Geophysical Research Abstracts
Vol. 13, EGU2011-9649, 2011
EGU General Assembly 2011
© Author(s) 2011

Validating water resources described in WaterML 2.0

Jonathan Yu, Peter Taylor, Simon Cox, and Gavin Walker
CSIRO, Australia ({jonathan.yu, peter.taylor, simon.cox, gavin.walker}@csiro.au)

Water observations data is a key element of a water resources information system. A data transfer standard is desirable to support ingestion and update into archives and caches, data delivery interfaces from archives and caches, and development of applications with a standard data binding or import/export mechanism. An international standard encoding for transfer of water information (WaterML 2.0) is being developed to support data transfer requirements at all phases of the ingestion, data fusion, and delivery process, and suitable for adoption internationally by the water data community. WaterML 2.0 is harmonization of CUAHSI WaterML with the OGC/ISO Observations and Measurements (O&M) data model and encoding and OGC Sensor Observation Service (SOS) interface.. Harmonization would enable delivery and consumption of water observations data using systems developed to conform to the SOS standard, enabling integration of water observations data with information from closely related domains in environmental sciences, such as geology and meteorology, where OGC-conformant systems are being deployed.

The WaterML 2.0 XML Schema is complemented by additional constraints described using Schematron, for enabling contextual constraints and cardinality checking by allows rules to incorporate assertions of XPath expressions to access and constrain element content.

A WaterML 2.0 validation service is being developed. This uses an architecture similar to the Water Data Transfer Format (WDTF) validation service, which is used in a narrower national context by Australian Bureau of Meteorology. The validation service adopts the two-pass approach. First, syntax and structural validation is performed by standard XML Schema validation tools. Second, validation of contextual constraints and code list checking is performed using a hybrid method combining context-sensitive rule-based validation (allowing the rules to be expressed within a given context) and semantic vocabulary services. A key feature of the architecture is that the vocabulary services are independent of the validation service, and accessed via a SPARQL end-point.

The distinction between the structural validation and contextual validation reflects a separation of concerns. This separation thus allows an organisation's business rules to be moved out of the XML schema definition and the XML schema to be reused by other businesses with their own specific rules. Furthermore, this approach decouples the governance and use of controlled vocabularies through a distributed architecture, as the controlled vocabularies may have wider governance implications. The proposed validation architecture allows for vocabularies to be defined and published for consumption independently of validation rules.

For WaterML 2.0 validation, there is a tension in determining the level of content rules to provide. In providing a more comprehensive set of content rules, there is a trade-off in flexibility of the format, which can restrict its usage and be more prescriptive of the use than is required. Providing a loose set of content rules will result allows for the converse, that is, greater flexibility however, less constraints on the format. Implementers will then be left to determining specific rules that apply in their use of the format. Further work is required to determine the right level of the content rules to include in the WaterML 2.0 validation service.

Also, the WaterML 2.0 validation requires a set of reference vocabularies for a use in a wider international context to be selected. Further work is required to determine these set of reference vocabularies.

Acknowledgements:

This work has, for the most part, been funded through a water information research and development alliance between CSIRO's Water for a Healthy Country Flagship and the Australian Bureau of Meteorology. The work has also been supported by The Consortium for the Advancement of Hydrological Sciences Inc. (CUAHSI).

The work is operating under the joint World Meteorological Organisation (WMO) and Open Geospatial Consortium (OGC) Hydrology Domain Working Group. The authors would like to thank the AUSCOPE team for their development and support provided of the vocabulary services.