Data Levels / Qualifiers Flags

Notes by Jason Downing

- Difference in data levels
  - Key is to identify level
  - Level one has 2 pieces (QC and then gaps issues)
- Need to define levels in conjunction with data release policy
- Idea of a local data set with more extensive flags (without removing anything suspicious)
- Defined levels may be difficult to apply across al the site – publishing needs to be independent of quality
- Can we understand and agree on data levels; maybe just agree on concepts if we can not agree on level identifiers
- Existing ClimDB Flags (was it really vetted and is it comprehensive enough)
  - Addition of a verified, accepted, suspicous, missing, estimated?)

Level 0
Raw Data (unfiltered)

- Raw streaming data
  - Initial conversion to a meaningful units (deg C v. millivolts)
    - Is a datalogger calculated value a legitimate level 0? (debate here)
    - NO FLAGS
- Datalogger flags (-6999) may be removed?

Level 1A (Provisional)
Some simple level of QA/QC

- Application of some flags
- Range checking (sensor range v. historical range)
- Flags would not be a final flag but more an internal flag to be used in additional QA work
- Provisional v. publishable versions of the data (related to data access policy)
- Any flags used need to be well defined
- Need to tag data sets with the quality level description
- Levels can be a confusing issue…these need to well defined and understood among all participants
- Need to adopt an existing standard or develop an additional labeling system?
  - Commonalities within the available options
- Which groups can (have the flexibility) to adopt other standards

Level 1B (Published)
Ready for Public Consumption

- QA/QC is a site applied feature
- Develop a list of data quality qualifiers relevant to the data level
  - Level 1 flags v. final published flags
  - A possibility
- Level 1b is where we move beyond a ‘provisional’ data set
- Unlikely to change

Level 2
Quality Enhanced Data

- Gap filling introduced
- Careful attention has been applied to the data to make a clean an useful data set
- There may be several versions of level 2 used in practice

MISC.

- Where do derived data fit into this scheme?
  - Use CHUASI approach?
- This could conceivably be applied to streaming data directly into the NIS?
  - EML questions about this process

    Best practices defined to handle this