A multi-disciplinary list of data elements for environmental monitoring: can we all speak the same language?

Revital Katznelson, Ph.D.
Berkeley, CA

Today's focus

Background: Development of data elements lists for separate monitoring disciplines

Why should we integrate data elements lists?

How should we integrate: identification of challenges, common elements, and unifying concepts

Organization and contents of the integrated list of data elements

Our audiences

Why am I **talking** about it?

Have you ever watched a **chemist** and a **fluvial geomorphologist** explain their work to each other in the presence of a **benthic macroinvertebrates specialist** (and being eavesdropped by a **toxicologist**?).

Yes? Send in the **microbiologist**, summon the **hydrologist,** gather the rest of the "ists"…

And let's have a dialog!

Why am **I** talking about it?

Been involved in environmental monitoring for >3 decades

Have interacted with folks from a variety of Agencies, Programs, disciplines, and Areas of Inquiry

Developed guidance, training tools, and data management tools

Taught numerous workshops and courses

Spent a lot of time compiling data elements lists

And I want to keep the **Information Management Dialog** moving forward.

Background Story, Part 1

I started putting Results of **field measurements** into an Excel spreadsheet in the mid-nineties, included tracking entities (Station ID, Sample ID, Instrument ID) and other essential bits of information, e.g., base flow of rain runoff

Created more spreadsheets for bits of information on location, instruments, calibration and accuracy checks records, Standards, etc.; everything the QA/QC officer wants to see

Added spreadsheets for laboratory results, toxicity testing, data loggers, etc.

Added bits of information that will allow data users to group, sort, and filter the data based on what each Result represents in the environment

Ended up having an Excel workbook with fully-documented data, and called it “the Project File”

(Available at http://www.waterboards.ca.gov/water_issues/programs/swamp/cwt_toolbox.shtml)

The Project File has placeholders for all bits of information

What

How Good?

(worksheets)

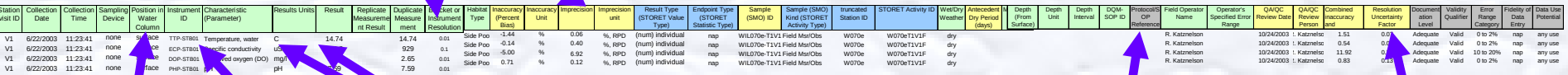
Who

Where

When

How

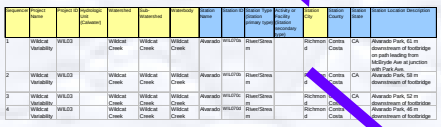
Why



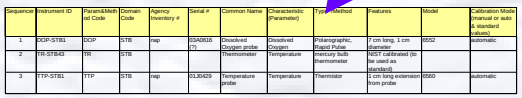
Station Visit ID	Collection Date	Collection Time	Sampling Device	Position in Water Column	Instrument ID	Characteristic (Parameter)	Results Units	Result	Replicate Measurement Result	Duplicate Measurement	Bucket or Instrument Resolution
V1	6/22/2003	11:23:41	none	surface	TTP-STB01	Temperature, water	C	14.74	14.74	14.74	0.01
V1	6/22/2003	11:23:41	none	surface	ECP-STB01	Specific conductivity	uS/		929	929	0.1
V1	6/22/2003	11:23:41	none	surface	DOP-STB01	Dissolved oxygen (DO)	mg/l		2.65	2.65	0.01
V1	6/22/2003	11:23:41	none	surface	PHP-STB01	pH	pH	7.59		7.59	0.01

Habitat Type	Inaccuracy (Percent Bias)	Inaccuracy Unit	Imprecision	Imprecision unit	Result Type (STORET Value Type)	Endpoint Type (STORET Statistic Type)	Sample (SMO) ID	Sample (SMO) Kind (STORET Activity Type)	Truncated Station ID	STORET Activity ID	Wet/Dry Weather	Antecedent Dry Period (days)	Depth (From Surface)	Depth Unit	Depth Interval	DQM-SOP ID	Protocol/SOP Reference	Field Operator Name	Operator's Specified Error Range	QA/QC Review Date	QA/QC Review Person	Combined Inaccuracy and	Resolution Uncertainty Factor	Documentation Level	Validity Qualifier	Error Range Category	Fidelity of Data Entry	Data Use Potential
Side Poo	-1.44	%	0.06	%, RPD	(num) individual	nap	WIL070e-T1V1	Field Msr/Obs	W070e	W070eT1V1F	dry							R. Katznelson		10/24/2003	I. Katznelso	1.51	0.01	Adequate	Valid	0 to 2%	nap	any use
Side Poo	-0.14	%	0.40	%, RPD	(num) individual	nap	WIL070e-T1V1	Field Msr/Obs	W070e	W070eT1V1F	dry							R. Katznelson		10/24/2003	I. Katznelso	0.54	0.01	Adequate	Valid	0 to 2%	nap	any use
Side Poo	-5.00	%	6.92	%, RPD	(num) individual	nap	WIL070e-T1V1	Field Msr/Obs	W070e	W070eT1V1F	dry							R. Katznelson		10/24/2003	I. Katznelso	11.92	0.01	Adequate	Valid	10 to 20%	nap	any use
Side Poo	0.71	%	0.12	%, RPD	(num) individual	nap	WIL070e-T1V1	Field Msr/Obs	W070e	W070eT1V1F	dry							R. Katznelson		10/24/2003	I. Katznelso	0.83	0.13	Adequate	Valid	0 to 2%	nap	any use

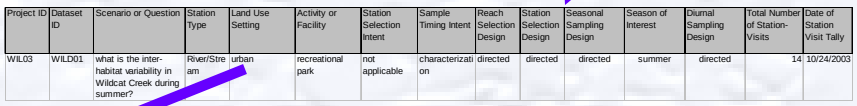
STORET Organization ID	Organizational Entity Name	Team Name	Organizational Entity Category	Organizational Entity Type	Organizational Entity Contact person ID	Contact Last name	Contact First name	Contact Role	Contact Status (volunteer or staff)	Address Line 1	Address Line 2	City
AWCM	Berkeley Ecology Center	RK Crew	Watershed Organization	Not-Profit Resource Center		Leff	Penny	Trainer	Volunteer			
AWCM	Wildcat Creek Monitors	RK Crew	Watershed Organization	Volunteer Group		Katznelson	Revital	Technical Leader	Volunteer			Berkeley



Sequence	Project Name	Project ID	Hydrologic Unit (Delaware)	Watershed	Sub-Watershed	Waterbody	Station Name	Station ID	Station Type (Station primary type)	Facility or Facility Division (Station secondary type)	Station City	Station County	Station State	Station Location Description
1	Wildcat Variability	WIL03		Wildcat Creek	Wildcat Creek	Wildcat Creek	Alvarado	WIL070e	River/Stream		Richmond	Contra Costa	CA	Alvarado Park, 61 m downstream of footbridge on path leading from McBryde Ave at junction with Park Ave.
2	Wildcat Variability	WIL03		Wildcat Creek	Wildcat Creek	Wildcat Creek	Alvarado	WIL070b	River/Stream		Richmond	Contra Costa	CA	Alvarado Park, 58 m downstream of footbridge
3	Wildcat Variability	WIL03		Wildcat Creek	Wildcat Creek	Wildcat Creek	Alvarado	WIL070d	River/Stream		Richmond	Contra Costa	CA	Alvarado Park, 52 m downstream of footbridge
4	Wildcat Variability	WIL03		Wildcat Creek	Wildcat Creek	Wildcat Creek	Alvarado	WIL070d	River/Stream		Richmond	Contra Costa		Alvarado Park, 46 m downstream of footbridge



Sequence	Instrument ID	Parameter Method Code	Domain Code	Agency Inventory #	Serial #	Common Name	Characteristic (Parameter)	Type	Features	Model	Calibration Mode (manual or auto & standard values)
1	DOP-STB1	DOP	STB	nap	03A0616 (?)	Dissolved Oxygen probe	Dissolved Oxygen	Polarographic, Rapid Pulse	7 cm long, 1 cm diameter	6562	automatic
2	TR-STB43	TR	STB			Thermometer	Temperature	mercury bulb thermometer	NIST calibrated (to be used as standard)		
3	TTP-STB1	TTP	STB	nap	D1JD429	Temperature probe	Temperature	Thermistor	1 cm long extension from probe	6560	automatic



Project ID	Dataset ID	Scenario or Question	Station Type	Land Use Setting	Activity or Facility	Station Selection Intent	Sample Timing Intent	Reach Selection Design	Station Selection Design	Seasonal Sampling Design	Season of Interest	Diurnal Sampling Design	Total Number of Station-Visits	Date of Station Visit Tally
WIL03	WILD01	what is the inter-habitat variability in Wildcat Creek during summer?	River/Stream	urban	recreational park	not applicable	characterization	directed	directed	directed	summer	directed	14	10/24/2003

What does it represent?

Background Story, Part 2

Organized the Data Fields as “Data Elements” in a list, by subject matter

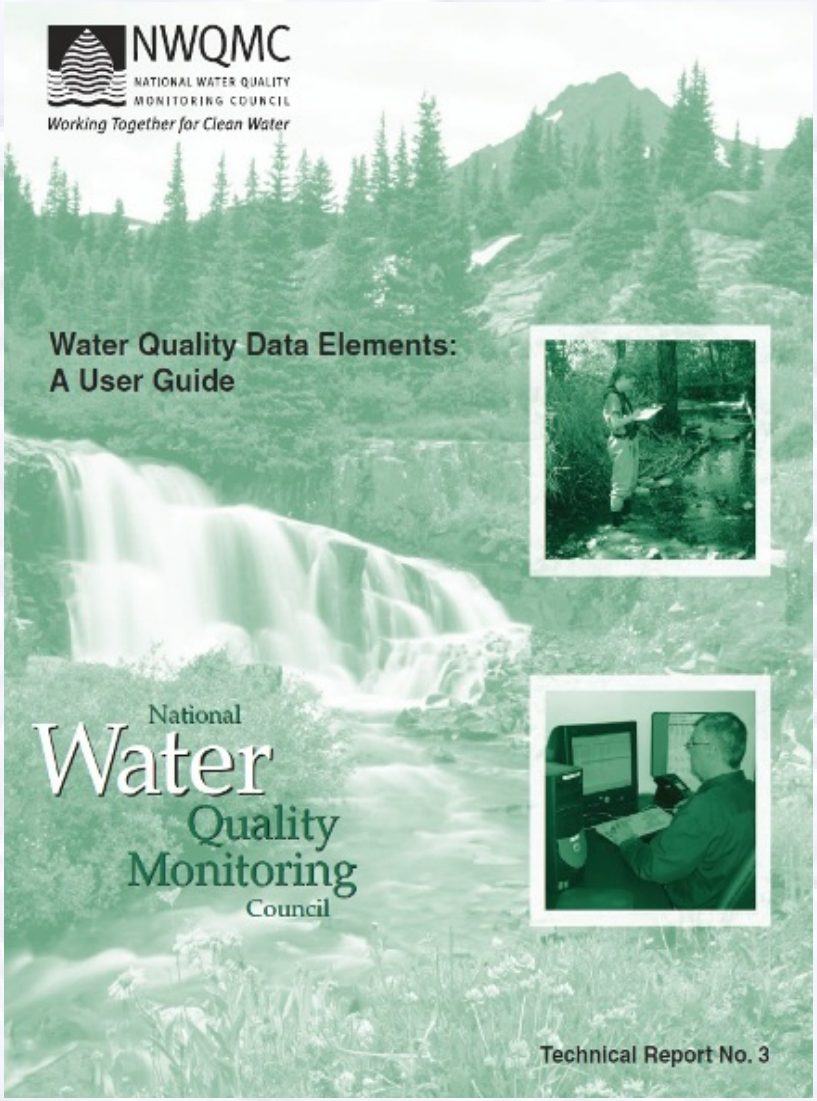
Started an array of Pick-Lists, a.k.a. Lookup Tables, for selected Data Fields

Discovered that the Water Quality Data Elements workgroup (WQDE) of MDCB has already created a core data elements list for chemistry and microbiology

Joined the WQDE workgroup!

History

The Water Quality Data Elements (WQDE) Workgroup of the Methods and Data Comparability Board created a basic list of core data elements and augmented it by supplements.



2001 WQDE for analytical chemistry & microbiology
2004 WQDE for toxicity tests
2005 WQDE for biological populations & communities
[Guidance document published 2006]
2008 WQDE for physical habitat

Problem: the four lists have not been integrated.

In fact, everyone has their own point of view!

Kingfisher: **Bummer!**

Minnow: **Shelter!**

Lateral pool with LWD & root wad

Angler: **Snag!**

Water Strider: **Shade!**

Geomorphologist: **Scour!**

Restoration Designer: **Opportunity!**

Why integrate?

Integration of data elements can facilitate integration of monitoring and assessment data from various areas of inquiry, including:

- chemical analyses,
- discrete field measurements,
- continuous (time series) field measurements
- bacterial counts,
- toxicity testing,
- biological assessments such as fish, benthic macroinvertebrate, or periphyton assemblages, and
- physical habitat assessments.

How can we integrate?

All of us environmental monitoring folks devise Measurement Systems, and run Quality Checks, and apply Sampling Design Principles, and visit monitoring Stations, and obtain Monitoring Results - in all our disciplines and areas of inquiry.

One way to integrate the lists is based on these concepts, and many others, which unify various areas of inquiry.

In fact, a high percent of data elements ARE common to all our disciplines and areas of inquiry.

Where do folks think they are unique?

The way they conduct their Field Activities – evaluation of vegetation density seems very different from measuring temperature

The way they report results and calculate Result Endpoints [descriptive statistics, metrics, indices, percentiles, etc.]

The way they use their measurement systems

The way they run their Quality Checks and the way they report the outcomes

‘Activity’ can be a thing!

(-) Pete Ruhl

Activity – a.k.a. Sample/Measurement/Observation – is what we do in the field to generate data or start a data generation process. It can be an action or an object.

- **Evaluative,** using eyes/brain/experience (e.g., categorical observation, estimate, score)
- **Measured,** using an instrument *in situ* or counting (e.g., discrete/time series field measurement, elevation survey, count, etc.)
- **Collected,** by transferring medium into containers for processing elsewhere, or capturing biota (Sample ***sensu strictu***)

Result Types

Result - the outcome of an evaluation, measurement, or analysis.

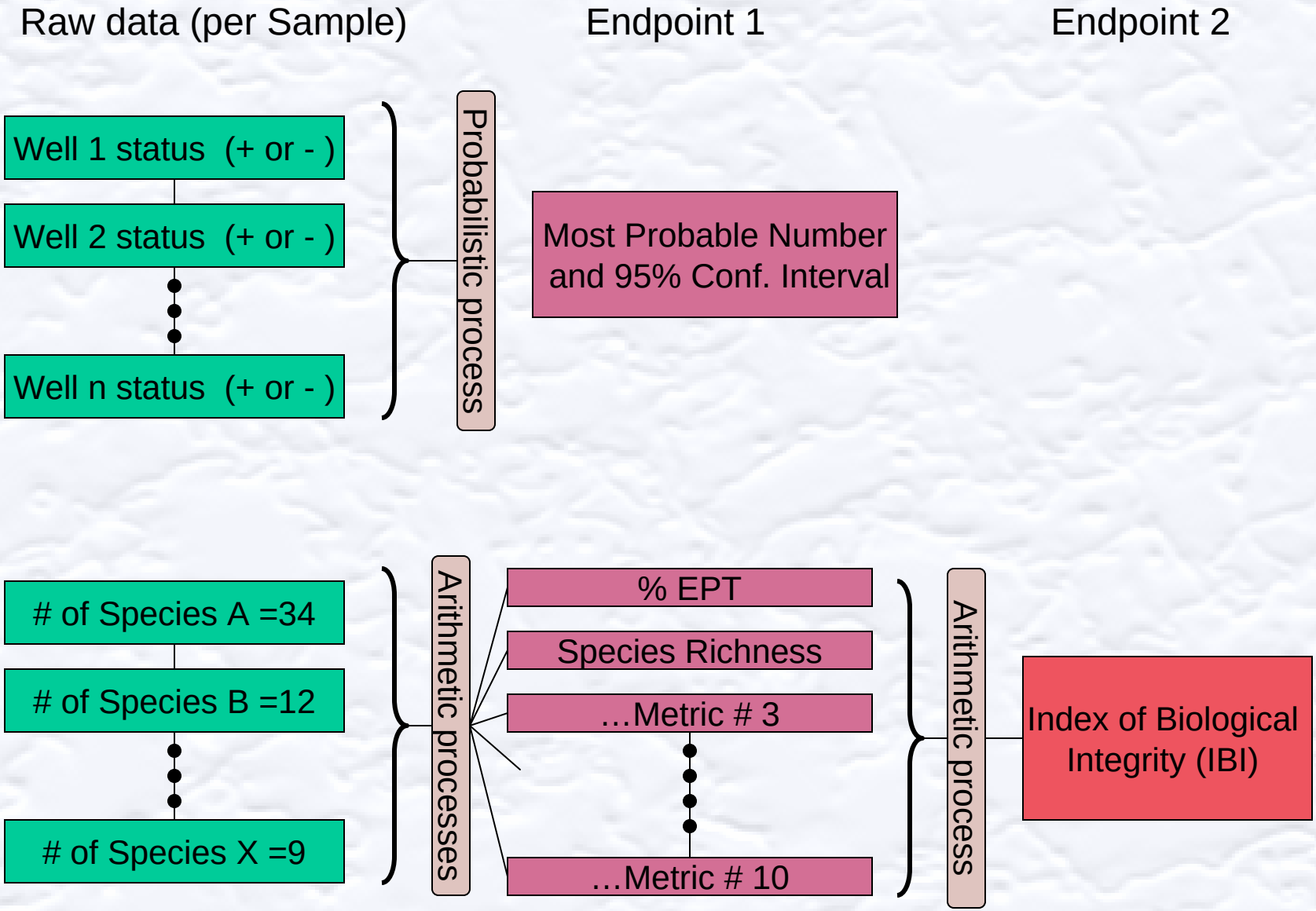
Verbal Category (e.g., 'murky')
Individual Value (e.g., 18 C)
Estimated number or numeric range (e.g., 20-50% embeddedness)
Count
Score
Calculated "Endpoint"

Result Endpoint examples

Endpoint – a.k.a. Calculated Endpoint - a catch-all term for metrics, indices, descriptive statistics (min, max, mean, and various percentiles), etc.

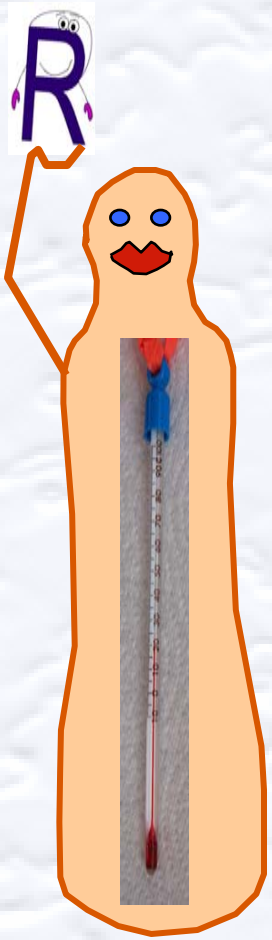
- **Derived Endpoint** for **one** Sample (e.g., LC50=0.4ug/L; E. coli=126 MPN/100mL; % tolerant taxa=24) [1 char, dependent]
- **Aggregated Endpoint** for **one** Sample or **spatial entity** (e.g., PEC toxicity quotient, Average stream width, median particle size, Index of Biological Integrity, etc.) [possibly >1 char, independent]
- **Descriptive Statistic** for **many** data points (Endpoint types: moving weekly average, 5-wk geometric mean, max, MWAT, etc)
- **Rank** for many data points

Results and Endpoint possibilities



“Measurement Systems” used to be only….

Devices and/or procedures used for quantitation of environmental characteristics, including instruments used for field measurements and sampling & analysis processes.



Some have one simple step, others have many steps

(= Result)

But they also include physical habitat assessments

(Also think about:

Densiometer;
Stadia rod;
Level;
Field data sheet for observations;
and more)

And an array of Sensors' Measurement Systems

Platform Installations:
- Water intake/circulation devices
- Effects of flood, debris, etc.

Output generation:
- The sensor
- The probe
- The display unit
- The data-logging and/or telemetry unit
- The connections/algorithms/apparati

that enable (manual or automatic) corrections for Temperature, Pressure, and Salinity (TaPaS)

Post-Output:
- Random measurement error
- Fouling error
- Drift error (systematic error)

Every Measurement System needs Quality Checks

- Matrix Spike/Matrix Spike Duplicate,
- sample arrival temperature,
- repeated categorical observations, or
- a survey loop closure

…are very different from each other, but all of them are Quality Checks.

Quality Checks examples

Quality Checks - actions designed to validate monitoring data, and to assess the error associated with the data

- **Comparisons** to Standard solutions, instruments, natural points (accuracy checks) (e.g., Buffer, NIST Thermometer, freezing point)
- **Repeats** of measurements or observation (precision checks)
- **Spikes** of analytes or surrogates (e.g., Certified Reference Material, internal standards, Matrix spikes, etc.) (recovery checks; accuracy)
- **Blanks** for containers and equipment (lack of contamination)
- **Positive/Negative Controls** for organisms or effects (e.g., microbial cultures, reference toxicants, etc.) (validity check)
- **Survey Loop** for elevation measurements (e.g., Survey Loop closure) (accuracy check)
- **Inspections/Verifications** for sample batches (e.g., custody seal, arrival temperature) (sample integrity, lack of deterioration)

The Power of Categorization and Typing

- Organizations, Results, Activities, Quality Checks, and many other entities that are common to all manner of environmental monitoring are unifying concepts.
- Categorizing and typing the variants contained in each of these entities clearly show how variants from different areas of inquiries belong together.
- If variants belong together, they can be placed on the same page in terms of information contents.
- Category/Types matrices can also provide the Pick-Lists for descriptors of these entities.

Integration Challenges

Cultural:

Data gatherers from different areas of inquiry ‘think’ differently

Intended users:

“Short list” (of core elements only, for data ‘miners’) versus “Long list” for use at the Project level

Structural:

Are all elements equivalent to Data Fields (i.e., column headers in a database table), or not?

Are the Pick-Lists included within the WQDE list, or are they organized as a separate list?

Decisions

Scope:

Water and sediment monitoring

Intended users:

Create the integrated list as a “Long list”, with the majority of data elements needed at the Project level, and later distill a “Short list” of **core** data elements for data ‘miners’

Structure:

Maintain the data elements as Data Fields (i.e., column headers in a database table),

Maintain Pick-Lists of cell-contents under a separate cover in an accompanying document

Organize the data element groups to fit the 2001 WQDE Modules

Construct the list based on unifying concepts.

Example: another unifying concept

Horizontal ….

All these items in purple cells are **characteristics**

Station ID	date	time	Specific Conductance	Diss. oxygen	H2O temp	pH	Turbidity
			µS	mg/l	Celsius		JTU
SLC-WB	3/29/1997	10:45	570	6.4	14.5	7.8	ND
SLC-WB	**4/26/1997**	10:30		4.6	17.5	6.8	ND
SLC-WB	6/7/1997	10:55	530	10.2	19		5
SLC-WB	6/21/1997	11:20		15	18	8.2	5
SLC-WB	8/2/1997	10:50	660	1	18	7.4	ND

↓ …. to Vertical

Station ID	date	time	Instrument ID	Reso - lution	Characteristic	units	Result	Qual	Comment
SLC-WB	3/29/1997	10:45	EC-SLC3	10	Sp. Conductance	µS	570		
SLC-WB	3/29/1997	10:45	DOW-SLC1	0.2	Dissolved oxygen	mg/l	6.4		
SLC-WB	3/29/1997	10:45	TR-SLC1	1	H2O Temperature	°C	14.5		
SLC-WB	3/29/1997	10:45	PHEL-SLC3	0.1	pH	pH	7.8	J	
SLC-WB	3/29/1997	10:45	TUJ-SLC1	5	Turbidity	JTU		ND	
SLC-WB	**4/26/1997**	10:30	EC-SLC3	10	Sp. Conductance	µS			dead battery

The “characteristic” field is a unifying entity

Bonus: you can add any number of descriptors to each Result

The new list is hierarchical

Data Fields transposed and organized by subject matter

Category number	Category name	Group number	Group name	Subject number	Subject name	Field ID	Field name	Pick-List examples
2	Result	1	Result (how much?)	2.1.1	Value	2.1.1.1	Result Value	7.8 (pH), murky, 8 (liter/sec)
2		1		2.1.2	Unit	2.1.2.1	Result Unit	mg/L, uS,
2		1				2.1.2.2	Measurement Basis	wet weight; ash-free dry weight
2		2	Characteristic (of what?)	2.2.1	Name	2.2.1.1	Characteristic	Temp, DO, Sp.Cond., pH, Turb
2		2		2.2.2	Identifier - Chemical	2.2.2.1	Substance Unique ID	
						2.2.2.2	Substance Registry System Name	**CAS, WQX code,**
2		3	Result descriptors	2.3.1	Result Type	2.3.1.1	Result Type	
2		3				2.3.1.2	Endpoint Type	
2		3		2.3.2	Result Qualifiers	2.3.2.1	Result Qualifier Code	
2		3				2.3.2.2	Batch Qualifier Code	J, R
2		3		2.3.3	Result-Specific procedures /outcomes	2.3.3.1	MDL	
2		3				2.3.3.2	Reporting Limit or PQL	
2		4						

Complete Data Element list: familiar categories

Category number	Category (Module) name	Group number	Group name
1	Monitoring Project ("Contact")	1.1	Project Identifiers
		1.2	Organization identifiers
		1.3	Project Contacts
2	Result	2.1	Result (how much?)
		2.2	Characterisitc (of what?)
		2.3	Result descriptors
3	Study Dataset ("Reason")	3.1	Identifiers
		3.2	Intent
		3.3	Design (…more)
4	Site Visit ("Date/time")	4.1	Trip
		4.2	Visit
5	Location		Site identifiers
			Site Description
			Site Location
6	Activity-Field ("sample collection")		Identifiers, type
			Spatial descriptors
			Sample (in a jar)
7	Measurement System - field&lab ("Sample Analysis")		Instruments and lab batches
			Method
			Quality checks

(plus "Housekeeping" data fields

The new list can provide for…

(a) Integration of data elements from different areas of inquiry related to aquatic systems

(b) Effective ‘crosswalks’ between different terminologies

(c) Organization by subject matter, independent of database structure and data models

(d) Standardization of pick-lists (a.k.a. Lookup tables)

(e) Comprehensiveness with flexibility: nitty-gritty for Project personnel, **core** elements for data users

Current status

A comprehensive list of Data Elements for aquatic sensors (covering field measurements and time-series monitoring) has been drafted, based on the new integrated list. The Sensors list is expanded updated by the Aquatic Sensors Workgroup in collaboration with WQX and WaterML2 of the OGC.

A PDF version of the Sensors list can be found on the Aquatic Sensors Workgroup (ASW) site

http://www.watersensors.org/

For the most current versions, please contact Revital at

revitalk@sbcglobal.net