

Assignment of data quality level - references (compiled March 28, 2013)

Current outline http://wiki.esipfed.org/index.php/EnviroSensing_Cluster

Identify data level - maintain/provide products at different data levels

- Level 0 - Raw data, no QC, no data qualifiers applied (data flags)
 - Preservation of original data streams
 - Datalogger conversion of units and formats may be acceptable
- Level 1 - QC'd, calibrated data, qualifiers added
 - Provisional level (near real-time preparation)
 - Published level (delayed release)
- Level 2
 - Gap-filled, estimated, or aggregated data
 - Involves interpretation – multiple algorithms possible

Hubbard Brook NERC/SensorNIS workshop notes:

http://im.lternet.edu/sites/im.lternet.edu/files/DataLevels_BreakoutReport_20111027.pdf

NEON: (<http://www.neoninc.org/documents/513>)

Table 1: Definition of Data Product Levels for NEON.

- Level 0: Raw data from instrumental or human observations.
- Level 1: Calibrated data generally from a single instrument, observer, or field sampling area. These data may include information on data quality.
- Level 2: Combinations of level 1 data used to create a gap filled data stream that may replace a level 1 product. Generally, products at this level this will reflect a stream from a single instrument, observer, or field sampling area. Annotations will indicate the gap filling approach employed.
- Level 3: Level 1 and /or 2 data mapped on a uniform space-time grid.
- Level 4: Derived products using levels 1, 2 and/or 3 data. Products at this level may combine observations from more than one instrument, observer, and/or sampling area.

NEON Data Product Levels

NEON will publish data products for science, education, outreach, and decision support. The data products published by NEON will include raw measurement data and scientific data products from calibrated measurements(including the associated calibration data)through high-level products that combine multiple measurements and may involve considerable assumptions and interpretation. In order to organize the production and presentation of the NEON data products, NEON will use a parsimonious series of data product levels. This minimum number of levels is sufficient to inform the data users of the nature of a product and to guide internal management and cyberinfrastructure processes. A description of data product levels is presented in Table 1. It is not necessary for all data products to proceed uniformly from Level 0 to Level 4.

Data at Level 0 is the data received by the NEON cyberinfrastructure, whether it is generated by an automated instrument or entered by a human observer. The Level 0 data is provided in the original (engineering) measurement units. The Level 1 observatory data are generally the lowest level of

scientifically useful data. Those data have been through a quality control process and have been calibrated to scientific units. Higher level data products are used along with Level 1 data in scientific research. Data products at levels 2 and 3 are generally derived from Level 1 data. Level 4 data products may be derived from a combination of data products at levels 1, 2, and/or 3.

CUAHSI

<http://his.cuahsi.org/documents/ODM1.1DesignSpecifications.pdf>, pp. 19-20, 57-58

QualityControlLevelCode = "0" - Raw Data

Raw data is defined as unprocessed data and data products that have not undergone quality control. Depending on the data type and data transmission system, raw data may be available within seconds or minutes after real-time. Examples include real time precipitation, streamflow and water quality measurements.

- QualityControlLevelCode = "1" – Quality Controlled Data

Quality controlled data have passed quality assurance procedures such as routine estimation of timing and sensor calibration or visual inspection and removal of obvious errors. An example is USGS published streamflow records following parsing through USGS quality control procedures.

- QualityControlLevelCode = "2" –Derived Products

Derived products require scientific and technical interpretation and include multiple-sensor data. An example might be basin average precipitation derived from rain gages using an interpolation procedure.

- QualityControlLevelCode = "3" –Interpreted Products

These products require researcher (PI) driven analysis and interpretation, model-based interpretation using other data and/or strong prior assumptions. An example is basin average precipitation derived from the combination of rain gages and radar return data.

- QualityControlLevelCode = "4" –Knowledge Products

These products require researcher (PI) driven scientific interpretation and multidisciplinary data integration and include model-based interpretation using other data and/or strong prior assumptions. An example is percentages of old or new water in a hydrograph inferred from an isotope analysis.

These definitions for quality control level are stored in the QualityControlLevels table. These definitions are recommended for use, but users can define their own quality control level system. The QualityControlLevels table is not a controlled vocabulary, but specification of a quality control level for each data value is required. Appendix B of this document provides a discussion of how to handle data versioning in terms of quality control levels (using the levels defined above), data series editing, and data series creation.

Other data quality level references:

Ameriflux:

<http://public.ornl.gov/ameriflux/available.shtml>

Earth Science Reference Handbook

http://eosps.nasa.gov/ftp_docs/2006ReferenceHandbook.pdf p.31

ILRS Data products: (CODMAC - Committee on Data Management, Archiving and Computing)

http://ilrs.gsfc.nasa.gov/about/reports/9809_attach7b.html