Modelling an Evolving, Geospatial SKOS Code List

Ole Voldsæter
Norwegian Social Science Data Services (NSD), Bergen, Norway
ole.voldsater@nsd.uib.no

Abstract: The success of Linked Data relies on canonical concepts and code lists that enable data from different sources to be connected and compared. But how do we create code lists that accurately describe a rapidly changing world? This paper describes a solution to the more specific problem of creating a code list that keeps track of administrative regions that may come and go.

Keywords: SKOS, code list, linked data, concept, geospatial

1. Introduction

A large number of Norwegian datasets are available on the web from several providers, such as NSD's collection of Norwegian surveys[i], the Norwegian Institute of Public Health[ii], Statistics Norway[iii] and others. This includes both micro data and aggregated (cube) data.

Many of these datasets have a geographical variable/dimension with units on two levels: "Fylke" and "Kommune", which may be translated to "county" and "municipality" respectively[iv]. There are 19 counties and 428 (as of 2013) municipalities in Norway, arranged in a hierarchical structure. Each county and municipality is designated by a name and a number. Counties are numbered 1 through 20[*] while municipalities are numbered by two digits appended to the number of the county to which the municipality belongs.

Considering the trend towards Linked Data, it would be extremely useful to have access to a canonical code list in machine readable form. Although ontologies such as GeoNames[v] exist, something more powerful is needed in order to deal with the ongoing, rapid evolution of the county/municipality hierarchy. Failure to take this into account would make it hard to compare data from different points in time.

Furthermore, for the sake of interoperability with vocabularies such as Data Cube[vi], the code list should be expressed in SKOS[vii].

2. Constant evolution

The county/municipality hierarchy undergoes constant evolution. When modelling this there are two important facts to consider:

- 100% of the land area is allocated to municipalities and counties at all times.
- The available land area has not changed for several generations, and is not expected to change in the foreseeable future.

This means that there is a limited set of changes that may occur. New municipalities cannot appear out of thin air as there is no available land. Nor can they disappear into thin air as the entire land must be covered at all times. The changes that can take place are:

- Two or more municipalities merging into a larger one
- A municipality splitting into two or more smaller ones
- A municipality being transferred to a county to which it borders
- A municipality or county changing name

Changes take place almost every year, so a given version of the code list will stay current for at most a couple of years.

[*] 13 is no longer used

3. How to model it in SKOS

We decided to design an experimental SKOS code list for which the main goal was to ensure compatibility across versions. Since changes take place almost every year, and always on January 1st, we decided to create one version per year for simplicity, even though some consecutive versions will be identical.

The code list is a SKOS concept scheme that relies on an OWL[viii] ontology that extends SKOS. The ontology contains subclasses of `skos:Concept` that represent counties and municipalities, and sub properties of `skos:RelatedMatch` that describe relations between counties and municipalities across versions.

A crucial feature is that the ontology is repeated for each yearly version. This was chosen because it is then possible to enumerate all counties and municipalities that exist in each version. It also allows the relational properties to target counties and municipalities from certain versions.

The concept scheme has one top concept, namely Norway as a whole. The 19 counties are narrower concepts relative to Norway as a whole, and each municipality is a narrower concept relative to a county. All concepts are linked to the corresponding concept in the preceding and the following version of the code list.

In most cases, a concept is linked to the next and previous versions with the `skos:exactMatch` property. This property is by definition transitive, so an arbitrary number of consecutive exact matches constitute an exact match:

```
nsd-geo:kommune903
        a       nsd-geo:KommuneNO ;
        skos:exactMatch nsd-geo-next:kommune903 ,
        nsd-geo-prev:kommune903 ;
```

When a municipalities are merged or split up, they are linked by custom properties `splitInto`, `splitFrom`, `mergedInto` and `mergedFrom`:

```
nsd-geo:kommune1248
      a       nsd-geo:KommuneNO ;
      nsd-geo:mergedInto nsd-geo-next:kommune1201 ;
      skos:exactMatch nsd-geo-prev:kommune1248 ;
```

When a municipality is renamed or transferred to a neighbouring county, it is linked by the `skos:closeMatch` property:

```
nsd-geo:kommune701
      a       nsd-geo:KommuneNO ;
      skos:changeNote "This region changed name from Borre"@en ;
      skos:closeMatch nsd-geo-prev:kommune701 ;
      skos:exactMatch nsd-geo-next:kommune701 ;
```

The advantage of this pattern is that a municipality (or county) in one version may be traced to its corresponding municipality in another version, and that it may be established whether it's exactly the same entity or the result of one or more reorganisations.

4. Applications

This pattern can be used to describe any geospatial classification system, but there are other potential use cases. It could be used to describe any evolving classification where concepts of the past, present and future are related. A medical classification of diseases, for example, evolves when our understanding of biology improves.

5. Acknowledgements

The author would like to thank Ørmulf Risnes and Archana Bidargaddi at the NSD for suggestions and encouragement

Appendix – Sample for 2002 in Turtle notation

```
@prefix nsd-geo:  <http://www.nsd.uib.no/lod/2002/geo.ttl#> .
@prefix nsd-geo-base:  <http://www.nsd.uib.no/lod/geobase.ttl#> .
@prefix nsd-geo-next:  <http://www.nsd.uib.no/lod/2003/geo.ttl#> .
@prefix nsd-geo-prev:  <http://www.nsd.uib.no/lod/2001/geo.ttl#> .
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd:     <http://www.w3.org/2001/XMLSchema#> .
@prefix owl:     <http://www.w3.org/2002/07/owl#> .
@prefix rdf:     <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix skos:    <http://www.w3.org/2004/02/skos/core#> .

nsd-geo:
      a       owl:Ontology , skos:ConceptScheme ;
      rdfs:label "Code list for counties and municipalities in Norway in
      2002"@en ;
      rdfs:seeAlso nsd-geo:RegionNO ;
      owl:priorVersion nsd-geo-prev: ;
      skos:hasTopConcept nsd-geo:norge ;
      skos:prefLabel "Code list for counties and municipalities in Norway in
      2002"@en .

# Classes for county and municipality. They are themselves SKOS concepts

nsd-geo:FylkeNO
      a       owl:Class , skos:Concept ;
      rdfs:subClassOf nsd-geo:RegionNO ;
      owl:disjointWith nsd-geo:KommuneNO ;
      skos:exactMatch nsd-geo-base:FylkeNO ;
      skos:prefLabel "Fylke"@no .

nsd-geo:KommuneNO
      a       owl:Class , skos:Concept ;
      rdfs:subClassOf nsd-geo:RegionNO ;
      owl:disjointWith nsd-geo:FylkeNO ;
      skos:exactMatch nsd-geo-base:KommuneNO ;
      skos:prefLabel "Kommune"@no .

# Enumerated superclass, specifies all valid instances for this version

nsd-geo:RegionNO
      a       owl:Class ;
      rdfs:comment "Enumerated class representing valid values for region
      scheme"@en ;
      rdfs:seeAlso nsd-geo: ;
      rdfs:subClassOf skos:Concept ;
      owl:oneOf (nsd-geo:norge nsd-geo:fylke7 nsd-geo:kommune2030 nsd-
      geo:kommune2028 nsd-geo:kommune2027 nsd-geo:kommune2025 nsd-
      geo:kommune2024 nsd-geo:kommune2023 nsd-geo:kommune2022 nsd-
      geo:kommune2021 nsd-geo:kommune2020 nsd-geo:kommune2019 nsd-
      geo:kommune2018 nsd-geo:kommune2017 nsd-geo:kommune2015 nsd-
      geo:kommune2014 nsd-geo:kommune2012 nsd-geo:kommune2011 nsd-
      geo:kommune2004 nsd-geo:kommune2003 nsd-geo:kommune2002 nsd-
      geo:kommune1943 nsd-geo:kommune1942 nsd-geo:kommune1941
(…)
```

```
# Custom properties used to track evolution of regions

nsd-geo:mergedFrom
      a       owl:InverseFunctionalProperty , owl:ObjectProperty ;
      rdfs:domain nsd-geo:RegionNO ;
      rdfs:label "Subject is the result of a merger"@en ;
      rdfs:range nsd-geo-prev:RegionNO ;
      rdfs:subPropertyOf skos:relatedMatch ;
      owl:inverseOf nsd-geo:mergedInto ;
      skos:definition "Subject is the result of a merger"@en .

nsd-geo:mergedInto
      a       owl:FunctionalProperty , owl:ObjectProperty ;
      rdfs:domain nsd-geo:RegionNO ;
      rdfs:label "Subject was merged with another region the following
      year"@en ;
      rdfs:range nsd-geo-next:RegionNO ;
      rdfs:subPropertyOf skos:relatedMatch ;
      owl:inverseOf nsd-geo:mergedFrom ;
      skos:definition "Subject was merged with another region the following
      year"@en .

nsd-geo:splitFrom
      a       owl:FunctionalProperty , owl:ObjectProperty ;
      rdfs:domain nsd-geo:RegionNO ;
      rdfs:label "Subject is the result of a split"@en ;
      rdfs:range nsd-geo-prev:RegionNO ;
      rdfs:subPropertyOf skos:relatedMatch ;
      owl:inverseOf nsd-geo:splitInto ;
      skos:definition "Subject is the result of a split"@en .

nsd-geo:splitInto
      a       owl:InverseFunctionalProperty , owl:ObjectProperty ;
      rdfs:domain nsd-geo:RegionNO ;
      rdfs:label "Subject was split up the following year"@en ;
      rdfs:range nsd-geo-next:RegionNO ;
      rdfs:subPropertyOf skos:relatedMatch ;
      owl:inverseOf nsd-geo:splitFrom ;
      skos:definition "Subjektet ble splittet opp påfølgende år"@no ,
      "Subject was split up the following year"@en .

# Listing of concepts, starting with Norway, then counties, then
# municipalities. skos:notation denotes county/municipaliry number
# while skos:prefLabel denotes name

nsd-geo:norge
      a       nsd-geo:RegionNO ;
      skos:exactMatch nsd-geo-next:norge , nsd-geo-prev:norge ;
      skos:notation "0"^^xsd:token ;
      skos:prefLabel "Norge"@no , "Norway"@en ;
      skos:topConceptOf nsd-geo: .

nsd-geo:fylke1
      a       nsd-geo:FylkeNO ;
      skos:broader nsd-geo:norge ;
      skos:exactMatch nsd-geo-prev:fylke1 , nsd-geo-next:fylke1 ;
      skos:inScheme nsd-geo: ;
      skos:notation "1"^^xsd:token ;
      skos:prefLabel "Østfold fylke"@no .
```

```
nsd-geo:fylke2
      a       nsd-geo:FylkeNO ;
      skos:broader nsd-geo:norge ;
      skos:exactMatch nsd-geo-next:fylke2 , nsd-geo-prev:fylke2 ;
      skos:inScheme nsd-geo: ;
      skos:notation "2"^^xsd:token ;
      skos:prefLabel "Akershus fylke"@no .


nsd-geo:fylke3
      a       nsd-geo:FylkeNO ;
      skos:broader nsd-geo:norge ;
      skos:exactMatch nsd-geo-prev:fylke3 , nsd-geo-next:fylke3 ;
      skos:inScheme nsd-geo: ;
      skos:notation "3"^^xsd:token ;
      skos:prefLabel "Oslo fylke"@no .
(…)

nsd-geo:kommune714
      a       nsd-geo:KommuneNO ;
      skos:broader nsd-geo:fylke7 ;
      skos:exactMatch nsd-geo-prev:kommune714 , nsd-geo-next:kommune714 ;
      skos:inScheme nsd-geo: ;
      skos:notation "714"^^xsd:token ;
      skos:prefLabel "Hof"@no .

# result of a merger
nsd-geo:kommune716
      a       nsd-geo:KommuneNO ;
      nsd-geo:mergedFrom nsd-geo-prev:kommune718 , nsd-geo-prev:kommune716 ;
      skos:broader nsd-geo:fylke7 ;
      skos:changeNote "This region changed name from Våle"@en ;
      skos:exactMatch nsd-geo-next:kommune716 ;
      skos:inScheme nsd-geo: ;
      skos:notation "716"^^xsd:token ;
      skos:prefLabel "Re"@no .
(…)
```

[i] http://nsddata.nsd.uib.no/webview/?language=en
[ii] http://khs.fhi.no/ (in Norwegian only)
[iii] http://ssb.no/en
[iv] http://en.wikipedia.org/wiki/Municipalities_of_Norway
[v] http://www.geonames.org/ontology/
[vi] http://www.w3.org/TR/vocab-data-cube/
[vii] http://www.w3.org/2004/02/skos/
[viii] http://www.w3.org/2001/sw/wiki/OWL