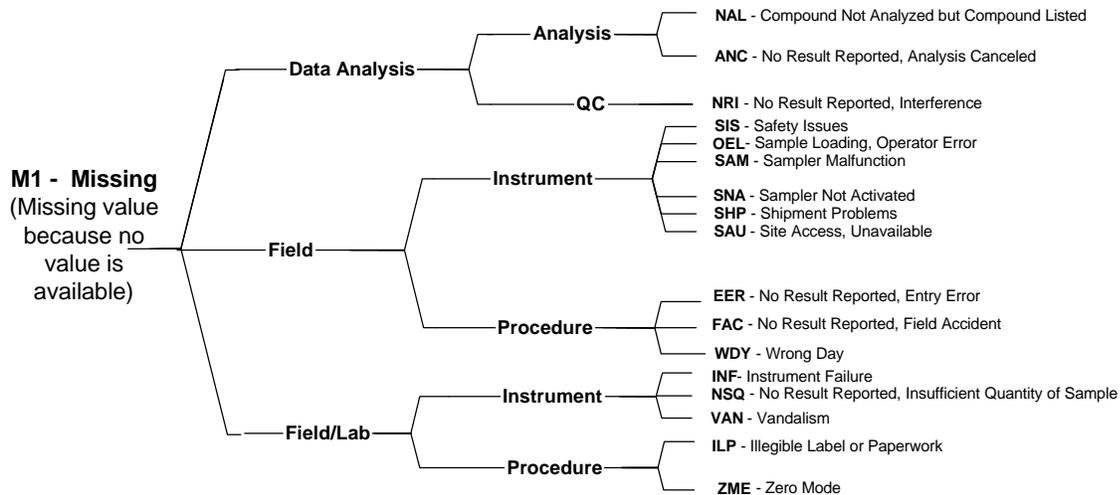
A project *may* use exceptional event flags (Attachment 3) to flag a set of measurements. Exceptional event flags may be assigned to a sample as needed to indicate the possible influence of local or larger scale events that may impact the representativeness of the sample. The flag would be entered in a column that precedes the sample measurement values and is appropriately identified as applying to the entire sample/row.

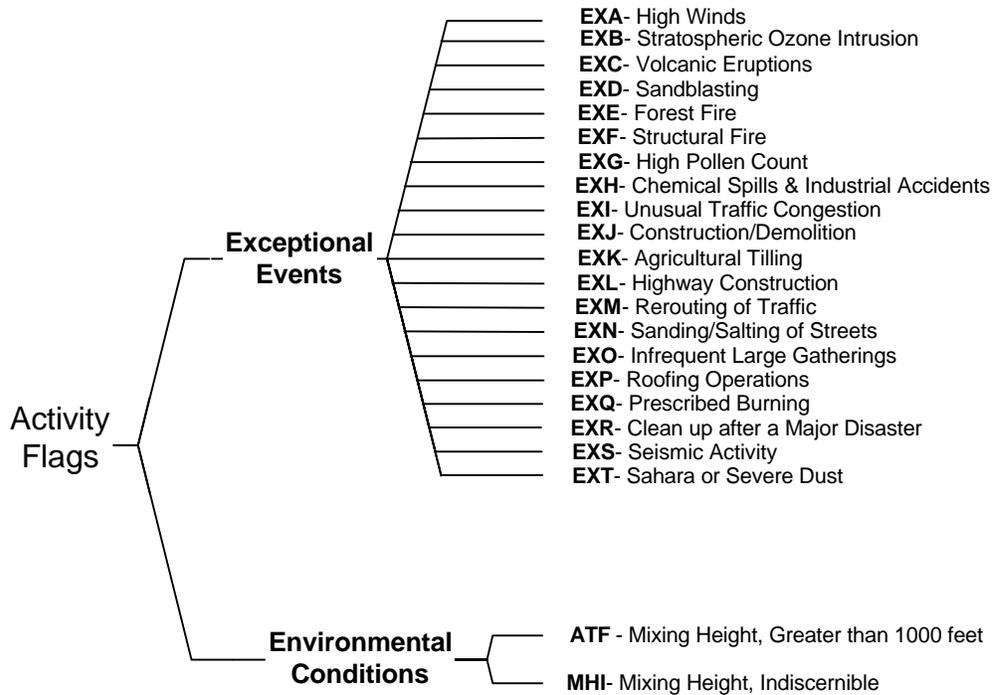
Supersite Detailed Data Quality and Exceptional Event Flags

The list of Supersite detailed data quality flags, their definitions, useful categorizing fields, and the corresponding NARSTO qualification flags are maintained on the Supersites web site in spreadsheet format (Excel 97, and also comma-separated value (.csv) format). Below is a hierarchical presentation of the detailed flags: subdivided first by the higher level NARSTO flags; next, taking into consideration the process points or generally where sampling, measurement, and analysis activities occur; and then the categories of quality checks and evaluations that are performed and that lead to assigning a particular flag. **Based upon these evaluations, flags might be assigned to a single measurement value, all results of a particular analysis, selected results across a set of samples, or all results for a sample.** In the final diagram, the exceptional event flags can be used to document the possible influence of local or larger scale activities that may impact the representativeness of a sample.









Changes and Additions to the Supersite Detailed Flag List

Site investigators and data users are encouraged to work with their Site Data Management Coordinators to suggest improvements in and additions to the list. Additions or changes to the Supersite detailed flag list or explanatory information should be recommended by a Site Data Management Coordinator to the Data Management Working Group for discussion and consensus.

Assigning Multiple Flags

Investigators may assign multiple detailed quality flags to a given data value as needed to convey an accurate sense of the "quality of a value" to users. If possible, the investigator is encouraged to check for another, perhaps more general flag that would be sufficient. If an adequate flag does not exist, consider suggesting a change or addition to the preferred list.

Separate multiple flags by a semicolon (;) without a space. Example: WEA;REP

When assigning multiple data originator flags, place the most important flag first, followed by the second most important, etc. Consider that only one NARSTO data qualification flag will be allowed when the data are permanently archived. Multiple flags must be resolved to a single NARSTO flag. If NARSTO data qualification flags are not also provided, the order of the detail flags may be used by the Site Data System to determine the appropriate single NARSTO flag.

Applying Flags to Averaged Data

Either of two flags in the preferred Supersite detailed list will apply to most of the averaged data to be reported: "DAV" or "UHA". These flags indicate, respectively, whether the percentage of valid values in the average is greater than or less than 75% of the possible values for the averaging period. The "DAV" flag maps to a NARSTO flag of V0 (valid) while "UHA" maps to V2 (valid estimated).

The minimum number of valid values that must be part of an average is not specified. The individual researcher must decide when to invalidate an average value (e.g., perhaps < 50%).

These flags should be applied to the results of data that are being averaged for the first time. For example, instruments may have a measurement frequency of every 1 minute but report only an hourly average. The 1-minute measurement results are not saved, but the instrument "knows" if any are "missing" and flags average values with less than 75% valid measurements. Sites may want to recommend retaining high-resolution data or, alternatively, capturing the number or proportion of valid non-missing measurements contributing to the averaged value.

Additional quality flags may be applied as noted in the section, Assigning Multiple Flags.

If the desired averaging period changes, ideally the investigator would return to the original measurements, aggregate the data over the new time period, and apply the "DAV" or "UHA" flags based on their defined criteria.

There may be times when it is necessary/desirable to average averaged values. No invalid values should be used in such subsequent steps. Their values should be set to missing. Flagging these averaged average values may be problematic.

- If the number or proportion of valid non-missing values was retained with the original averages, the percent of valid values for the new averaging period can be calculated and "DAV" or "UHA" applied.
- Consider applying "DAV" and "UHA" based on the flagging of the original averages. Using analogous criteria, if more than 25% of the original averages were flagged "UHA" then the new average is flagged "UHA" also. If less than 25% were "UHA", then flag the new average "DAV" -- the newly derived average is composed of at least 75% original averages that had at least 75% of the total data points possible for the original averaged time period.

REFERENCES

Great Lakes Environmental Monitoring Database (GLENDa), 2000, The US EPA Lake Michigan Mass Balance Project. (<http://www.epa.gov/glnpo/lmmb/database.html>) and (<http://www.epa.gov/glnpo/lmmb/metadata.html>).

NARSTO Data Management Handbook (ORNL/CDIAC-112/R2), 2000, S.W. Christensen, T.A. Boden, L.A. Hook, and M.-D. Cheng. (Preparation and electronic publishing by NARSTO Quality Systems Science Center) (<http://cdiac.esd.ornl.gov/programs/NARSTO/>). Oak Ridge National Laboratory, Oak Ridge, TN, USA.

Supersites Consensus Metadata Standard: Variable Naming (current version at <http://cdiac.esd.ornl.gov/programs/NARSTO/metadatastandards>)

DEVELOPMENT BACKGROUND

The EPA Supersites program recognized, in March 2000, a need for standardization of key metadata, including data quality flags. The Supersites Data Management Working Group has discussed flags in many conference calls (see <ftp://cdiac.esd.ornl.gov/private/narsto/ssdmwg/minutes>) The preferred flag list was derived from the Great Lakes Environmental Monitoring Database (2000) and many of the flags and descriptions were extracted by CCAQS and directly incorporated into this list. The Working Group decided to include the U.S. EPA AIRS Exceptional Event flags into this standard because (1) many Sites will be making extensive use of AIRS data, (2) the information content of such a sample-level flag is useful for interpreting present study data collected during an event, and (3) future evaluations of data collected during events may be instructive. The standard was sent for PI review on 20010309 and approved 20010410 by the Working Group.

Send comments to Les Hook (hookla@ornl.gov)

Attachment 1. NARSTO Data Qualification Flags

Flag_Type	NARSTO_Flag_Code	NARSTO_Flag_Definition	NARSTO_Flag_Description	Applicability
NARSTO	V0	Valid value	Flag valid data values.	Applies to all measurement data types.
NARSTO	V1	Valid value but comprised wholly or partially of below-MDL data	Applies to both single and averaged data values. Measured value reported, even when below MDL. Define MDL (minimum detectable limit).	Applies to all measurement data types. See related V7 flag.
NARSTO	V2	Valid estimated value	Can apply to calculated values, approximate/out-of-range values, and measured values with, for example, EPA "J" flag (see footnote).	Applies to all measurement data types, model input and output data products, and gridded data products.
NARSTO	V3	Valid interpolated value	Valid interpolated value. Provide interpolation method in documentation.	Applies to all measurement data types, model input and output data products, and gridded data products.
NARSTO	V4	Valid value despite failing to meet some QC or statistical criteria	Apply this flag based on evaluation of field and laboratory QC sample data and subsequent statistical outlier tests on the entire data set.	Applies to all measurement data types.
NARSTO	V5	Valid value but qualified because of possible contamination	Apply this flag for possible contamination of blanks and regular samples.	Applies to all measurement data types.
NARSTO	V6	Valid value but qualified due to non-standard sampling conditions	Provide description of sampling conditions or variance from SOP in documentation.	Applies to all measurement data types.
NARSTO	V7	Valid value but set equal to the detection limit (DL) because the measured value was below the DL	Applies to both single and averaged data values. The measurement was below DL and the Principal Investigator lacks confidence in it and the DL was substituted in its place. Define MDL (minimum detectable limit).	Applies to all measurement data types. See related V1 flag.
NARSTO	M1	Missing value because no value is available	Use this flag when no result was reported. Identify in documentation the missing value code that is used in the result field.	Applies to all measurement data types.
NARSTO	M2	Missing value because invalidated by Data Originator	Use this flag when the result reported to a site database was invalid. Invalid results are not sent to the NARSTO archive. Identify in documentation the missing value code that is used in the result field.	Applies to all measurement data types.
NARSTO	H1	Historical data that have not been assessed or validated	Use this flag when, for example, historical data may have been used for preliminary characterization or range finding purposes. It will not be used in subsequent analyses but is part of the project record.	Applies to all measurement data types, model input and output data products, and gridded data products.

Reference the NARSTO Data Management Handbook for NARSTO flags.

For EPA CLP flags reference URL: <http://www.epa.gov/superfund/programs/clp/guidance.htm#org>

USEPA Contract Laboratory Program, National Functional Guidelines for Organic Data Review, EPA-540/R-99-008 (PB99-963506), October 1999.

USEPA Contract Laboratory Program National Functional Guidelines for Inorganic Data Review EPA-540/R-94-013 (PB94-963502), February 1994.

Attachment 2. Supersite Detailed Data Quality Flags

The Supersite detailed data quality flag list is located at <http://cdiac.esd.ornl.gov/programs/NARSTO/metadastandards> and named SS_detailed_flags_yyyymmdd.xls.

The Excel 97 spreadsheet prints on approximately 6 legal size pages.

Attachment 3. Exceptional Event Flags

The exceptional event flag codes and definitions are U.S. EPA AIRS flag codes and definitions for compatibility.

U.S. EPA AIRS Database, Air Quality Subsystem (AQS)
Exceptional Event Flags (Re-Engineered AQS, <http://www.epa.gov/ttn/airs/>)

Exceptional_Event_

Supersite Flag Code	AIRS Flag_Code	Code Definition: The reason for an abnormal observation.
EXA	A	HIGH WINDS
EXB	B	STRATOSPHERIC OZONE INTRUSION
EXC	C	VOLCANIC ERUPTIONS
EXD	D	SANDBLASTING
EXE	E	FOREST FIRE
EXF	F	STRUCTURAL FIRE
EXG	G	HIGH POLLEN COUNT
EXH	H	CHEMICAL SPILLS & INDUST. ACCIDENTS
EXI	I	UNUSUAL TRAFFIC CONGESTION
EXJ	J	CONSTRUCTION/DEMOLITION
EXK	K	AGRICULTURAL TILLING
EXL	L	HIGHWAY CONSTRUCTION
EXM	M	REROUTING OF TRAFFIC
EXN	N	SANDING/SALTING OF STREETS
EXO	O	INFREQUENT LARGE GATHERINGS
EXP	P	ROOFING OPERATIONS
EXQ	Q	PRESCRIBED BURNING
EXR	R	CLEAN UP AFTER A MAJOR DISASTER
EXS	S	SEISMIC ACTIVITY
EXT	T	SAHARA OR SEVERE DUST

[Formerly called Validity Flag in AIRS AQS
(http://www.epa.gov/aqspubl1/validity_flag.html)]