LETTING MONITORING DATA SPEAK FOR THEMSELVES

Revital Katznelson

State Water Resources Control Board, Sacramento. CA. Office: San Francisco Bay Regional Water Quality Control Board, 1515 Clay Street, Suite 1400, Oakland, CA 94612, USA

Biographical Sketch of the Author

Revital Katznelson received her Ph.D. from the Hebrew University of Jerusalem, Israel, in 1984. She has extensive experience in performing, interpreting, and assuring quality of field and laboratory analyses of chemical, biological and bacteriological water quality parameters. She also has hands-on experience in using enzyme-linked immunosorbent assay (ELISA) kits and in toxicity testing with a variety of test organisms. Dr. Katznelson has led many ecological studies in aquatic systems including creeks, ponds, lakes, marshes, lagoons, and reservoirs, and has developed innovative methods and experimental systems for her environmental research. Currently she is employed as a Regional Coordinator for the Citizen Monitoring Program of the State Water Resources Control Board, and provides technical liaison between citizen monitoring groups and various data users, including Regional Water Quality Control Board staff.

Abstract

Users of monitoring data appreciate reliable, defensible, and usable data, but the tools to communicate these quality attributes often lack clarity and consistency, especially where field activities are concerned. In the absence of unambiguous communication tools, assumptions about data quality are often made on the basis of who collected the data, whether they were adequately trained, whether they used established protocols, and whether they had an approved Quality Assurance Project Plan in place. This "programmatic" approach, which relies on external perception of merit, does not provide the data user with the relevant facts regarding the actual quality of specific data sets or individual results. A data quality management (DQM) system has been developed to provide for the primary data management functions of documentation and quality assurance in a way that allows each data point to "speak for itself." This DQM system, designed for individual monitoring projects, includes a Project File Database and instructions for its use. Essentially, the Project File is a simple Microsoft Excel workbook that provides placeholders for all descriptors of the monitoring result (i.e, the outcome of a measurement or analysis), the measurement (i.e., the identity, features, and specifications of the instrument or kit used), and the quality of the measurement (i.e., instrument-specific calibration, accuracy checks, and precision records). The Project File also contains placeholders for "data retrieval handles" that allow the user to sort, filter, and pool individual results based on the monitoring intent, sampling design (e.g., probabilistic or deterministic), station type (e.g., outfall or creek), or conditions during sampling (wet or dry weather). Finally, the results can be accompanied by qualifiers that inform the user about the range of associated error, whether they have been validated, and whether they are supported by adequate documentation. The Project File has been successfully used by citizen monitors and by agency staff.

1.0 Introduction

Knowledge about watersheds is essential for watershed management. Decision makers and regulators in charge of watershed management and protection want “usable, reliable, scientifically defensible data of known quality” to support sound decisions. In the regulatory and watershed management arena, the process of data gathering and use involves more than one person. Communication between people requires a language that is unambiguous and specific. That language is very complex, requires an enormous amount of detail, and spans multiple scientific disciplines. This paper is focused on water quality monitoring, including field measurements and laboratory analyses. It presents a suggestion for a language that lets the data users know, for each monitoring Result point, how good it is and what it represents in the environment.

2.0 Terms and expressions of data quality

What are DATA? "Data" are bits of information. In the context of water quality monitoring, “data” are the monitoring Results, i.e., the outcomes of our measurements and analyses. The Result for a specific water quality parameter is what the field operator produces and what the data user will use, and everything else revolves around that. A Result can be numerical or verbal. It can be an individual value (e.g., pH 7.5), a calculated endpoint (e.g., 3 cubic ft per second), a verbal category (e.g., murky), or a numeric range category (e.g., 25-50% canopy cover). In rare situations a result can be a narrative statement, i.e., a sentence or a paragraph. Results are used with other bits of information that describe them; some “essential” bits (e.g., sampling date and time) are often included in the term “data” while other bits are called ”metadata”, which means “data about the data”. Many descriptors are linked to a Result point through a unique entity (e.g., Station ID).

What are GOOD data? This question has been posed to a large number of people, and their responses provided a large number of answers. Table 1 shows a list of attributes compiled from people’s feedback. It is not a surprise that different people mean different things when they say "good data"; it just confirms the need to find some generally accepted terms to describe data quality. The first step can be arranging the list in a way that assigns these attributes into major groups of features or to distinct aspects, as suggested in Table 1. Is "data of known quality" equivalent to "data of good quality"? Not necessarily, but data cannot be of good quality if people do not know their quality. Known quality requires documentation, and for that people need to use words.

Because the words and terms people use for data information exchange carry a lot of significance, people that wish to communicate clearly about data quality attributes need to use one word for one meaning, avoid use of one word in more than one meaning, avoid using a word in a wrong sense, and minimize the use of synonyms to relate to the same thing. For example, the word “accuracy” is problematic when used to mean ‘realistic representation’ - as in “our sample ‘precisely and accurately’ represents the conditions in the creek”, or when used by equipment manufacturers that specify the ‘accuracy’ of the device they sell – and that statement has nothing to do with what the instrument reads in a Standard that represents the “true value”, or how it drifts from the calibrated state. Accuracy was also used to indicate that data was copied or entered correctly; the word “fidelity” can be used instead to describe the same thing. In Table 1, and the rest of this paper, accuracy is reserved to describe the measurement or analysis – and only for that.

Data reliability is often confused or mixed with other aspects of data quality, usually in conjunction with judgment about the type of instrument used to collect that data. Many people have the notion that sophisticated instruments give data that are more reliable than data collected with simple kits. It takes effort to explain that the more sophisticated the instrument the more can go wrong with it, and that without documentation that the instrument – or the simple kit for that matter – were working properly, no data is reliable.

Table 1 also provides clear distinction between measurement quality and sample integrity. This distinction facilitates the interpretation and reporting of QA/QC measures such as duplicates (which are used to assess precision), or Standard recovery records (which indicate measurement accuracy) – separately from blanks (which prove lack of contamination, but have nothing to do with the accuracy of the measurement itself). When people say that “the Result is inaccurate because the sample was contaminated”, they probably mean to say that the Result is not correct. Measurement quality and sample integrity are fundamentally distinct from how well the sample represents the environment it was collected in (i.e., “representativeness”). Sample representativeness can be shared with other people only through a complex array of information bits that describe the reasons for monitoring and for selecting the particular sampling time, place, and conditions. In other words, representativeness is about the intent and the design of the study. This topic is the central theme of Section 5 below.

What is Data Quality Management? As mentioned above, known data quality requires documentation. Documentation means keeping records of all relevant Result descriptors and this requires placeholders, or a “data management system”. Because quality depends on documentation, quality-related information has to be managed in a way that links it to the Results. The term “data quality management” captures the concept that there is no way to know the data quality without management, and that a data management system that does not link data with data quality information cannot provide data of known quality. Data quality management is different from Quality Assurance and Quality Control (QA/QC), which relies on a series of actions and tests. Essentially, a data quality management system can provide placeholders for the outcomes of QA/QC procedures, and thus link them to the data they are intended to support. The next section describes how it works.

3.0 The Data Quality Management (DQM) System and its Project File

This section describes the new Data Quality Management (DQM) system developed by the author for use by the Clean Water Team (CWT), the Citizen Monitoring Program of the State Water Resources Control Board, in California. The DQM system features an array of forms, spreadsheets, and dictionaries for information transfer and communication between operators within a scientifically-based data collection effort. It supports environmental monitoring Projects focused on water quality and related watershed information. A **"Project"** is defined as a data collection effort that is limited in space and time (e.g., routine monitoring of one creek over one year, or a special study to identify the source of a particular constituent, or a Snapshot Day monitoring in multiple locations). DQM revolves around the **"Project File",** a Microsoft Excel workbook with multiple spreadsheets that include all the Results and all the supporting documentation relevant to one Project.

Figure 1 shows how all these descriptors are linked within the DQM Project File. As mentioned above, many descriptors are linked to a Result point through an "information bit" that has a unique identification number and represents a unique entity (e.g., Sample ID, Station ID, etc.). Information bits that describe these entities are sometimes packaged in separate tables for convenience, for example the Station is described in detail in a table called "Location" by a large number of information bits such as landmark description, latitude-longitude coordinates & datum, etc. Instruments are described in their own table (called "Instrumeth" to combine instruments and methods) and also link the calibration and accuracy records to the Results through their unique Instrument ID and the ID of the Standard used for calibration.

The Project File tables are organized in the separate workbooks' spreadsheets as "flat database tables", with fields having a homogeneous content and records that are independent of each other. In addition, the Project File may contain specialized Worksheets for processing of raw data to derive an endpoint. For example, worksheets are essential for calculation of flow discharge from multiple measurements of depth and velocity. The endpoint (i.e., the Result) has to be copied into a Result Table. Worksheets can be constructed as "flat database tables" as well, or be constructed in blocks of cells with internal relationships in both dimensions. Identical blocks of cells representing different samples or tests can be stacked one underneath the other in one Worksheet. Worksheets are not intended to be compatible with other data management structures, but their output can be organized in ways that will facilitate copying and pasting endpoints into the Result table.

It must be noted that some unique identifiers established for different activities serve the same descriptive purpose regarding the Results, for example the entities Instrument ID, Lab batch ID, Toxicity batch ID, ELISA run ID, etc. serve the same purpose of linking the measurement/analysis/test/run with the Results generated by that entity. Similarly, Results are the outcomes of measurements, observations, or analyses of Samples; each entity has the same relation to the Result.

The term "Sample ID" is often used generically for both field measurements (or observations) and samples collected in a container and analyzed offsite (see Glossary for definitions), However the users will need to know if the result they see in the database, e.g., pH, was measured in the field or later in the lab. The DQM Project File can be constructed with separate fields for Sample ID, Measurement ID, or Observation ID, or it can have one field called "Sample (SMO) ID" which will be used for all three kinds, and have another field to distinguish between them. The Sample (SMO) ID represents a unique combination of time and space, and can generate numerous results for different parameters. In the realm of water quality monitoring, a visit of a monitoring team to a Station to conduct several sampling and measurement activities is a natural increment of time that can be captured by a point of time and space, even if some of the activities take longer than a few minutes or are performed more than a few minutes apart from each other. A Station-Visit is a very important entity, in fact it is the natural unit of counting when considering the power of a dataset (see Section 5 below).

Using the STORET concepts of Trip ID and Station Visit ID, the DQM suggests constructing the Sample (SMO) ID from the Station (e.g., SLC1), Trip (e.g., T3), and Visit (e.g., V1). Thus, a pH measurement conducted at Station SLC1 during the first station-visit of trip T3 could have a Sample (SMO) ID like "SLC1-T3-V1", and if a sample was collected for analysis elsewhere, a container number (e.g., A, B, C) can be added (e.g., "SLC1-T3-V1C"). Having the same Sample (SMO) ID at the base of all the activities conducted in one visit will make it easier for the data users to associate concomitantly collected results of different parameters, for example the pH value measured at the creek when the sample for ammonia was collected.

The DQM Project File contains over 250 placeholders, or "fields", for information bits that support one or more Result points. A list of fields was developed by the author in parallel to the development of the Project File. This "Information Needs List" is organized by subject matter (rather than in tables); fields are easy to find when asking "what does this bit of information describe? The Station? The Instrument? The Measurement?". Some fields' content describe an individual result, and some are provided for a group of results (e.g., packaged for a Sample, a Batch, a Dataset, or an entire Project). The "Information Needs List" has been compared to and augmented by other Metadata lists (e.g., the Water Quality Data Elements list compiled by the National Water Quality Monitoring Council, STORET, the US Army Corp of Engineers' Electronic Deliverable Format (EDF), and many other lists; references not provided).

The descriptors, identifiers, and linkages established for laboratory work have been in use by environmental scientists for decades and were easily complied from existing data management systems. However, data quality management tools for field measurements are sparse. The concept of Instrument ID was probably used by some professional field operators for a long time, but none of the agency staff and citizen groups that this author has seen before year 2000 were tracking or recording which Instruments were used to generate their Results. Introduction of the Instrument ID concept brought about a huge improvement in the reliability of citizen monitoring data and paved the way to helping groups generate data of known measurement quality, i.e., data that can speak for themselves!

4.0 Data Qualifiers: How good is it?

There are at least two levels, or two major steps, in the process of QA/QC review and data validation. The first review step is mechanistic, possibly partially automated, and should be performed “as close to the field” as possible by a person that knows the Project detail. This first step is about verifying Sample IDs, flagging eyebrow-raisers (e. g., dissolved >> total, or BOD >> COD), calculating measures of accuracy and precision, and comparing QA results to measurement quality objectives and other performance criteria specified in the contract or the QA Plan. The outcome of this first QA/QC review step is a set of data qualifiers. Table 2 shows a selection – and definitions - of qualifiers that can be attached to each Result point directly (or to small batches of results that share the same combination of all three qualifiers). Measurement error is derived in different ways for different Instruments, and the formal way to do this has to be specified in the SOP for that specific Instrument. Accuracy and precision can be combined either additively or “probabilistically”; this author suggests the additive approach to let the user know the worst-case scenario of how far the Result can be from the truth. The ranges of cumulative error offered in the menu in Table 2 appear to be appropriate for most users (and this information accompanies the Results into the central database), however the data users can always get the exact numbers by accessing the Project File itself. Most of the qualifiers shown in Table 2 can be generated easily given appropriate guidance and checklists, and do not require professional judgment (for example, assigning Validity status of “Unknown”, “Not checked”, “Not valid”, or “Valid”). However, a decision to assign a qualifier such as “Estimated” takes professional experience. The second step of the QA process involves professional judgment and can be done later in the timeline and with larger batches of data. In addition to finalizing the qualifier assignments, this review step examines how usable the Results are and whether the results actually make sense when one tries to interpret them. The four fields shown in Table 2 provide information on the data that span four independent aspects of data quality, enabling the users to pick and choose according to which aspect is most important to them.

Beyond the direct qualifiers shown in Table 2, the DQM Project File also contains a placeholder for “data use potential” based on programmatic attributes of the data collection effort. The three level item names offered as options are parallel to a set of three levels, or scores, of programmatic qualifiers developed for citizen groups (Bridget Hoover, personal communication).

(a) The lowest level, corresponds to “Minimal use of the data” which means that the data gathering was done for educational purpose only, not for use; this is equivalent to programmatic Level 1 (no planning, no protocols, demonstration rather than training, minimal record keeping).

(b) The second level, called “screening” – which means data can be used with caution and important findings need a follow-up testing - is like programmatic Level 2 (internal monitoring plan, written protocols, short training, no QAPP, sporadic calibrations, data in field data sheet and spreadsheet).

(c) The third level, called “Any use”, means that the data can be used for any suitable purpose and corresponds to programmatic Level 3 (approved sampling plan, established SOPs, multiple training sessions, approved QAPP, routine calibration, comprehensive documentation with data).

5.0 Intent and Design descriptors: What does it represent?

The DQM Project File introduces a new set of placeholders for information about the reason for monitoring and about what a group of results, packaged as a "Dataset", actually represent. A dataset is defined as a collection of results that share the same intent and design attributes, i.e., all were collected to answer one study question. The number of Station-Visits in a dataset is a measure of its statistical power. Each Result belongs to a given Sample and a given Dataset, and each Sample belongs to a given Dataset. A Station can serve different Datasets. There may be several Datasets, or studies, in one monitoring Project. Combined with information about Station type and sampling conditions, intent and design information provide "handles" for retrieving, sorting, pooling, and filtering data from a central database. Table 3 shows the Project File placeholders for intent, design, and power descriptors and how these descriptors have been applied to a number of study-question scenarios. This type of table was also a very good planning tool when given to citizen monitoring groups (together with the Lookup dictionary shown in Table 4) because it helps to focus the study question and design. Table 4 is a Lookup Table with a menu of items one can use to describe the study; it also provides definitions for most of them. Note that the lists of items that describe Station type, Station selection intent, and Sample timing intent are very short and provide just a few examples; the full list is under development and augmentation with items from other sources.

6.0 Overview of DQM implementation

Data Quality Management (DQM) system consists of **manageable pieces** of guidance and tools (= DQM materials), organized to accomplish the following:

- provide instructions tailored to the different **roles** people play within the data collection effort
- support the various **phases** of a typical monitoring Project
- enable flow of **information** between all roles, in all directions
- enable flow of **fully-documented monitoring results** from the data gatherers to the data users".

The DQM provides tools and instructions that help any person communicate information and data in a consistent way through all planning, monitoring, and data reporting tasks of a typical monitoring Project. These tools and instructions are organized for three types of operators that play three distinct "roles" in the Project, using guidance and forms or templates that are specifically tailored to each of the roles. Project roles include the Field Operator, the Trainer, and the Technical Leader. There are additional roles that the DQM materials address (the Member of the Public, the Data User, and the Technical Expert); these roles are not an integral part of the Project but persons in these roles provide input during the planning and designing phase of a Project.

Project tasks start with the Project planning phase (question formulation, parameter package & sampling design development, data quality objectives development, and monitoring methods selection). Planning aids include the "Dataset" table and a few other fields in the Project File (in a table similar to Table 3), and a template Monitoring Tasks list that defines roles and responsibilities as related to all Project tasks and deliverables.

The next phase includes all the tasks of data acquisition (field measurements, sampling , and off-site analyses). The last phase is transfer of the Results, through all the data validation and qualifying process, from the data collectors (Field operators, lab technicians) to the data users.

The DQM is based on the premise that communication of scientific contents has to be very specific. Each term or bit of information has to mean one thing only, and each word has to have only one meaning. The DQM also recognizes that operators need very detailed and clear guidance to assure clear, unambiguous communication. The DQM provides the specificity and level of detail that leaves nothing to interpretation. Another special feature of the DQM is that it is comprehensive. It contains placeholders for more than two hundred bits of information that describe one Result point.

The DQM system deals with very complex scientific content, which cannot be easily simplified, by organizing it in small pieces that are complementary to each other and support one comprehensive system. There information bits have simple, straightforward relationships, and are connected by linkages that are intuitive and conform with simple criteria (e.g., what entity does this bit of information describe"). This flexibility allows for arrangement of the information in any database structure.

The DQM Project File itself is a small, easily manageable package. It is totally transparent – no automation or macros – and can be viewed directly, but can also be combined with automated data entry forms and/or queries. Any person with minimum spreadsheet skills can use it for data entry. It can be sent by email and be easily transported via a 56K modem. These features make the file useful for low-tech and high-tech organizations alike.

7.0 Summary

The Data Quality Management (DQM) system presented in this paper provides for

- Comprehensive capture of information and effective linkages of descriptors with Results
- A language for unambiguous, specific communication
- New ways of communicating complex concepts of sampling design and dataset power in a database-ready format, using the entity "Dataset".
- Breakdown of the information into high level of detail that allows the data users, be they scientists, managers, or regulators, make very fine distinctions when they selectively retrieve, sort, or filter the data.
- Tools that can be used by any person and yield well-documented data.

Acknowledgements
This paper was supported by the California State Water Resources Control Board, division of Water Quality. The author wishes to thank Steve Weisberg, Dave Paradies, Karen Worcester, and Bridget Hoover for their inspiring comments, and the numerous State Board and Regional Broad colleagues who provided valuable feedback on the DQM materials.

Disclaimer

The content of this paper does not necessarily reflect the views and policies of the State Water Resources Control Board, nor does mention of commercial products constitute endorsement or recommendation for use.

References

Please contact the author at rk@rb2.swrcb.ca.gov for the latest URLs and DQM material.

Glossary

Accuracy: The extent of agreement between an observed value (measurement result) and the accepted, or true, value of the parameter being measured.

Blank (Sample): A sample that contains pure water and is analyzed concomitantly with a set of environmental samples. Blanks usually include field blanks and trip blanks to assure that there was not contamination during sampling and shipping, as well as method blanks and reagent blanks tested within the analytical procedures

Calibration: The action of adjusting the readings of an instrument to have them match a “true” value as represented by known natural conditions (e.g., freezing point) or a laboratory standard.

Contamination: Inadvertent addition of an analyte or interfering compounds to a sample from the sampling equipment, sample container, etc. Contamination may cause false positive results or higher result values.

Concomitant (sampling) – accompanying; one activity related to a given spot is happening at the same time as another activity at the same spot, e.g., a sample for analysis of Ammonia is collected from the creek at the same time the pH is measured in creek water.

Data - bits of information. The core of DATA is the RESULT.

Descriptors (aka Metadata) – A myriad of descriptive items that provide information about the data. There are about 200 descriptors needed to make one result "bit" usable, reliable, and of known quality.

Endpoint: (1) A numerical value representing the result of a measured parameter that has been calculated from a number of individual measurements (e.g.: flow discharge in cubic feet per second, or bacterial concentration in MPN/100 ml.). (2) That stage in titration at which an effect, such as a color change, occurs, indicating that a desired point in the titration has been reached.

Measurement – (1) one of the sample/measurement/observation (SMO) entities, conducted in situ (with probe) or in a sampling device, immediately (not hauled away). (2) generic term for any quantitation activity

Measurement Quality Objectives: Statements about the tolerated error and desired sensitivity of a measurement. They include extent of values for the measures of precision, accuracy, detection limit, and resolution. MQOs are a subset of Data Quality Objectives (DQOs).

Metadata: "Data about data." Information that describes the result of each measurement (i.e., the “how much”) in terms of what, when, where, why, how and by whom that result was collected. This information is essential for data validation and helps others understand exactly how the data was obtained.

Observation - one of the sample/measurement/observation (SMO) entities, an on-site estimate or evaluation of a parameter not measured. Results are expressed categorically (word from menu, e.g., "murky", or numeric range category, e.g., "1 to 20 l/sec”)

Parameter: A property or substance to be measured within a medium. Parameters include properties such acidity (pH) or electrical conductivity, particulates such as suspended solids or bacteria, and analytes such as ammonia or heavy metals.

Precision: A measure of how close repeated trials are to each other

Quality Assurance/Quality Control (QA/QC): The total integrated program for assuring the validity of monitoring and measurement data.

Representativeness: A data quality indicator, representativeness is the degree to which data portray the actual or true environmental condition measured.

Resolution: The smallest increment that can be discerned on the scale of a measuring device, or the capability of a method to discriminate between measurement responses.

Result – the outcome of a measurement, analysis, or observation. Results are always linked to a parameter, and are linked to units in most situations. A result for a specific parameter is what the collector produces and what the data user will use.

Sample - one of the sample/measurement/observation (SMO) entities, a sample is collected in a container, hauled away, and analyzed elsewhere

Sample (SMO) ID – a unique identifier for a sample/measurement/observation (SMO) entity.

Water Quality Parameters: Any of the measurable properties, qualities or contents of water.

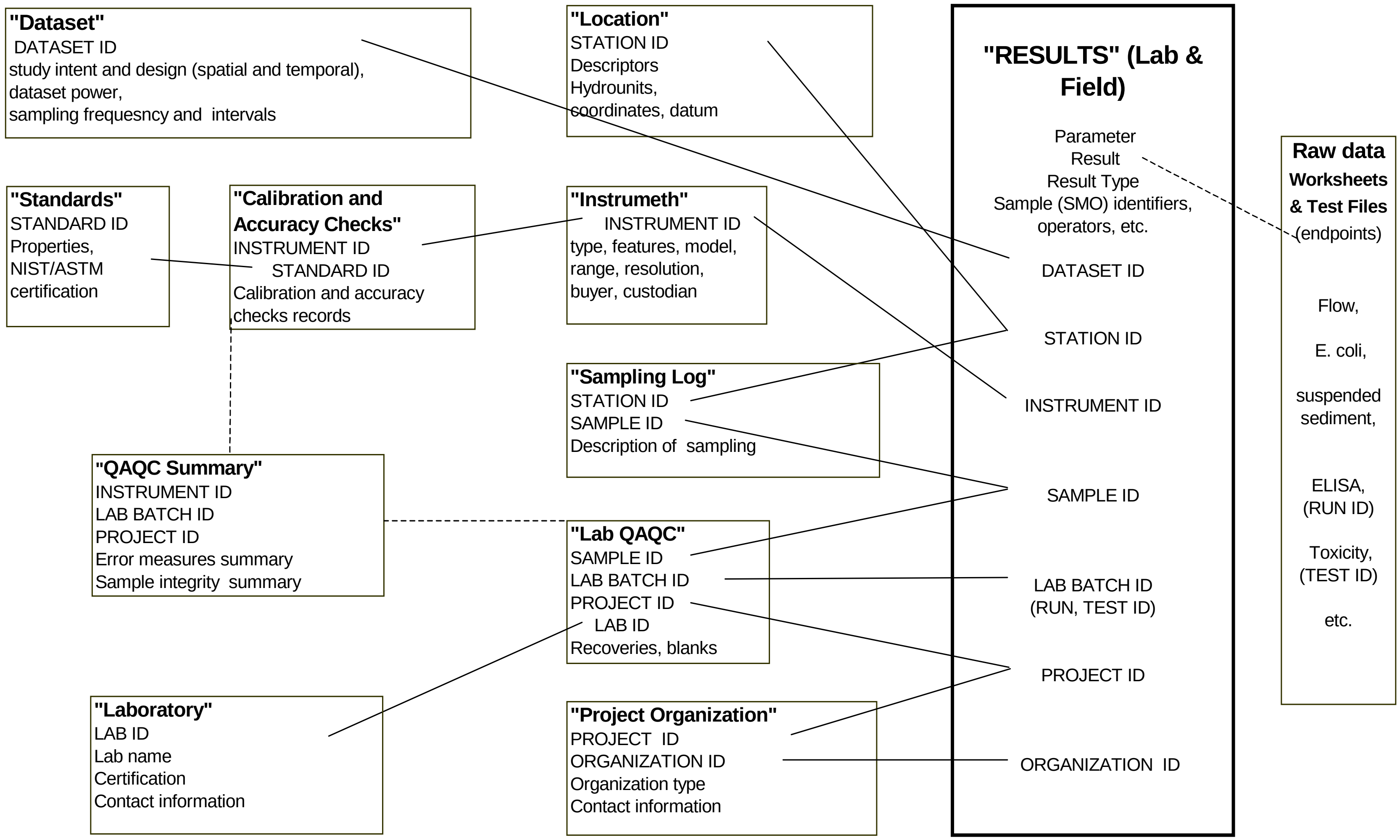

Figure 1: Linkages within the Data Quality Management (DQM) Project File

Table 1: Attributes of GOOD DATA

USABILITY of the data –

*	Capable of answering questions and supporting management decisions
*	Complete Parameter Package: all supporting parameters included
*	Comparable to other data sets in terms of sampling design and data quality
*	Adequate Measurement Quality Objectives in terms of sensitivity and tolerated error
*	Adequate Statistical Power of the dataset (number & replication of samples)
*	Scientifically defensible, including in court
*	Reported and documented in formats that can be easily read, understood, and transformed by others

RELIABILITY (Credibility) of the data –

1	High probability that the reported value indeed falls within the range of error specified for it, and
2	Complete documentation is provided (all the information on location, sampling design, measurement, QAQC, etc), and
3	Honest reporting by field operators

VALIDITY of the data –

1	Compliance with data quality objectives has been confirmed, and
2	The test, assay, or analysis used to collect the data was valid

KNOWN QUALITY of the result: Three DISTINCT aspects -:

A	Quality of the measurement itself (accuracy, precision, detection limit, resolution)
B	Sample integrity (lack of deterioration, lack of contamination)
C	Representativeness of the measured value (how does the sample represent “true” conditions, across time & space: issues of sampling design and sample collection method in relation to inherent variability)

FIDELITY of the data –

*	Consistent interpretation of "menu options" for verbal categories or numeric range categories
*	Correct transfer of information from observer to "scribe"
*	Correct recording, copying, and data entry into electronic formats

CERTIFICATION –

*	Laboratory analyses were made by a certified laboratory
*	Laboratory work was checked against certified Standards
*	Field work was done by certified operators, survey work done by certified surveyors
*	Field work done with or calibrated against certified instruments and Standards

Table 2: Data qualifiers

Field name	menu item Name	menu item definition
Documentation Level		
	Unknown	
	Minimal	Poor Station location description, unknown instrument or lab batch, etc.
	Partial	By best professional judgment (BPJ) – vital information pertaining to the measurement quality or to what the results represent (or both) is missing
	Adequate	By best professional judgment (BPJ) – there is sufficient information to allow for use of the result without restrictions
Validity Status		
	Unknown	
	Not Checked	Data quality has not been reviewed
	Not Valid ("R")	("R" for rejected) existing information indicates that the result was obtained in an analytical run or toxicity test that were not acceptable, or with the use of malfunctioning instrument
	Estimated ("J")	"J"; by best professional judgment (BPJ) - not valid but flaw not detrimental; result can be used but with caution
	Valid	Analytical run or toxicity test were acceptable; recoveries or reference toxicant test results were within appropriate control chart
Measurement Error Range category		
	Unknown	
	0 to 10%	Combined measure of measurement accuracy, precision, and resolution was within the range of 0 to 10% of measured value
	10 to 20%	"
	20 to 50%	"
	50 to 100%	"
	100 to 200%	"
	>200%	"
Fidelity of data recording or entry		
	Unknown	There is no measure of confidence in the fidelity of the entered result
	50% or less correctly entered	Spot-checks or double-entry indicate that less than 50% of the results in the dataset have been recorded or entered correctly
	50-90% correctly entered	"
	more than 90% entered correctly	"

Table 3: Use of dataset intent, design, and power descriptors for monitoring scenarios examples

Field name	Scenario # (see description of the scenarios below)			
	1	**2**	**3**	**4**
Spatial descriptors				
Station Type	creek	creek	outfall	creek
Station Selection Intent	Impact assessment	source ID	not applicable	characterization of refuge areas
Reach Selection Design	directed	directed	not applicable	directed
Station Selection Design	directed	directed	Anecdotal (Wherever)	directed
Temporal descriptors				
Storm runoff flows (wet) or base flow (dry) weather	not applicable	dry	dry	dry
Sample Timing Intent	not applicable	not applicable	worst case	worst case
Seasonal Sampling Design	directed	not applicable	Directed or Anecdotal (Whenever)	directed
Season of interest	any	any	any	summer
Diurnal Sampling Design	not applicable	Systematic	Directed or Anecdotal (Whenever)	directed
Dataset power descriptors				
Total # of Station-visits	8/dataset (2 Stations 2 years)	240/dataset (10 Stations 1 year)	120/dataset (5 Stations 1 year)	40 (2 Stations 1 year)
Sampling Frequency (at each Station)	2/year	24/year	24/year	20/year (2 per day at 0500 and 1400, 10 days/year)
Sampling Interval	6 months	2 weeks	2 weeks	1 week

Scenarios:

1. Bioassessment to determine if a Treatment Plant discharge into creek has impact on macroinvertebrates
2. Routine turbidity measurements above confluences to identify tributaries contributing suspended sediments during non-storm conditions
3. Detergent measurements during dry weather at discharging outfalls
4. Monitoring of temperature and dissolved oxygen (at dawn and in the afternoon) to determine if salmonid fish can survive in a given urban creek

Table 4: Selected examples of dataset descriptors and their definitions in a "Lookup Table Dictionary"

Field name	menu item name *(Note a)*	menu item definition
Station Type		
	overland flow (gutter, swale)	Sampling activity conducted at a land-based geographic feature that was temporarily covered with runoff water.
	outfall	End of a pipe that transports effluent from a Facility to a discharge point on or in a body of water
	creek	
	lake	
	ocean	
Station Selection Intent (reason for selecting that location to sample)		
	Source ID	Identifying the source of a given constituent within a river network or land use activities
	characterization of refuge areas	Identifying and characterizing habitat areas having the best-case-scenario in term of extreme conditions; i.e., the least impacted habitats in a reach
	impact assessment	Monitoring to determine whether an impact to a given ecosystem has occurred.
	permit compliance monitoring	Monitoring for the purpose of comparison with water quality benchmark specified in a discharge permit to check if that discharge is in compliance
Sample Timing Intent (reason for monitoring at the selected time)		
	routine monitoring	Repeated monitoring at fixed time intervals to provide long-term data
	snapshot	One-time monitoring of multiple Stations
	dry weather discharge	Monitoring during dry weather to characterize non-storm flow
	storm runoff monitoring	Monitoring storm runoff events at different water levels and phases during the event
	worst case scenario	Monitoring during the times anticipated to represent the most critical or the most extreme conditions within the natural fluctuations.
Stream Reach Selection Design (design principle used to select stream reach) *(Note b)*		
	systematic	Deterministic approach, points selected deliberately at fixed-intervals of area, length, or time
	directed (targeted)	Deterministic approach, points selected deliberately based on knowledge of their attributes of interest
	stratified random	Probabilistic approach, deliberate, points selected at random from a population stratified by specific attributes
	Anecdotal	Non-of-the-above, non-deliberate; points selected causally or whenever/wherever, or by given constraints

Note a: The table presents a small subset of the actual menus (except for the design fields).
Note b: The same menu items also apply to Station Selection Design, Seasonal Sampling Design, and Diurnal Sampling design

Paper source

Katznelson, R. (2002). Letting Monitoring Data Speak for Themselves. in: Proceedings of the third National Monitoring Conference of the National Water Quality Monitoring Council, "Building a Framework for the Future", May 19-23, 2002, Madison, WI.

For further information feel free to contact the author

Revital Katznelson, Ph.D.
Regional Citizen Monitoring Coordinator (retired 2006)
State Water Resources Control Board

Current (2012)
Revital Katznelson, Ph.D.
Environmental Scientist
Berkeley, California
revitalk@sbcglobal.net
510 406 8514