International Conference on Hydroinformatics

8-1-2014

A Harmonized Vocabulary For Water Quality

Simon J.D. Cox

Bruce A. Simons

Jonathan Yu

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

A HARMONIZED VOCABULARY FOR WATER QUALITY

S J D COX, B A SIMONS & J YU

Environmental Information Systems, CSIRO, Highett, Australia

Interoperability of water quality data depends on the use of common models, schemas and vocabularies. However, terms are usually collected during different activities and projects in isolation of one another, resulting in vocabularies that have the same scope being represented with different terms, using different formats and formalisms, and published in various access methods. Significantly, most water quality vocabularies conflate multiple concepts in a single term, e.g. quantity kind, units of measure, substance or taxon, medium and procedure. This bundles information associated with separate elements from the OGC Observations and Measurements (O&M) model into a single slot.

We have developed a water quality vocabulary, formalized using RDF, and published as Linked Data. The terms were extracted from existing water quality vocabularies. The observable property model is inspired by O&M but aligned with existing ontologies. The core is an OWL ontology that extends the QUDT ontology for Unit and QuantityKind definitions. We add classes to generalize the QuantityKind model, and properties for explicit description of the conflated concepts. The key elements are defined to be sub-classes or sub-properties of SKOS elements, which enables a SKOS view to be published through standard vocabulary APIs, alongside the full view. QUDT terms are re-used where possible, supplemented with additional Unit and QuantityKind entries required for water quality. Along with items from separate vocabularies developed for objects, media, and procedures, these are linked into definitions in the actual observable property vocabulary.

By formalizing the model for observable properties, and clearly labelling the separate concerns, water quality observations from different sources may be more easily merged and also transformed to O&M for cross-domain applications.

INTRODUCTION AND PROBLEM STATEMENT

Observations on water resources concern a wide variety of parameters, covering both supply and quality. However, it is conventional for each agency, programme or project to manage its own list of terms denoting observation parameters, usually with little or no reference to those maintained by other organizations. As a consequence, investigations that use data from multiple sources are immediately challenged by the requirement to harmonize the terminology.

Lists of water quality parameters are available from national and local agencies. There is typically a lot of overlap between these lists, but it is hard to determine matches automatically. The vocabularies are at different levels of detail, so a single value in one list may correspond with a set of values in another. Some vocabularies conflate more than one concern in each value (e.g. chemical species + unit of measure + sampling protocol + sensor), but these may only be identified in an extended term name. Reference to entries in other vocabularies is inhibited by the absence of identifiers that can be used external to the systems within which they reside.

The goal of the work described in this paper is to enable transparent description of water observation parameters, in which the definitions have an explicit structure, with cross-references to related vocabularies where appropriate, formalized consistently with contemporary knowledge representation technology and aligned with existing systems where possible, and published through multiple web interfaces.

OBSERVATIONS AND OBSERVABLE PROPERTIES

The concept of 'observable property' is formalized in the Observations and Measurements (O&M) model and ontology [1], [2]. O&M defines a set of terms for data and metadata associated with the act of observation, the key ones being procedure, observed-property, feature-of-interest, phenomenon-time, result and result-time (Figure 1). The attributes and associations of the Observation class separate the different concerns involved in fully describing an act of observation and its result. These separate concerns can often be discerned in the conflated parameter definitions found in most of the parameter vocabularies.

In practice, a report of an observation formalized using O&M will make reference to externally managed descriptions of its observation procedure (often a sensor), the feature of interest, and the observed property. In turn, this requires separate services to deliver descriptions of these, one of which delivers an observable-property vocabulary, being a set of definitions of properties whose values may be estimated by an act of observation using a suitable procedure.

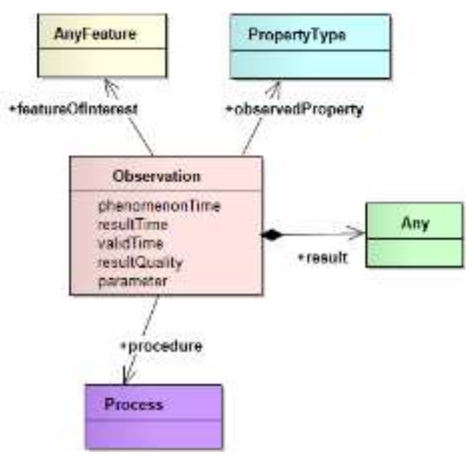


Figure 1 - Main elements of OGC Observations and Measurements model [1]

PREVIOUS MODELS FOR PROPERTIES AND QUANTITIES

The O&M model uses a class `PropertyType` to provide the target of the observed-property association (Figure 1). This class is formally defined for the ISO/TC 211 General Feature Model [3], but in O&M it is left abstract, on the understanding that a model (sometimes 'implicit') will be developed in the context of an application domain.

Nevertheless, a generic model for property-types was provided in Annex D of OGC O&M Part 1 [4], and further developed for INSPIRE [5] and by the World Meteorological

Organization [6] (Figure 2). This model allows for (a) compound property types composed of more primitive elements (b) property types in which specific constraints are applied to a base, for example by fixing some key associated parameter. In this study we do not consider the former pattern. A typical example of the latter would be 'sea surface temperature' where the base property is temperature, with the realm or feature of interest constrained to be the sea-surface. The `base` property determines the core semantics and dimensionality, and thus units of measure in the case of quantities. Consistent with the General Feature Model and O&M, the `PropertyType` class includes both quantitative and non-quantitative properties.

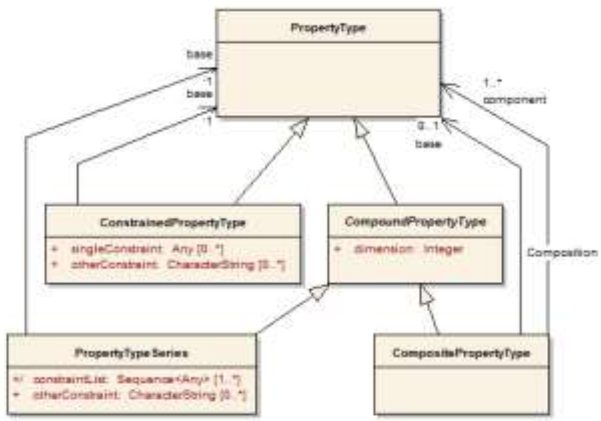


Figure 2 - Model for observed properties from OGC O&M v1 [4]

Meanwhile, Hodgson et al. [7] developed an OWL ontology [8] for systems of quantities, units, dimensions and types (QUDT). The primary goal of QUDT is to support re-scaling or unit-conversion, in an engineering context where a variety of metric and conventional units of measure are used. The analysis of quantities and units is based on the principles that underlie the SI system of units [9] and the International Vocabulary of Metrology [10]. The core of the model is shown in Figure 3, showing required relationships between a quantity value (a scaled number), its unit of measure, and the quantity kind for which it provides a value. The class `qudt:QuantityKind` captures the semantics and dimensionality, and is equivalent to the subset of property-types whose values are amounts expressed as scaled numbers. The property `qudt:generalization` links a quantity-kind to a more general one that determines the dimensionality, and is thus similar to the `base` association in the O&M property-type model (Figure 2).

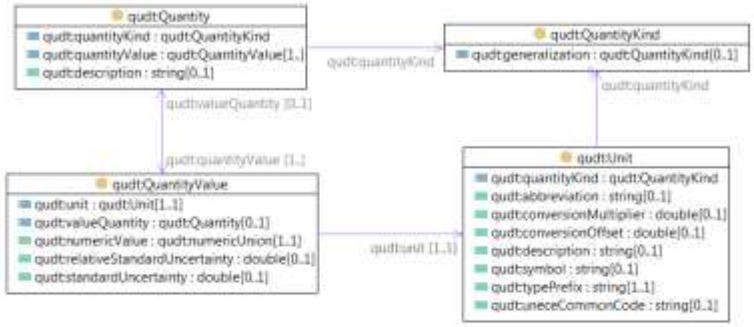


Figure 3 - Core QUDT model [7]

A hierarchy of quantity-kind classes are defined in QUDT, relating to the domain of application, along with a set of members of each class (Figure 4). The chemistry quantity-kinds are shown in Figure 4, highlighting `qudt:Concentration` which would be expected to cover a large subset of water-quality parameters. QUDT does not consider specific chemicals, but QUDT can be used without treating it as exhaustive, as the use of RDF/OWL enables other providers to extend both the ontology (the set of classes and properties) and the vocabularies (the set of individual members).

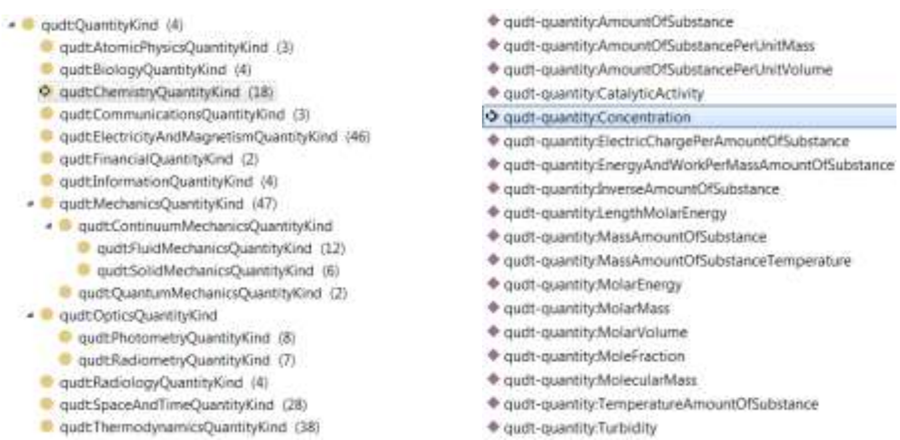


Figure 4 - Hierarchy of QUDT quantity-kind classes, and the set of chemistry quantity-kinds [7]

OBSERVABLE PROPERTY ONTOLOGY

We have defined an ontology for observable properties ("OP") which extends QUDT, incorporating some of the requirements identified in the O&M model and its successors. The core model is given in Figure 5. The class `op:ScaledQuantityKind` uses the property `qudt:unit` to link to a suitable or preferred unit of measure. `op:ScaledQuantityKind` is asserted to be equivalent to `qudt:QuantityKind`, and may be merged in a future version. A disjoint class `op:QualityKind` is defined for observable properties whose values are not scaled numbers, with the vocabularies from which values may be taken indicated with the `op:applicableVocabulary` property. The class `op:PropertyKind` generalizes all the observable-property classes, and also has RDF properties to match various concerns often included in parameter definitions. The property `op:constraint` matches the O&M v1.0 model, while `op:featureOfInterest` and `op:procedure` are similar to the observation characteristics shown in Figure 1, and `op:matrix` and `op:objectOfInterest` match other facets seen in existing vocabularies. The class `op:SubstanceOrTaxon` is the set of chemical substances and biological taxa which appear in the definition of some observable properties. The OP ontology is published at <http://environment.data.gov.au/def/op>.

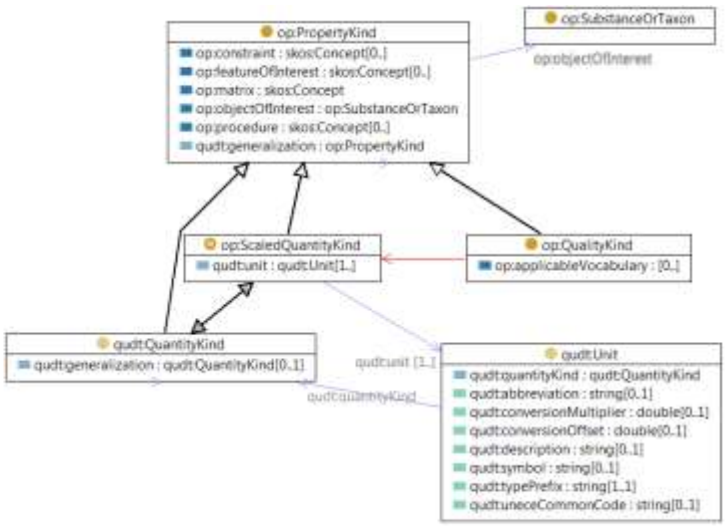


Figure 5 - Core classes in the observable property ontology

As an example of how this might be applied, a parameter in the NERC vocabulary service [11], [12], identified http://vocab.nerc.ac.uk/collection/P01/current/PCBBPCC5/, is named:

Concentration of 2,2',3,3',4,4',5,5',6-nonachlorobiphenyl {PCB206 CAS 40186-72-9} per unit wet weight of biota {Platichthys flesus (ITIS: 172894: WoRMS 127141) [Size: length >280mm Subcomponent: liver]}

This might be deconstructed using the OP ontology as follows:

generalization	= qudt-quantity:Concentration
objectOfInterest	= 2,2',3,3',4,4',5,5',6-nonachlorobiphenyl {PCB206 CAS 40186-72-9}
featureOfInterest	= biota {Platichthys flesus (ITIS: 172894: WoRMS 127141) Size: length >280mm }
matrix	= liver
constraint	= per unit wet weight

We also align the OP classes to SKOS [13], such that the observable property vocabulary may be accessed using a generic SKOS-based vocabulary API or interface, such as SISSVoc [14]. The SKOS alignment also facilitates mappings to members of other vocabularies published using SKOS, such as the NERC vocabulary service [12].

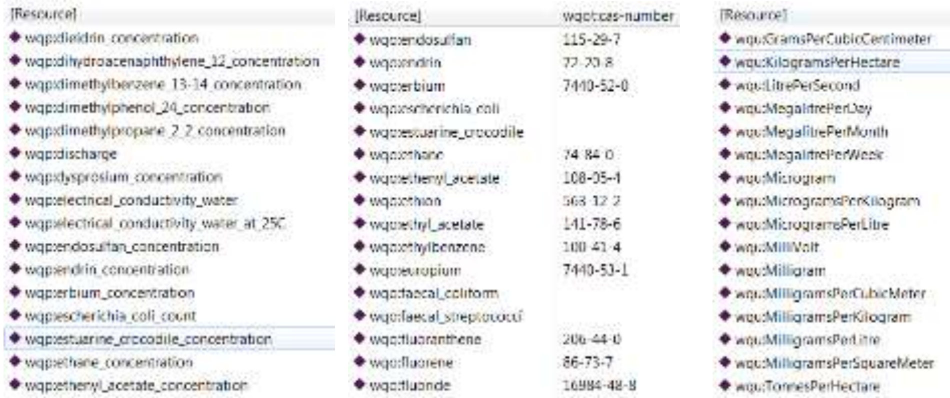
A VOCABULARY FOR OBSERVABLE WATER QUALITY PROPERTIES

We have used the OP ontology to formalize vocabularies used in some Australian water quality monitoring projects, and published them as follows:

http://environment.data.gov.au/water/quality/def/property/	- a set of observable properties
http://environment.data.gov.au/water/quality/def/object/	- substances and taxa
http://environment.data.gov.au/water/quality/def/unit/	- units of measure

http://environment.data.gov.au/water/quality/def/feature/ - features that appear in some observable property definitions.

Some representative subsets of the vocabularies are shown in Figure 6.



[Resource]	[Resource]	wqo:cas-number	[Resource]
wqp:dieldrin_concentration	wqo:endosulfan	115-29-7	wqu:GramsPerCubicCentimeter
wqp:dihydroacenaphthylene_12_concentration	wqo:endrin	72-20-8	wqu:KilogramsPerHectare
wqp:dimethylbenzene_13_14_concentration	wqo:erbium	7440-52-0	wqu:LitrePerSecond
wqp:dimethylphenol_24_concentration	wqo:escherichia_coli		wqu:MegalitrePerDay
wqp:dimethylpropane_2_2_concentration	wqo:estuarine_crocodile		wqu:MegalitrePerMonth
wqp:discharge	wqo:ethane	74-84-0	wqu:MegalitrePerWeek
wqp:dysprosium_concentration	wqo:ethenyl_acetate	108-05-4	wqu:Microgram
wqp:electrical_conductivity_water	wqo:ethion	563-12-2	wqu:MicrogramsPerKilogram
wqp:electrical_conductivity_water_at_25C	wqo:ethyl_acetate	141-78-6	wqu:MicrogramsPerLitre
wqp:endosulfan_concentration	wqo:ethylbenzene	100-41-4	wqu:MilliVolt
wqp:endrin_concentration	wqo:europium	7440-53-1	wqu:Milligram
wqp:erbium_concentration	wqo:faecal_coliform		wqu:MilligramsPerCubicMeter
wqp:escherichia_coli_count	wqo:faecal_streptococci		wqu:MilligramsPerKilogram
wqp:estuarine_crocodile_concentration	wqo:fluoranthene	206-44-0	wqu:MilligramsPerLitre
wqp:ethane_concentration	wqo:fluorene	86-73-7	wqu:MilligramsPerSquareMeter
wqp:ethenyl_acetate_concentration	wqo:fluoride	16984-48-8	wqu:TonnesPerHectare

Figure 6 – Selections from the observable properties, objects, and units vocabularies.

In Figure 7 we show some details of a single member of the /property/ vocabulary, published using a SISSvoc service [14], which provides a uniform interface that allows the user to search for items in the vocabularies, using their labels or words in their labels, and to navigate using links. The data is provided in various formats, including HTML, JSON, RDF/XML, Turtle, XML and text, which may be requested using standard HTTP methods.

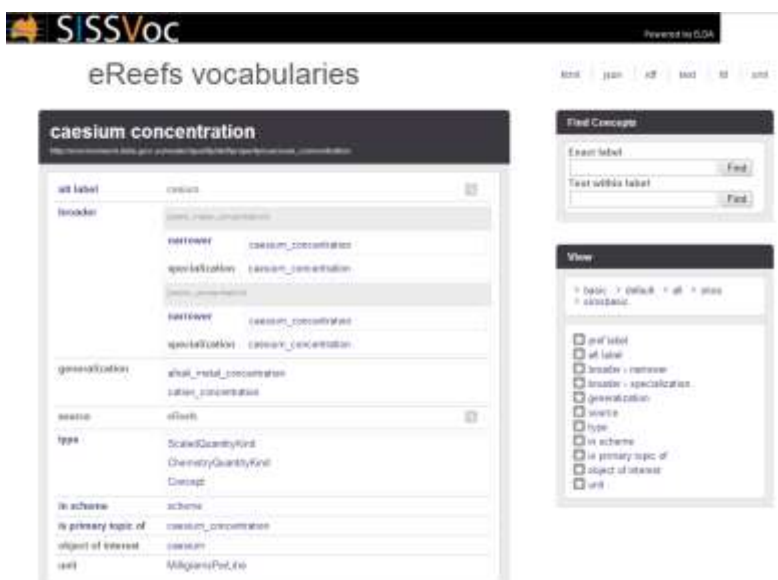


Figure 7 – Screenshot of vocabulary service user-interface

SUMMARY AND CONCLUSIONS

We have developed an OWL ontology for observable properties, which extends the QUDT ontology by the addition of

- Classes for non-quantitative properties, and scaled quantities that indicate a preferred unit of measure
- A class for substances and taxa, which often feature in observable property definitions
- RDF properties that support both general and specific constraints on property definitions, aligned with the O&M model for observations
- Sub-class and sub-property axioms to align with SKOS, to enable publication through standard vocabulary APIs.

We have used the ontology to formalize a linked set of observable property and related vocabularies, based on parameter sets used in Australian water quality projects, and have published these in a reliable and authoritative domain. Use of semantic web technology enables easy reference to the vocabularies, and will encourage adoption, supporting interoperability between data from projects that adopt these as a standard. It also enable transparent mapping to related vocabularies that are also published as web resources.

REFERENCES

[1] S. J. D. Cox, Ed., *ISO 19156:2011 - Geographic information -- Observations and measurements*. Geneva: International Organization for Standardization, 2011.

[2] S. J. D. Cox, "An explicit OWL representation of ISO/OGC Observations and Measurements," in *Proceedings of the 6th International Workshop on Semantic Sensor Networks co-located with the 12th International Semantic Web Conference (ISWC 2013)*, 2013, pp. 1–18.

[3] ISO/TC-211, *ISO 19109:2005 - Geographic information -- Rules for application schema*. Geneva: International Organization for Standardization, 2005.

[4] S. J. D. Cox, Ed., *Observations and Measurements - Part 1 - Observation schema*, 1.0 ed. Wayland, Mass.: Open Geospatial Consortium, 2007, p. 73 + xi.

[5] INSPIRE Working Group on Observations & Measurements, *Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development*, 1.0 ed. European Commission, 2011, p. 32.

[6] World Meteorological Organization, "WMO Observable Property Model," 2013. [Online]. Available: http://wis.wmo.int/AvXML/AvXML-1.0RC2/index.htm?goto=2. [Accessed: 19-Mar-2014].

[7] R. Hodgson, P. J. Keller, J. Hodges, and J. Spivak, "QUDT - Quantities, Units, Dimensions and Types v1.1," 2013. [Online]. Available: http://www.qudt.org/. [Accessed: 19-Mar-2014].

[8] J. Bao, E. F. Kendall, D. L. McGuiness, and P. F. Patel-Schneider, "OWL 2 Web Ontology Language Quick Reference Guide (Second Edition)." 2012.

[9] Organisation intergouvernementale de la Convention du Mètre, *The International System of Units*, 8e ed. Sèvres, France: Bureau international des poids et mesures, 2006, p. 186.

[10] Joint Committee for Guides in Metrology (JCGM/WG 2), *International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM) 3rd Edition*, no. Vim. Joint Committee for Guides in Metrology, 2006, pp. 1–127.

[11] A. Leadbetter, R. Lowry, and D. O. Clements, "The NERC Vocabulary Server: Version 2.0," in *Geophysical Research Abstracts 14, Proceedings EGU General Assembly 2012v*, 2012.

[12] "NERC Vocabulary Server version 2.0 (NVS2.0) at BODC." [Online]. Available: http://vocab.nerc.ac.uk/. [Accessed: 13-Feb-2014].

[13] A. Miles and S. Bechhofer, "SKOS Simple Knowledge Organization System Reference." World Wide Web Consortium, 2009.

[14] S. J. D. Cox, J. Yu, and T. Rankine, "A Linked Data API for SKOS vocabularies." *Semant. Web J.*, vol. submitted, 2014.