1 24/02/2022 DAYCLI Message : Data Quality

1.1 Code Table for Data Quality Control information

0	Data checked and declared good
1	Data checked and declared suspect
2	Data checked and declared aggregated
3	Data checked and declared out of instrument range
4	Data checked and declared aggregated and out of instrument range
5	Parameter is not measured at the station
6	Daily value not provided
7	Data unchecked
8	
9	-
10	-
11	-
12	-
13	-
14	
…	…
255	Missing (QC info not available)

1.2 Design of the Data Quality Control information

The intention is to have a maximum of “Data checked and declared good”, code **0**.

The assumption is that the data provider, generally a National Meteorological and Hydrological Service (NMHS), is the entity having the best knowledge on its data status.

The strategy with the DAYCLI message is to share mainly checked data and avoid sharing data that is highly suspect. Nevertheless, suspect data could be shared if the data provider considers that this could have an interest.

The DAYCLI message is intended to be simple to manage and has been designed to do not include all the complexity of the data quality management in place in most Climate Data Management Systems.

The Data Quality Control information is designed for both machine-to-machine exchanges and systems using manual processes.

In the case of a data obviously false, without any interest for DAYCLI users, it is recommended that the data provider does not share this value and declares it missing e.g.

TMin	**Associated QC information**	**QC code**
Missing	Daily value not provided	6

In the case of a suspect data that the data provider wants to share e.g.

TMin	**Associated QC information**	**QC code**
32.5	Data checked and declared **suspect**	1

In the case that the data provider did not have enough time to control the data, then the data could be coded unchecked, e.g.

TMin	**Associated QC information**	**QC code**
11.2	Data unchecked	7

In the case that the data provider performed the data control, then made a modification (estimation/reconstruction/interpolation/etc.) on the data, then he could share this data with a code "Data checked and declared good".

TMin	**Associated QC information**	**QC code**
14.5	Data checked and declared good	0

1.3 Codes explanation

1.3.1 Parameter is not measured at the station

The station does not have the instrument to be able to measure this parameter. E.g. no instrument for measuring snow.

it is also possible that the instrument is non-operating at that date.

1.3.2 Daily value not provided

The data is missing for several reasons: transmission problem, observer absent, a data highly suspect that has been removed by the data provider, etc.

1.3.3 Data unchecked

No data control processes have been performed on the data.

Most data should have been undergone with format test (see WMO-No. 1269 “Guidelines on Surface Station Data Quality Control and Quality Assurance for Climate Applications”, 2021).When assigning “Data unchecked” to a data means no control has been performed on the data except format or constraint tests. E.g. consistency, statistical, spatial controls, etc.

1.3.4 Data checked and declared suspect

Data has been controlled (tests on consistency, statistical, spatial controls, etc.) and has been declared suspect by one of them.

1.3.5 Data checked and declared aggregated

Data has been controlled and has been declared aggregated.

That is typically the case when the amount of precipitation assigned for a day represents the amount of precipitation of several days. . This can also occur for snowfall amount.

In that case, the number of days from which the aggregated value corresponds to should be known.

1.3.6 Data checked and declared out of instrument range

This covers the special case where the total amount of precipitation cannot be determined, but at least exceeds the capacity of the gauge.

1.3.7 Data checked and declared aggregated and out of instrument range

This covers the situation when the amount of precipitation assigned for a day represents the amount of precipitation of several days, accompanied by an overflowed/underestimated case.

1.3.8 Data checked and declared good

Data checked by one or several processes (tests on consistency, statistical, spatial controls, etc.) and declared good by the data provider.

2 Use cases

Note: it is important to distinguish between **Parameter** (e.g. snowfall) not measured, **daily value** not provided, and **QC** not provided.

2.1 Use case: Parameter not measured at the station

Parameter not measured, e.g. snow in a tropical lowland country, or snow is only measured seasonally. In that case, QC code is 1. QC cannot be carried out.

Depth snow value	Associated QC information	QC code
Missing	Parameter is not measured at the station	5

2.2 Use case: Parameter measured but the value missing

Parameter measured, but daily **value** is missing. In this case, QC code is 2. QC cannot be carried out.

TMax	Associated QC information	QC code
Missing	Daily value not provided	6

2.3 Use case: QC not provided

Parameter is measured, daily value is present, but daily QC missing. There is no precision on the QC code. Therefore, QC code is 15. Note that this is separate from the cases where there is an aggregation, as in Use case 5.

TMax	Associated QC information	QC code
32.5	QC not provided	15

2.4 Use case: Overflowed/ Out of instrument range

Daily value is measured, but there is an overflow for a raingauge, or perhaps a blocked gauge in the case of snow, and therefore the best we can say is that we can specify a minimum amount, but no more than that. In that case, the QC code is 7.

In that case, we can understand with high confidence that the measurement exceeds a known minimum, but we don't know what the actual is/was.

Date	Precipitation	Associated QC information	QC code
Day 1	200 mm	Data checked and declared out of instrument range	3

2.5 Use case: Aggregation

Aggregation: In this case, observations are not available for some individual days, but the aggregated total (or highest/lowest value in the case of temperature) are known.

Precipitation: because of an observer problem the rain gauge has accumulated a precipitation over 4 days (from DAY 2 to DAY 5 in the below tables). The precipitation amount for this 4 days are reported at DAY 5 (32.5 mm). From DAY 2 to DAY 5 the QC information is set to 5 " Data checked and declared aggregated". That allows specifying on how many days the aggregation corresponds to. For Day 6, the problem is solved and there is no precipitation.

Date	Precipitation	Associated QC information	QC code
Day 1	200 mm	Data checked and declared out of instrument range	3
Day 2	Missing	Data checked and declared aggregated	2
Day 3	Missing	Data checked and declared aggregated	2
Day 4	Missing	Data checked and declared aggregated	2
Day 5	32.5	Data checked and declared aggregated	2
Day 6	0	Data checked and declared good	8

2.6 Use case: Aggregation with instrument in maintenance

Note also the case made by Marcus: if we know that missing data over five days, with a total on the sixth, included a period where the station was undergoing maintenance, then we could distinguish in the daily messages between "data missing" during the maintenance period, where observations did not occur (say, on Days 1-2), and aggregated (where we know that on days 3-5 measurements did occur, but we don't know the daily split). Or if the maintenance took place on Days 2-3 the string of flags would be: Day1 -aggregated; days 2-3 missing; days 4-5 aggregated (with the total aggregation on Day 5).

Date	Precipitation	Associated QC information	QC code
Day 1	Missing	Data checked and declared aggregated	2
Day 2	Missing	Parameter is not measured at the station	5
Day 3	Missing	Parameter is not measured at the station	5
Day 4	Missing	Data checked and declared aggregated	2
Day 5	32.5	Data checked and declared aggregated	2

2.7 Use case: Aggregation with overflow/out of instrument range

Here the case with aggregated precipitation and at the end an amount of precipitation that exceeds the raingauge capacity.

Aggregation from DAY 2 to DAY 5, with DAY 5 overflowed.

Date	Precipitation	Associated QC information	QC code
Day 1	3.5	Data checked and declared good	0
Day 2	Missing	Data checked and declared aggregated	2
Day 3	Missing	Data checked and declared aggregated	2
Day 4	Missing	Data checked and declared aggregated	2
Day 5	200 mm	Data checked and declared aggregated and out of instrument range	4

2.8 Use case: Missing values for proper computation

Here the case where some values are missing to be able to compute "perfectly" a daily value. It could be because:

- ✓ the AWS has had a problem and on the 1440 Minute values (60mn*24) needed to compute the daily value, some are missing;
- ✓ on the 24 values of the day, 1 or several values are missing to compute the daily value;
- ✓ on the 8 main synoptic hours of the day, 1 or several values are missing to compute the daily value.

In that case, several possibilities are offered for the data provider:

1. the data provider considers that the data is good, e.g. on the 1440 Minute values 60 are missing, he will code it as **0** : "Data checked and declared good"

TMin	Associated QC information	QC code
11.2	Data checked and declared good	0

2. the data provider considers that the data is suspect but interesting to share worldwide, e.g. on the 24 Hourly values of the day 4 are missing, he will code it as 4 "Data checked and declared suspect"

TMin	Associated QC information	QC code
11.2	Data checked and declared suspect	1

3. the data provider considers that the data is highly suspect, e.g. on the 8 main synoptic hours 6 are missing, he will assume that his data does not represent the reality and he decides to do not share this value with the world. He will assign Missing to the data with code 6 “Daily value not provided”.

TMin	**Associated QC information**	**QC code**
Missing	Daily value not provided	6