QualityML: a dictionary for quality metadata encoding

Miquel Ninyerola (1), Eva Sevillano (1), Ivette Serral (2), Xavier Pons (1), Alaitz Zabala (1), Lucy Bastin (3), and Joan Masó (2)

(1) Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain, (2) CREAF, Cerdanyola del Vallès , Spain, (3) School of Engineering and Applied Science, Aston University, Birmingham , UK

The scenario of rapidly growing geodata catalogues requires tools focused on facilitate users the choice of products. Having quality fields populated in metadata allow the users to rank and then select the best fit-for-purpose products. In this direction, we have developed the QualityML (http://qualityml.geoviqua.org), a dictionary that contains hierarchically structured concepts to precisely define and relate quality levels: from quality classes to quality measurements. Generically, a quality element is the path that goes from the higher level (quality class) to the lowest levels (statistics or quality metrics). This path is used to encode quality of datasets in the corresponding metadata schemas. The benefits of having encoded quality, in the case of data producers, are related with improvements in their product discovery and better transmission of their characteristics. In the case of data users, particularly decision-makers, they would find quality and uncertainty measures to take the best decisions as well as perform dataset intercomparison. Also it allows other components (such as visualization, discovery, or comparison tools) to be quality-aware and interoperable.

On one hand, the QualityML is a profile of the ISO geospatial metadata standards providing a set of rules for precisely documenting quality indicator parameters that is structured in 6 levels. On the other hand, QualityML includes semantics and vocabularies for the quality concepts. Whenever possible, if uses statistic expressions from the UncertML dictionary (http://www.uncertml.org) encoding. However it also extends UncertML to provide list of alternative metrics that are commonly used to quantify quality.

A specific example, based on a temperature dataset, is shown below. The annual mean temperature map has been validated with independent in-situ measurements to obtain a global error of 0.5 ˚C.

Level 0: **Quality class** (e.g., *Thematic accuracy*)

Level 1: **Quality indicator** (*e.g., Quantitative attribute correctness*)

Level 2: **Measurement field** (*e.g., DifferentialErrors1D*)

Level 3: **Statistic or Metric** (*e.g., Half-lengthConfidenceInterval*)

Level 4: **Units** (*e.g. Celsius degrees*)

Level 5: **Value** (*e.g.*0.5)

Level 6: **Specifications**. Additional information on how the measurement took place, citation of the reference data, the traceability of the process and a publication describing the validation process encoded using new 19157 elements or the GeoViQua (http://www.geoviqua.org) Quality Model (PQM-UQM) extensions to the ISO models.

Finally, keep in mind, that QualityML is not just suitable for encoding dataset level but also considers pixel and object level uncertainties. This is done by link the metadata quality descriptions with layers representing not just the data but the uncertainty values associated with each geospatial element.