OULUN YLIOPISTO
UNIVERSITY of OULU

DATA QUALITY FLAG SCHEMES
IN DATA QUALITY ASSURANCE (QA): REVIEW AND SYNTHESIS

Mika Ruusunen
FLEX^e, D2.3-12-3
30.6.2016
Contact: mika.ruusunen@oulu.fi

REVIEWED STANDARDS AND STATE-OF-ART PRACTICES

- **MID**, DIRECTIVE 2004/22/EC, on measuring instruments
- **ISA100.11a**, standard on wireless networking technology
- **IEC 61850**, standard on substation automation
- **NOAA**, U.S., handbook of automated data quality control checks and procedures
- **Ocean Data Standards**, multiple quality flag schemes
- **EuroGOOS**, Recommendations for in-situ data real time quality control
- **Nordklim**, quality control of meteorological observations
- **Kuffel et. al**, real time simulation and testing using IEC 61850
- **Campbell et.al**, 2015, Bioscience, Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data
- **Taylor et.al**, 2013, Biogeosciences, Automated quality control methods for sensor data
- **Rönkkö et.al**, 2015, Quality control of environmental measurement data with quality flagging
- **Beckman Coulter Inc.**, accuracy and precision flags

OULUN YLIOPISTO
UNIVERSITY of OULU

MID

- MI-001, water meters
- MI-002, gas meters
- MI-003, active electrical energy meters
- MI-004, heat meters (water)
- MI-005, measuring systems for quantities of other liquids than water
- …
  - Qualifiers:
    - Range limits (flow, temperature, pressure, AC/DC supply)
    - Maximum permissible error 2-5% (MPE) for flowrate depending on temperature level
    - Permissible effect of disturbances:
      "The critical change value is the smaller of the two following values:
      — the quantity corresponding to half of the magnitude of the MPE in the upper zone on the measured volume;
      — the quantity corresponding to the MPE on the quantity corresponding to one minute at maximum flowrate."
    - Durability tests: flow (<critical value) and temperature (<0,1 °C)

OULUN YLIOPISTO
UNIVERSITY of OULU

ISA100.11A

- Focused on wireless measurements

- Related qualifiers:
  - Jitter (latency)
  - Punctuality (timeliness)
  - Reliability (missing values/h)

- Sensor classes: 0 to 5, according to timeliness requirements

OULUN YLIOPISTO
UNIVERSITY of OULU

IEC 61850-7-3

• Qualifiers:

Bit(s)	IEC 61850-7-3		Bit-String	
	Attribute name	Attribute value	Value	Default
0-1	Validity	Good	0 0	0 0
		Invalid	0 1	
		Reserved	1 0	
		Questionable	1 1	
2	Overflow		TRUE	FALSE
3	OutofRange		TRUE	FALSE
4	BadReference		TRUE	FALSE
5	Oscillatory		TRUE	FALSE
6	Failure		TRUE	FALSE
7	OldData		TRUE	FALSE
8	Inconsistent		TRUE	FALSE
9	Inaccurate		TRUE	FALSE
10	Source	Process	0	0
		Substituted	1	
11	Test		TRUE	FALSE
12	OperatorBlocked		TRUE	FALSE

Redrawn from Kuffel, Rick, Dean Ouellette, and Paul Forsyth. 2010. "Real Time Simulation and Testing Using IEC 61850." In *Modern Electric Power Systems (MEPS), 2010 Proceedings of the International Symposium*, 1–8. IEEE.

OULUN YLIOPISTO
UNIVERSITY of OULU

NOAA

- Two level flagging scheme:
- Primary, general description
- Secondary, tests
- Based on evaluation and synthesis of 16 widely used schemes in ocean data standards

Value	Primary-level flag short name	Definition
1	Good	Passed documented required QC tests
2	Not evaluated, not available or unknown	Used for data when no QC test performed or the information on quality is not available
3	Questionable/suspect	Failed non-critical documented metric or subjective test(s)
4	Bad	Failed critical documented QC test(s) or as assigned by the data provider
9	Missing data	Used as place holder when data are missing

Example quality control tests / data processing history
Globally impossible value
Monthly climatology standard deviation test
Excessive spike check
Excessive offset/bias when compared to a reference data set
Excessive data uncertainty
Unexpected X/Y ratio (e.g., chemical stoichiometry or property-property X to T, S, density, among others)
Excessive spatial gradient or pattern check ("bullseyes")
Below detection limit of method
Interpolated value (not measured)
Data offset corrected value relative to a reference data
Expert review

Redrawn from NDBC Technical Document 09-02, Handbook of Automated Data Quality Control Checks and Procedures, 2009.

OULUN YLIOPISTO
UNIVERSITY of OULU

OCEAN DATA STANDARDS

- Example of reviewed ocean data standards

Flag Description	ODV	GTSPP	ARGO	SEADATANET	ESEAS	WOD	WODSTATION	WOCEBOTTLE	WOCECTD	WOCESAMPLE	QARTOD	BODC	PANGAEA	SMHI	OceanSITES	IODE
no quality control (QC) was performed	1	**0**	0	0	0	0	0	2	2	2	0	Q	*	blank	0	2
QC was performed; good data	0	**1**	1	1	1	0	0	2	2	2	3	blank	blank	blank	1	1
QC was performed; probably good data	0	**2**	2	2	1	0	0	2	2	2	3	blank	blank	blank	2	1
QC was performed; probably bad data	4	**3**	3	3	3	4	3	3	3	7	2	K	?	?	3	3
QC was performed; bad data	8	**4**	4	4	4	4	3	4	4	7	1	K	/	B	4	4
the value was changed as a result of QC	1	**5**	5	5	2	0	0	2	2	2	3	R	*			5	2
the value is missing	1	**9**	9	9	9	0	0	5	5	5	9	N	*	B	9	9

Oceanographic quality flag schemes and mappings between them
Version: 1.4, Authors: Reiner Schlitzer & Alfred Wegener

OULUN YLIOPISTO
UNIVERSITY of OULU

EUROGOOS

Code	Definition
0	No QC was performed
1	Good data
2	Probably good data
3	Bad data that are potentially correctable
4	Bad data
5	Value changed
6	Below detection limit
7	In excess of quoted value
8	Interpolated value
9	Missing value
A	Incomplete information

Recommendations for in-situ data Real Time Quality Control
Authors: Sylvie Pouliquen and the DATA-MEQ working group

- Proposed by EuroGOOS for real-time, in-situ flagging scheme

OULUN YLIOPISTO
UNIVERSITY of OULU

NORDKLIM | RÖNKKÖ ET. AL

Flag	Original interpretation	Generic interpretation
0	No check performed	Value not checked
1	Observation is ok	Approved value
2	Suspected small difference	Suspicious value
3	Suspected big difference	Anomalous value
4	Calculated value	Corrected value
5	Interpolated value	Imputed value
6		Erroneous value
7		Frozen value
8	Missing value	Missing value
9	Deleted value	Deleted value

Redrawn from Rönkkö, Mauno, Okko Kauhanen, Markus Stocker, Harri Hytönen, Ville Kotovirta, Esko Juuso, and Mikko Kolehmainen. 2015. "Quality Control of Environmental Measurement Data with Quality Flagging." In International Symposium on Environmental Software Systems, 343–350. Springer. http://link.springer.com/chapter/10.1007/978-3-319-15994-2_34.

OULUN YLIOPISTO
UNIVERSITY of OULU

QUALIFIERS / DATA QUALITY FLAGS
SUMMARY

Known qualifiers/check tests for flagging:

Accuracy
Uncertainty
Date and time:
       jitter
       timeliness
       chronology
Calibration:
       due
       time limits
Range limits
Detection limit
Persistence (frozen, stuck, normal, …)
Change in slope
Change in variance
Change in operating related statistics
Internal consistency/related sensors
Spatial concistency /nearby sensors
Sensor energy:
       battery life
       AC/DC supply quality
Data outlier
Missing value:
       single data point
       multiple data points
       deleted value
Redundancy:
       analytical
       non-parametric
Data source (original, corrected, replaced,
       interpolated, estimated,
       calculated, …)
Expert knowledge

General measures of quality:

Accuracy
Completeness
Consistency
Timeliness
Believability
Value added
Interpretability
Accessibility

OULUN YLIOPISTO
UNIVERSITY of OULU

PROPOSED FLAGGING SCHEME
QUALITY VALUES

Severity

Value	Data flag name	Definition
1	Good	Passed all checks
2	Probably good	Failed any non-critical check(s)
3	N/A QC	Quality not checked/available
4	Suspect	Failed, but maybe correctable
5	Bad	Failed QC or flagged by data provider
6	Missing data	N/A data point(s)

Increase

Properties:
- Fixed number of values
- Any of the flags can be justified by secondary level checks
- Universal, for variety of data sources

OULUN YLIOPISTO
UNIVERSITY of OULU

PROPOSED FLAGGING SCHEME
QUALITY LABEL

- Priority
- Time criticality
- Automatic procedure

← Decrease →

- Primary level:
- Quality values 1-6

1 2 3 3 2 1

- Secondary level flags:
- 6 ? Not → 1
- Statistical etc. checks
- Accuracy
- Uncertainty
- Timeliness & calibration
- Expert flag

OULUN YLIOPISTO
UNIVERSITY of OULU