In Reply Refer To: July 12, 2002
Mail Stop 412

Office of Water Quality Technical Memorandum 2002.15

Subject: Use of the new data-quality-indicator (DQI) field in NWIS 4_1

DESCRIPTION OF THE PROBLEM

The *data-quality indicator* (DQI) is a new field added in the NWIS 4_1 release for marking the review status and quality of water-quality results. Detailed requirements for how the software processes and uses the new DQI field are described at the URL listed below. The detailed requirements also describe two additional new fields in the NWIS water-quality system, *value qualifiers* and *result comments*, which are also discussed in this document.

http://ok.water.usgs.gov/nawqa/phoenix

A consistent approach to setting the DQI must be followed by all users of NWIS so that non-local and non-USGS data consumers obtain data from the USGS that are useful and understandable. USGS managers and users of the NWIS database software need guidance for setting the DQI to achieve the goal of consistent nationwide implementation. The DQI was implemented to achieve two purposes:

1. **Prevent wide dissemination of poor-quality measurement results.**

   Data quality is not a categorical variable; however, the DQI presents a limited list of options. Measurement results may be suitable for some purposes and not for other purposes. In the most basic sense, it is useful to track an attempt to make a measurement for logistical and planning purposes, even when the measurement attempt failed to create a result useful for describing environmental conditions. In the case of measurement results affected by contamination in the field or the laboratory, the results indicate a maximum concentration that might have been in the environment at the time and place of sampling, although the magnitude of the result is biased high. These two examples show how the perceived quality of a result depends upon the purpose for which one retrieves data from the system.

   The new DQI provides a mechanism for gross screening of results during retrieval. Most results will be available to all retrievers; some results will be included or excluded based upon a gross-data-quality selection option made by the retriever; and some results will be excluded from retrievers with limited access.

2. **Prevent accidental overwriting of reviewed measurement results.**

   Users of the NWIS water-quality system have been plagued by alterations of measurement results that have been previously stored, retrieved, and reviewed. The data alterations have commonly been caused by batch reloading of results from the laboratory or another source.

   The DQI provides the ability to protect individual measurement results from being modified by the batch-update software. Users will have the ability to mark results as reviewed, which will prevent the batch-update software from changing these results unless the protection is overridden.

The successful implementation of data-quality indicators requires:

1. creation of data models and tools within the database software that provide the abilities to store and use data-quality designations resulting from the review process;
2. understanding of which DQI designations will be used for various purposes;
3. agreement on the circumstances when the DQI designation of a result is changed; and
4. painless integration of a DQI designation into water-quality data-reduction and analysis tasks.

Existing NWISWeb aggregation processes do not use the DQI as one of the means of filtering data. The current data-exclusion criteria for NWISWeb are described on the web page: http://ok.water.usgs.gov/nawqa/phoenix/reload/nwisweb_screens_summer02.html .
The use of DQI as a data-retrieval filter was implemented in NWIS release 4_1 and is planned for NWISWeb as described in this memorandum. NWIS release 4_2 changes for water-quality data processing also are described herein. Users of the database should assign DQI with the knowledge of how this information can be used currently, as well as how it will be used in the future.

ASSUMPTIONS REGARDING DATA-QUALITY DESIGNATIONS

This memo describes the thought processes and workflows involved in setting and using the DQI. Some fundamental assumptions must be identified to explain the new workflow.

Poor-quality results are flagged to prevent inadvertent retrievals of NWIS water-quality data that would lead to incorrect interpretations by remote and (or) relatively uninformed data users. Since data quality depends upon the interpretative purpose, we will assume that the purposes of environmental-data retrieval are:

1. to assess the temporal and spatial extent of constituents in the environment.
2. to determine basic statistical properties (maximum, minimum, median, range, etc.) of environmental constituent concentrations.

An analyst can make a decision to accept or reject results after data review has been completed. The data-quality indicator will be used to track the review status. Quality-control-sample data (blanks, spikes, replicates, blind samples, etc.) are retrieved for making inferences regarding the quality of corresponding environmental data. The DQI may be set for quality-control sample results. The purpose for setting DQI in these cases is to avoid erroneous interpretations about environmental data quality made from dubious quality-control sample results.

All retrievers will receive measurement identifiers, values and remark codes, and will have sufficient information to interpret these data. However, not all retrievers will retrieve value qualifiers, reporting limits, measurement method, or result comments; nor will all retrievers have the ability to retrieve and interpret internal memoranda that document changes in methods, data-quality problems discovered ex post facto, and sampling-network designs.

The USGS Water Resources Discipline uses nearly all available distributed water-quality data in national data "warehouses" and for web-based retrieval systems. National aggregation techniques automatically exclude data that are marked as "proprietary" [http://ok.water.usgs.gov/nawqa/phoenix/policy/proprietary_def.html].

NWIS users have the option to select which permutations of DQI-flagged results shall be included in retrieved data. Exclusion of "poor-quality" results from retrievals is the default, but users have an option to obtain results flagged via DQI as poor quality ('Q'). NWISWeb needs to provide this capability in the future. It shall be the users' responsibility for interpreting whether poor-quality results are appropriate and how these results should be interpreted. NWIS retrieval software provides exclusion criteria for data that are marked as "Awaiting review." or "Unapproved method" (release 4_2), although NWISWeb does not yet have these capabilities (results from unapproved methods are currently removed from NWISWeb using a parameter-method code list).

CONCLUSIONS ABOUT DATA-QUALITY DESIGNATIONS

1. Data should have the DQI flag set to "poor quality" when it can be determined that the results are misleading for the purposes of describing the spatial and temporal extent of the constituent or the range of environmental concentrations.
2. All results identified as "poor quality" via DQI also need additional information stored describing why the quality was considered poor. The additional information may be stored as a remark code, a qualifier code, a sample comment, and (or) a result comment.
3. Other situations where quality cannot be determined to be poor ("rejected") can and should be identified with additional information in the database, specifically using value-qualifier codes, sample and result comments.
4. The following specific scenarios have been identified as situations where results can be rejected (qualified as poor quality) using the DQI.

   Values affected by contamination.
   Sources of information: Laboratory set blanks, field blanks and field notes.
   Additional documentation: Contamination in laboratory blanks is identified by the laboratory with the 'v'-value-qualifier code. Systematic contamination should identified by the field analyst adding the 'V'-remark code (OWQ Technical Memorandum 97.08) to the affected result records.
   Example: Laboratory-set blank data indicate measureable contamination for the analyte for the set-blank sample associated with the field-sample set. A field-result comment is added that identifies the contamination problem, and the value-qualifier code 'v' is applied.

Values affected by either high or low bias that exceeds method variability.
Sources of information: Blind samples, laboratory set blanks and spikes, laboratory surrogates, documented SOP variations. (SOP variation does not, by itself, constitute a good reason for rejection of results. However, in conjunction with other supporting information, a pattern of SOP variation may be used to define a collection of results affected by bias).
Additional documentation: The laboratory will report a result affected by a variation from the Standard Operating Procedure (SOP) using the value-qualifier code 'l'. Field users should identify SOP variations using the value-qualifier code 'f'. The SOP problem and(or) the reason for the DQI assignment should be described in result or sample comments.
Example: Surrogate recovery for one schedule of a sample is poor. Result comments are added for these results to indicate that the results were rejected due to poor surrogate recovery.

Compound-identification uncertainty.
Sources of information: Rerun results.
Additional documentation: The laboratory will report a rerun result that was "non-confirmable due to interference" using the value-qualifier code 'u.' Or, the laboratory reports a rerun result where "different bottle results do not match" using the value-qualifier code 'y.'
Example: After a laboratory rerun, one of the results is a detection, and the other result is a non-detection, and holding times are within tolerance. The result is rejected due to the uncertainty, and the rerun results are annnotated in the field-result comment.

Excessive variability in numeric results.
Sources of information: Blind samples, standard-reference samples, surrogate recovery, laboratory spikes, and field spikes.
Additional documentation: The laboratory will supply value-qualifier code 'm' to indicate a consistently highly-variable compound (sometimes referred to as a method flake) with questionable precision. The laboratory will apply value-qualifier code 'w' to indicate high variability, questionable precision, and (or) accuracy.
Example: All results for "poorly performing compounds" of a method are flagged with the value-qualifier 'm' and the relevant Open-File Report or technical memorandum is cited in the result comment.

Sample-identification uncertainty.
Sources of information: Labeling errors, strong evidence of laboratory mixup, expected sample type not received by laboratory.
Additional documentation: Sample or result comments.
Example: Laboratory results for all the major ions indicate a water type that is incompatible with previous sampling and the geologic environment. All results for the major ion schedule are rejected and result comments are added to acknowledge the probable bottle switch.

5. There are two scenarios where the DQI should not be used to designate poor-quality results. In these scenarios, value qualifiers and result comments can and should be used to annotate the results, but the results should not be automatically excluded from retrievals using the DQI. In these two scenarios, one avoids prejudging the potential

utility of environmental results in "gray areas" where the data quality is suspicious, but the mechanisms of cause and effect are less certain.

Numeric result seems unusual, but there is no documented cause of bias or variability.
Sources of information: suspected poor quality from statistical analysis, spatial comparisons, graphical presentations, ratio computations.

Numeric result seems normal, but there were known deviations from standard protocols.
Sources of information: suspected poor quality from field notes, analytical worksheets, techniques, equipment, or personnel.

6. Alternatives are available for storing evaluations of data quality.

Table 1.--Alternatives for storing data-quality evaluations. [A new null-value qualifier ('x') is implemented at NWIS release 4_2 to indicate that a result failed quality-assurance review.]

Alternative	Implications for NWIS	Value shown on the Web?
Result comments and (or) value qualifier	Measurement attempt and value stored	Yes
DQI = 'Q'	Measurement attempt and value stored	Today: Yes Future: Optionally available, but not retrieved by default
DQI = 'I'	Measurement attempt and value stored	Today: Yes Future: No
Null-value qualifier	Measurement attempt stored	No
Deleted result	Nothing stored	No

The following examples show when these approaches may be applied.

- Nutrient sample results were not analyzed within the SOP holding time. Review of the data indicates bias in the nitrite and nitrate measurements. The numeric result is nullified and a 'u' null-value qualifier is stored.
- Anomalous analytical results are returned for a sample that do not correspond well with previous results for the site. A dilution error is suspected, and the DQI for the sample results is set temporarily to 'I' while a rerun is being processed.
- A laboratory pH value varies substantially from the field-measured pH. After consulting with the field technician, it is determined that the meter was not calibrated properly. The field-pH result is nullified and an 'e' null-value qualifier is stored.
- Specific-conductance values for some samples collected by a contract-observer vary substantially from historical data for a surface-water site. An interview with

the observer yields the admission that not all samples were collected from the flowing part of the stream. All results for these samples are deleted.

7. Another potentially confusing DQI scenario occurs when quality-control sample results indicate poor-quality environmental-sample results. The DQI of the environmental-sample results should be modified as a result of interpreting the quality-control sample results. Quality-control sample results should only be identified as poor quality due to unacceptable bias and variability of the quality-control sample results; not because interpretation of the quality-control results indicates poor-quality environmental data.

WORKFLOWS ASSOCIATED WITH DATA QUALITY

There are four general categories of data and corresponding DQI-aging progressions. In each case, there is no implicit assumption that the originally assigned DQI code will be changed within any particular timeframe. However, the benefits of changing the DQI code (overwrite protection and limits to dissemination) will behoove reviewers to utilize the system capabilities.

Table 2.--General cases of the aging of DQI coding. [Special cases will occur, such as an ex post facto review of QC data that causes a DQI change from ‘R’ to ‘Q.’]

Data classification	**Initial DQI assignment**	**Reviewed DQI assignments**
Routine: Most data as they are entered, or received from the laboratory.	Assumed satisfactory (‘S’)	Reviewed & OK (‘R’)
		Reviewed & Rejected (‘Q’)
Historical: Non-proprietary data in the database at the time of DQI implementation.	Not Reported (‘A’)	Reviewed & OK (‘R’)
		Reviewed & Rejected (‘Q’)
Access legally restricted: Data sealed by law, court order, government attorneys, or cooperative agreement; --or--, Data purchased with a license agreement that limits our distribution.	Proprietary, Not Reviewed (‘P’)	Proprietary: Reviewed & OK (‘O’)
		Proprietary: Reviewed & Rejected (‘X’)
Access withheld pending review: All uses of this code must be consistent with WRD policy on release of Water Resources Division Information Products--WRD Memorandum 99.07 Should not be used routinely, since data "awaiting review" are excluded from public access.	In Review (‘I’)	Reviewed & OK (‘R’)
		Reviewed & Rejected (‘Q’)

As data progress from the stages of initial entry to reviewed, different people and computer programs process the results. This progression is a water-quality data workflow, where the database system is integral to efficient processing. In the discussion below, program names are used to identify which computer programs are used at various stages of the workflow.

Table 3.--Programs used in Water-Quality Data Workflow.

Program Name	**qwdata Menu Options**	**Explanation**
QWLOGIN	1	Interactive program to create a new sample record and enter result data.
QWFIELD	2-1	Interactive program to enter field-measurement results for an existing sample.
QWINPUT	2-2	Interactive program to enter laboratory-measurement results for an existing sample.
QWEDIT	2-3	Interactive program to alter existing samples and edit or add results.
QWDQIFLAG	7-6	Interactive program to update DQI of existing results meeting user-specified criteria.
QWCKDQI	7-7	Interactive program inventory DQI settings of existing results.
QWCARDSIN	Env = 8-2 QC = 8-4	Batch program to update existing samples, and add or alter associated results.
QWENTER	Env = 8-3 QC = 8-5	Batch program to create or update samples, and add or alter associated results.
QWCARDSINXDQI	Env = 8-9 QC = 8-10	Batch program to update existing samples, and add or alter results protected by DQI.

1. Data are entered interactively during sample login (QWLOGIN) and given the default DQI value=‘S’ (or ‘R’ if field notes are already reviewed).
2. Data are batch loaded from the lab (QWCARDSIN/QWENTER), and assigned the DQI value stored in the batch file, normally ‘S.’ Provisional data, e.g. experimental methods, will be assigned ‘U’ in the batch file (NWIS Release 4_2).
3. For proprietary results: the batch file is modified to assign a DQI value of ‘P’ before loading (QWCARDSIN/QWENTER), or the DQI value is reset to ‘P’ after initial entry (QWDQIFLAG).
4. Data are added interactively for field measurements, e.g. coliforms, BOD (QWFIELD) or cooperator laboratory measurements (QWINPUT). The default DQI ‘S’ is assigned, or ‘R’ if measurements are already reviewed.
5. Data are reviewed by District personnel.
    a. Individual results that fail review criteria are assigned interactively the DQI ‘Q’ (QWEDIT).
    b. Groups of results that pass review criteria are assigned interactively the DQI ‘R’ (QWDQIFLAG).

6. Reload from the laboratory of data previously reviewed by District.
    a. A group of results is determined to have been transmitted incorrectly. A memorandum describing the data deficiencies is prepared and transmitted. New batch files are prepared for Districts.
    b. District personnel retrieve and process the new batch files (QWCARDSINXDQI) to update the local database.
7. Review of proprietary data by project personnel.
    a. Approved results are selected and DQI is changed from 'P' to 'O' (QWDQIFLAG).
    b. Rejected results are selected and DQI is changed from 'P' to 'X' (QWDQIFLAG).
8. Data review by NWQL, OWQ, BQS, or other focused task group.
    a. A well-defined group of analytes and methods, over a specified time period, is determined to be poor quality. A memorandum describing the data deficiencies is prepared and transmitted.
    b. District Water-Quality DBA's select the corresponding set of results and change DQI from 'A', 'S', 'R', or 'I' to 'Q' (QWDQIFLAG).
    c. Water-quality discipline reviewers may examine DQI settings in the District water-quality database. An inventory of DQI settings used in the database is prepared (QWCKDQI). The inventory summarizes results by "review" and "proprietary" status.

References:

DQI requirements-
http://wwwok.cr.usgs.gov/nawqa/phoenix/www/dqi.html

Value-qualifier requirements-
http://wwwok.cr.usgs.gov/nawqa/phoenix/www/rmk_qual.html

Result-comment requirements-
http://wwwok.cr.usgs.gov/nawqa/phoenix/www/resultcomment.html

Null-value qualifier requirements-
http://wwwok.cr.usgs.gov/nawqa/phoenix/www/nullresult.html

WRD policy on release of Water Resources Division Information Products--
WRD Memorandum 99.07

Remark Code (V) for Water-Quality Data--OWQ Technical Memorandum 97.08

Please email all questions to *phoenix@usgs.gov*.

Stephen K. Sorenson
Acting Chief, Office of Water Quality

This memorandum does not supersede any other Office of Water Quality Technical Memorandum.

Distribution: All WRD Employees

****

Stephen K. Sorenson
Acting Chief, Office of Water Quality
U.S. Geological Survey
412 National Center
Reston, VA 20192
703/648-6864
sorenson@usgs.gov