Geophysical Research Abstracts
Vol. 17, EGU2015-4392, 2015
EGU General Assembly 2015

Associating uncertainty with datasets using Linked Data and allowing propagation via provenance chains

Nicholas Car, Simon Cox, and Peter Fitch
Land & Water Flagship, CSIRO, Australia (nicholas.car@csiro.au)

With earth-science datasets increasingly being published to enable re-use in projects disassociated from the original data acquisition or generation, there is an urgent need for associated metadata to be connected, in order to guide their application. In particular, provenance traces should support the evaluation of data quality and reliability. However, while standards for describing provenance are emerging (e.g. PROV-O), these do not include the necessary statistical descriptors and confidence assessments. UncertML has a mature conceptual model that may be used to record uncertainty metadata. However, by itself UncertML does not support the representation of uncertainty of multi-part datasets, and provides no direct way of associating the uncertainty information – metadata in relation to a dataset – with dataset objects.We present a method to address both these issues by combining UncertML with PROV-O, and delivering resulting uncertainty-enriched provenance traces through the Linked Data API.

UncertProv extends the PROV-O provenance ontology with an RDF formulation of the UncertML conceptual model elements, adds further elements to support uncertainty representation without a conceptual model and the integration of UncertML through links to documents. The Linked ID API provides a systematic way of navigating from dataset objects to their UncertProv metadata and back again. The Linked Data API's 'views' capability enables access to UncertML and non-UncertML uncertainty metadata representations for a dataset.

With this approach, it is possible to access and navigate the uncertainty metadata associated with a published dataset using standard semantic web tools, such as SPARQL queries. Where the uncertainty data follows the UncertML model it can be automatically interpreted and may also support automatic uncertainty propagation . Repositories wishing to enable uncertainty propagation for all datasets must ensure that all elements that are associated with uncertainty (PROV-O Entity and Activity classes) have UncertML elements recorded.

This methodology is intentionally flexible to allow uncertainty metadata in many forms, not limited to UncertML. While the more formal representation of uncertainty metadata is desirable (using UncertProv elements to implement the UncertML conceptual model), this will not always be possible, and any uncertainty data stored will be better than none. Since the UncertProv ontology contains a superset of UncertML elements to facilitate the representation of non-UncertML uncertainty data, it could easily be extended to include other formal uncertainty conceptual models thus allowing non-UncertML propagation calculations.