Defining a water quality vocabulary using QUDT and ChEBI

International Congress on Modelling and Simulation 2013

Bruce Simons, Jonathan Yu, Simon Cox
The Problem

• Inconsistent Governance
  • same concepts appear in multiple vocabularies,
  • limited relationships between terms and collections from different data providers,

• Modularity
  • difficult to incorporate vocabularies not governed by that discipline (e.g. ‘units of measure’),

• Not interoperable
  • local, non-resolvable identifiers,
  • lack of an ontology describing the relationship between concepts

• Ambiguity
  • concepts are poorly defined,
  • multiple concerns merged into one
Ambiguity in labels

“nitrogen”
“dissolved nitrogen”
“Total nitrogen, water, filtered, milligrams per liter”
“Concentration of nitrogen (total) per unit volume of the water body [dissolved plus reactive particulate phase] by oxidation and colorimetric autoanalysis”
“Concentration of nitrogen (total) per unit mass of the water body [dissolved plus reactive particulate <GF/F phase] by filtration and high temperature Pt catalytic oxidation”
“Concentration (moles or mass) of total nitrogen (i.e. nitrogen in all chemical forms) in suspended particulate material per unit volume of the water column.”
“Concentration of nitrogen (total) {'PON'} per unit volume of the water body [particulate 2-10um phase] by filtration, acidification and elemental analysis”
“Dissolved total and organic nitrogen concentrations in the water column”
“TOTAL NITROGEN (N) CONTENT”
<table>
<thead>
<tr>
<th>cas_rn number</th>
<th>ANGDTS Code</th>
<th>ANGDTS Description</th>
<th>Units_used</th>
<th>WDTF Parameter</th>
<th>chemical name</th>
<th>ADWG name</th>
<th>IUPAC name</th>
<th>Group</th>
<th>Ion</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>EC</td>
<td>ease at which conduction current can be caused to flow through material in microSiemens/cmcentimetre</td>
<td>us/cms/cmms/cm</td>
<td>ElectricalConductivityAt25C_uScm</td>
<td>Electrical Conductivity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PH</td>
<td>pH</td>
<td>negative logarithm of hydrogen ion concentration in pH units</td>
<td>pH units</td>
<td>WaterpH_pH</td>
<td>pH</td>
<td>pH</td>
<td>pH, alkalinity, acidity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16887-00-6</td>
<td>16887-00-6</td>
<td>concentration of chloride as Cl in milligrams/litre</td>
<td>mg/L</td>
<td>Chloride</td>
<td>Chloride</td>
<td>Chloride</td>
<td>Anion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TDS</td>
<td>TDS</td>
<td>the portion of total solids that passes through filter and deemed to have been dissolved in sample in milligrams/litre</td>
<td>mg/L</td>
<td>Total Dissolved Solids</td>
<td>Total Dissolved Solids</td>
<td></td>
<td></td>
<td>Salinity</td>
<td></td>
</tr>
<tr>
<td>TOTALAL KALINITY</td>
<td>ALKT</td>
<td>concentration in milligrams/litre CaCO3 of titratable bases using a methyl-orange endpoint of about pH 4.3</td>
<td>mg/L</td>
<td>Total Alkalinity (as CaCO3)</td>
<td></td>
<td></td>
<td>pH, alkalinity, acidity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HARDNE SS_CACO3</td>
<td>HARD</td>
<td>the ability of water to precipitate soap and is sum of calcium and magnesium concentrations as milligrams/litre CaCO3</td>
<td>mg/L</td>
<td>Hardness (as CaCO3)</td>
<td>Hardness (as calcium carbonate)</td>
<td></td>
<td></td>
<td>Hardness (as calcium carbonate)</td>
<td></td>
</tr>
<tr>
<td>SAR</td>
<td>SAR</td>
<td>ratio of sodium to magnesium and calcium and used to assess risk of excess sodium in irrigation water</td>
<td>Ratio</td>
<td>Sodium Adsorption Ratio</td>
<td></td>
<td></td>
<td>Salinity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3812-32-6</td>
<td>ALKC</td>
<td>alkalinity ascribed to carbonate in milligrams/litre CO3</td>
<td>mg/L</td>
<td>Carbonate Alkalinity (as CaCO3)</td>
<td>Carbonate</td>
<td>pH, alkalinity, acidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NITRATE</td>
<td>14797-55-8</td>
<td>concentration of nitrate as N in milligrams/litre</td>
<td>mg/L</td>
<td>Nitrate</td>
<td>Nitrate and Nitrite</td>
<td></td>
<td></td>
<td>Nitrate and Nitrite</td>
<td></td>
</tr>
<tr>
<td>7439-89-6</td>
<td>7439-89-6</td>
<td>concentration of iron as Fe in milligrams/litre</td>
<td>mg/L</td>
<td>Iron</td>
<td>Iron</td>
<td>Metal</td>
<td>Cation</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Methodology

1. Assess the observed water quality concepts used in the Healthy Headwaters and NGIS databases;

2. Develop an ‘Observable Property’ model based on:
   - the Queensland Healthy Headwaters and groundwater data;
   - the Open Geospatial Consortium *Observations and Measurements* property-type model;
   - the *Quantities, Units, Dimensions, Data Types v1.1* (QUDT) unified model of units and quantity kinds;

3. Align the chemistry terms with the *Chemical Entities of Biological Interest* (ChEBI) OWL ontology;

4. Use SKOS and establish a vocabulary service using Spatial Information Services Stack Vocabulary (SissVoc).
An Observation is an action whose result is an estimate of the value of some property of the feature-of-interest, obtained using a specified procedure.
Water Quality and O&M

“Concentration”

“nitrogen concentration”

“nitrogen”

Quantities, Units, Dimensions and Types (QUDT)
Water Quality extension to QUDT

- **QUDT**
  - **qudt:quantityKind**
  - **qudt:unit** [1..]

- **Water Quality**
  - **wqop:objectOfInterest**
  - **wqop:SubstanceOrTaxon**

- **<owl:equivalentClass>**
Dissolved nitrogen concentration objects

SubstanceOrTaxon

- nitrogen
  - dissolved nitrogen concentration
  - nitrogen concentration
  - elemental nitrogen (CHEBI_33267)
  - MilliGramsPerLitre
    - MolePercent
      - Concentration
        - AmountOfSubstancePerUnitVolume
Standard ontology of chemicals

>36 000 chemical entities

http://purl.obolibrary.org/obo/chebi.owl
Mapping chemical substances to ChEBI

Alternative approach to map skos concepts to ChEBI classes

- Manually create skos:exactMatch statements
- Transform subset of owl:Class to skos:Concept
- Add prov:hadPrimarySource

Subset of SubstanceOrTaxon

Imports into:
- chebi.owl
- chebi_skos.ttl
- object-chebi-mapping.ttl
- object.ttl
- imports into
- wq.ttl

Manually create skos:exactMatch statements
Observable property vocabularies should be:

1. Standardized:
   - Remove ambiguity

2. Published
   - Interoperable
   - Appropriate governance

3. Extend / re-use existing vocabularies where possible:
   - Modular

http://sissvoc.ereefs.info/search

Establish ‘procedure’ and ‘feature of interest’ vocabularies
Thank you

Environmental Information Systems
Bruce Simons

+61 3 9252 6514
bruce.simons@csiro.au
www.csiro.au/science/Environmental-Information-Systems
WQ imports

‘wq’ – imports ‘object-chebi-mapping’, ‘property’

‘object-chebi-mapping’ maps chemistry terms to ChEBI and imports ‘object’, ‘chebi-skos’

‘property’ defines “ScaledQuantityKinds” and imports ‘object’, ‘unit’

‘unit’ defines additional units and imports ‘quantity-kind’

‘object’ defines “SubstanceOrTaxon” terms and imports ‘wqop’

‘chebi-skos’ SKOS version of ChEBI

‘wqop’ defines property “objectOfInterest”

‘quantity-kind’ imports QUDT

QUDT

qudt-dimensionalunit-1.1

qudt-quantity-1.1

qudt-unit-1.1

dimension-1.1
Add water quality ScaledQuantityKind (observed property) instances

Add water quality unit of measure instances

Add water quality QuantityKind instances

Add rdfs:subClassOf = skos:Concept for qudt:QuantityKind, and qudt:Unit

Add rdfs:subPropertyOf skos:semanticRelation

Remove skos:semanticRelation properties

Export skos:semanticRelation properties

Convert skos:semanticRelation properties to rdfs:seeAlso properties

Add dbpedia annotations
<table>
<thead>
<tr>
<th>SKOS:prefLabel</th>
<th>skos:broader</th>
<th>skos:narrower</th>
<th>skos:collection</th>
<th>qudt:unit</th>
<th>wqop:objectOfInterest</th>
</tr>
</thead>
<tbody>
<tr>
<td>water pH {@en}</td>
<td>wqp:PotentialHydrogen</td>
<td></td>
<td>wqp:physical_observables</td>
<td>wqu:pH</td>
<td>sweet:matrWater.owl#L</td>
</tr>
<tr>
<td>chloride concentration {@en}</td>
<td>wqp:anion_concentration</td>
<td>wqp:halide_concentration</td>
<td>wqp:chemistry_observables</td>
<td>wqu:MilliGramsPerKilogram</td>
<td>wqo:chloride</td>
</tr>
<tr>
<td>total dissolved solids {@en}</td>
<td>wqp:Salinity</td>
<td>wqp:solids-dissolved_organic</td>
<td>wqp:physical_observables</td>
<td>wqu:MilliGramsPerLitre</td>
<td>wqo:solid</td>
</tr>
<tr>
<td>total alkalinity as CaCO3 {@en}</td>
<td>wqp:Alkalinity</td>
<td></td>
<td>wqp:physical_observables</td>
<td>wqu:MilliGramsPerLitre</td>
<td>sweet:matrWater.owl#L</td>
</tr>
<tr>
<td>total hardness as CaCO3 {@en}</td>
<td>wqp:Alkalinity</td>
<td></td>
<td>wqp:physical_observables</td>
<td>wqu:MilliGramsPerLitre</td>
<td>sweet:matrWater.owl#L</td>
</tr>
<tr>
<td>water sodium adsorption ratio {@en}</td>
<td>wqp:Salinity</td>
<td></td>
<td>wqp:chemistry_observables</td>
<td>unit:Unitless</td>
<td>sweet:matrWater.owl#L</td>
</tr>
<tr>
<td>carbonate alkalinity as CaCO3 {@en}</td>
<td>wqp:Alkalinity</td>
<td></td>
<td>wqp:physical_observables</td>
<td>wqu:MilliGramsPerLitre</td>
<td>sweet:matrWater.owl#L</td>
</tr>
<tr>
<td>nitrate N concentration {@en}</td>
<td>wqp:nitrate_nitrite_concentration</td>
<td></td>
<td>wqp:minor-trace_element_observables</td>
<td>wqu:MilliGramsPerCubicMeter</td>
<td>wqo:nitrate</td>
</tr>
<tr>
<td>iron concentration {@en}</td>
<td>wqp:transition_met_al_concentration</td>
<td>wqp:iron_ferric_concentration</td>
<td>wqp:minor-trace_element_observables</td>
<td>unit:Percent</td>
<td>wqo:iron</td>
</tr>
</tbody>
</table>

Water Quality terms as SKOS

Water Quality Vocabulary | Simons, Yu, Cox
Conceptual Model of QUDT v1