Nordic Vector Tile Project – Data Model

# Introduction

The Arctic SDI Geoportal currently uses a topographic basemap with a styling that was developed years ago. The styling is described in a cartographic specification that lists layers and describes their individual styling which may depend on attributes of the underlying data. The cartographic specification is written in a spreadsheet.

The topographic basemap has been implemented with WMS in the Arctic countries, and it has been cached in a WMTS (in Norway). The styling of the map is fixed – it works as a general purpose basemap, but it does not provide additional styles for the variety of needs that different groups of users may have. E.g., there could be a need for a greyscale basemap with blue waterbodies, in order not to confuse overlays with the basemap.

The Nordic Vector Tile Project seeks to provide the basemap as vector tiles. Using vector tiles, the map is rendered by the client. Consequently, it will be possible to use the same vector tiles to provide basemaps with different styles that at tailormade for different groups of users.

The approach is to place data for the vector tile basemap in a central datastore in Norway, from which vector tiles are generated and made available in a service. The Nordic countries provide the data for the vector tiles in Geo­Packages in a common data model. The data model is in the form of a UML-model. It is based on the cartographic specification such that the layers of the WMS basemap are represented as feature types in the data model. The tool ShapeChange has been used to generate database tables (DDL) and a GeoPackage template from the UML-model.

This document summarizes the discussions and dialogue about the vector tile data model.

# Design choices

## Structure of the model

The UML-model consists of class diagrams. Each layer in the cartographic specification is represented by a class. The classes are placed in packages such that each group layer (theme) is represented by a package. The model uses inheritance such that there is a common abstract BasemapElement class and the classes representing layers inherit common attributes from this class.

When making the model, it was kept in mind that the data should contain the information that was needed for reproducing the current topographic basemap using vector tiles technology. Hence the attributes are related to the styling specified in the cartographic specification. Attributes typically have values in enumerations corresponding to the values mentioned in the cartographic specification.

In addition to styling attributes, the basemap elements should include identifiers, e.g., to make it possible to update data in the future in other ways than simply replacing the entire data set. The basemap elements should also include timestamps and licensing information as metadata embedded in the data.

In the cartographic specification, there are mandatory and optional layers. There are feature types for all layers in the model, but data may not always be provided – if the country does not have data that is meaningful for the feature type, it is left out. In the model, there is no distinction between mandatory and optional layers. See also the section about empty tables below.

## Definitions

When developing data models as UML class diagrams, it is good practice to carefully write definitions of the classes (feature types). The definitions describe the data and can be valuable information for developers and users. An attempt was made to find standard definitions from data specifications that could have been used when the cartographic specification was compiled. Since the cartographic specification shared some names with well-known data specifications that had been sources of inspiration, it was a worthwhile attempt to find definitions. However, it turned out that there is no fixed definition of the data in a layer of the current basemap. The layers are provided with an approach in which the Arctic countries provides the data that fits the layer names best. I.e. there is no completely harmonized data as such, it is a best-effort approach. This approach has been adopted for the vector tile data, and the data model does not include definitions.

# Issues

## Geometry types for geographical names

The group layer with geographical names consists of seven layers with names: HydrographyNames, ProtectedAreasNames, PopulatedPlacesNames, RailwayStationNames, PortsNames, HydroreliefObjectNames and OrographyObjectsNames.

Initially, they have all been given a geometry attribute of type GM\_MultiSurface in the UML-model, i.e. multipolygon in the database as a start. Currently, this geometry type is used for place names in Greenland, so it was an intentional choice, but it could be subject to change.

It turns out that (some) place names in Greenland used to be points, but there has been a transition to multipolygons. Looking at data, some of the place names have triangles as geometries that are quite small compared to other geometries. As part of the transition, the points were converted to small triangles, and then they should be turned into other geometries by hand later. However, a new system for place names has been implemented in Greenland, so this transition process is unclear at the moment. Some place names in Greenland have been linked to curves, since it is an advantage to have a curve when visualizing e.g. a river in a map and showing the name.

Data from Finland has point geometries for a number of feature types with geographical names. An approach to fitting this data into the model is to buffer the points in order to get small polygons.

In Danish place names, there are points, curves, and (multi)polygons. (They are not part of the Arctic data, but the approach is comparable). In connection with Swedish data, it has also been mentioned that different geometry types are in use for different feature types.

There is a Danish project working with vector tiles with data from Denmark, and they are considering only using points for names, i.e., placing the labels with the geometry instead of leaving the placement to the client. Lessons learned from e.g. the Norwegian vector tiles would be helpful.

*Suggestion: Introduce three variants of all feature types to allow for different geometry types. The geometry type is indicated with a suffix: \_s, \_p, and \_c for surface (multipolygon), point, and curve, respectively.*

## PopulatedPlacesNames

The PopulatedPlacesNames contains an attribute for styling based on the size or kind of the populated place. The categorization taken from the cartographic specification contains the following elements: cities500k-1mill, towns100k-500k, towns50k-100k, towns50k, urban, rural.

When updating the WMS basemap for Greenland, it turned out that this categorization was insufficient to style populated places accordingly. In areas with sparsely populated places, there is a need for using alternative intervals for population sizes.

*Suggestion: Add the following intervals to the categorization: >6.000, 3.000-6.000, 700-2.999, 200-699, 50-199, <60.*

This should also be noted in the cartographic specification, since small populated places in Greenland that are important will occur in the basemap, but they will not fit the cartographic specification.

The data for Greenland has used the elements urban and rural in order to fit the data model. However, this has introduced a discrepancy with the styling in the WMS basemap with the newest data.

## Additional data

The topographic basemap is meant to provide a frame of reference for other data. The data for vector tiles is to be provided in scale 1:250,000 as a starting point.

When looking at transport networks, it turns out that many roads in Greenland are quite small compared to the usual interpretation of main, regional and local roads. Consequently, these roads are “scaled” to provide a basemap with the important roads.

This also gives rise to a discussion about including smaller trails and footpaths in the vector tiles. Despite their insignificant physical size, they may be very important for users in the field and provide valuable information about access to an area.

In general, one could include more feature types in the vector tiles than in the topographic basemap (WMS). It is up to the client to request the data needed for the specific purpose.

## Empty tables

Feature types may not be available from all countries.

When delivering data from Greenland, empty tables where kept in the GeoPackage when data was not available.

Are empty tables acceptable, needed or optional?

*Suggestion: Some kind of metadata with check sums etc. could be delivered together with the GeoPackage.*

## Updating data

So far, initial data deliveries have been made available for download (Iceland) and/or has been uploaded to an FTP-server in Norway.

An approach to updating data has not been discussed further. The update frequency may vary and needs to be considered together.

# Observations

## Identifiers

An observation from providing data from Greenland is that the format and availability of identifiers may vary with the feature types. There is a variety of identifiers ranging from 4-digit numbers to UUIDs (some persistent and others generated on the fly for de data delivery).

## Place names

Place name polygons are not covering the surface.

# Summary

The issues and things to consider can be summarized in a short list:

* **Geometry types for geographical names**

*Suggestion: Introduce three variants of all feature types to allow for different geometry types. The geometry type is indicated with a suffix: \_s, \_p, and \_c for surface (multipolygon), point, and curve, respectively.*

* **PopulatedPlacesNames**

*Suggestion: Add the following intervals to the categorization: >6.000, 3.000-6.000, 700-2.999, 200-699, 50-199, <60.*

* **Additional data**

Include more feature types in the vector tiles than in the basemap, e.g. smaller trails and footpaths?

* **Empty tables**

Are empty tables acceptable, needed or optional?

*Suggestion: Some kind of metadata with check sums etc. could be delivered together with the GeoPackage.*

* **Updating data**

An approach to updating data should be discussed.

SDFI, March 21st, 2024