An analysis of the harmonized base model for Spatial Data in the Netherlands for applicability in a European context

Wilko Quak[1], Marian de Vries[1], Maarten Vermeij[1], Peter van Oosterom[1],
Ko van Raamsdonk[2], Marcel Reuvers[3]
[1]*Delft University of Technology/OTB, Delft, The Netherlands*
[2]*Kadaster, Apeldoorn, The Netherlands*
[3]*GeoNovum, Amersfoort, The Netherlands*

Within the Netherlands there are many public and private parties that exchange geo-information. More and more often the data is automatically processed, transformed or integrated in complex chains of data processing. This can only be done when the data is structured in a way that can be understood by automated processes. In order to achieve this various data providers agreed to make a harmonized base model for geo-information in the Netherlands. This model (NEN3610) consists of a collection of base classes that are described in the Unified Modelling Language (UML). The model defines properties and relationship between the different base classes on which all data providers agree. Data providers that wish to use the model can extend the base model with their own organization specific model. Such an extension of the base model is called a sector model. In Figure 1 an overview of how the sector models relate to the national and international models is given.

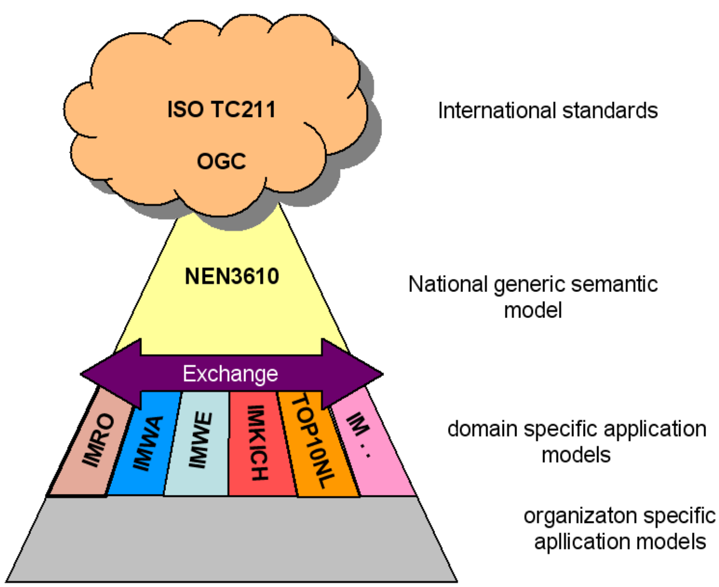

Figure 1: The NEN3610 base model for Geo-Information in relation with international standards and local implementations

This paper gives an overview of the main design issues of NEN3610 and how the base model can be extended to create sector models. In the end we discuss applicability of this model for use in the context of INSPIRE.

NEN3610, A BASE MODEL FOR THE NETHERLANDS

The development of the NEN3610 base model started in 2003 when it was decided that the previous base model (dating from 1992) was due for a revision. In the revision process shortcomings of the old model were overcome and new developments in geo-information were taken into account. A focus group of geo-information users was formed to review the work of the project group. Below we give the main characteristics of the model

The model adheres to international standards

One drawback of the old base model was that it was mainly based on national standards. As a consequence software specifically written for the Dutch market was needed to work with the model. To overcome this the new model needed to link to international standards (of OGC and ISO TC211). This model is described according to the ISO norm ISO19109 Rules for Application schema.

The model uses UML as a modeling tool

Inside the project UML class diagrams are used to describe the model. UML is a graphical modeling tool with well-defined semantics and an underlying computer model. The graphical representation of the model gives human users pictures to talk about; most figures in this document are UML diagrams. The underlying computer model enables software to process the model automatically thus enabling a model driven approach (MDA).

The model defines base classes that can be extended by sectors

The base class of the NEN3610 model is called GeoFeature and contains the characteristics that are shared by all features in the Netherlands. It is depicted in Figure 2. In the class several attributes are defined. Many of the attributes are optional; this lowers the minimum implementation burden for the model as they do not have to be implemented in a sector-model. If they are implemented however, the semantics can be understood by all parties.

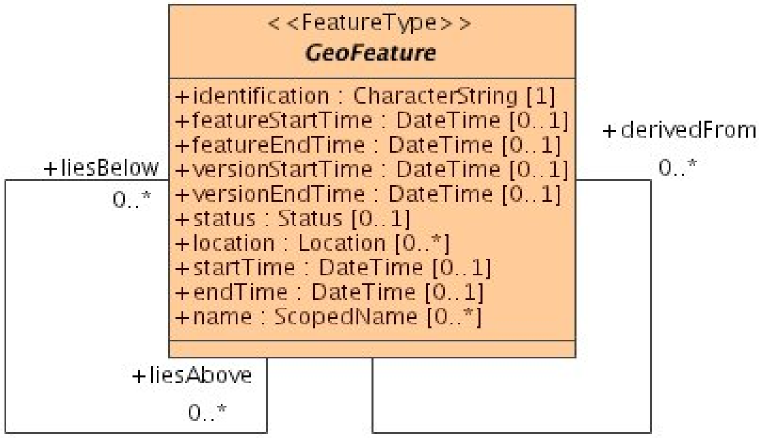

Figure 2: The base class of the NEN3610 Hierarchy

The class GeoFeature is at the top of a class hierarchy. Below the top level asubclasses are defined that represent objects that are commonly seen in the Netherlands (see Figure 3). These classes are the points where the sectors can connect to the base model by extending one of the classes in the base model. In case no class in the base model matches the needs of the sector, the base class GeoFeature should be extended.

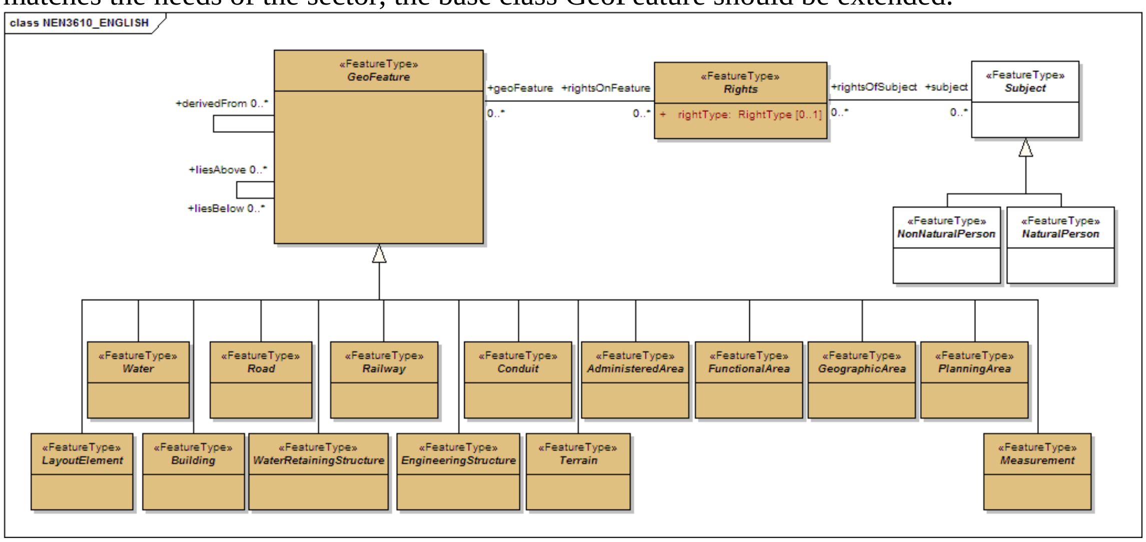

Figure 3: First level of the hierarchy of NEN3610 without details.

Every object has a unique nationwide identifier

In the base class of NEN3610 (GeoFeature) only the identification attribute is obligatory. This attribute should contain a value that is unique across all sector models in the Netherlands. The model describes a system that ensures the assignment of unique identifiers to all objects even across various data producers.

An GML application schema can be automatically derived from the model

Apart from using UML for harmonizing the data model, the model can also be used to derive a file format to exchange the data between the parties. The exchange format is defined by a GML application schema. This XML document can be automatically derived from the UML model using the rules in the ISO19136 GML document.

Every object can be traced through time

In the NEN3610 model every object has several temporal attributes. The different temporal attributes are used in different cases and have different semantics. The startTime and endTime attribute are used to track the feature in the real world whereas the other temporal (featureStartTime, featureEndTime, versionStartTime and versionEndTime) attributes are used to track the representation of the feature in the computer. Also the derivedFrom relationship in Figure 1 can be used to connect an object to its successors.

Formal constraints are used to model more semantics

In order to enhance the semantics contained within the model, it is extended with constraints. Constraints can be expressed in various manners, ranging from (informal) text to formal specifications. Part of the UML standard is the Object Constraint Language (OCL), which can be used to specify constraints in a formal manner. A (standard) XML schema does not currently support OCL constraints and the constraints are consequently not included in translating a UML schema into an XML schema. However, the constraints certainly have a function within the UML model. One approach to include the constrains in the XML encoding of the models is using Schematron (ISO/IEC 19757-3:2006). With the exchange of real data (at the level of sector models) constraints are even more important, therefore the sectors are encouraged to use constraints in de models.

Enumerations of well defined phenomena

For many attributes in the model, a predefined list of possible values is provided as part of the model in the form of enumerations or CodeLists. Depending on the level of agreement between the different parties in the model the list is a closed enumeration (only the given values are allowed, this is depicted by the <<enumeration>> stereotype) or the list is list of suggested values, that can be extended or changed in sub-models (with the <<CodeList>> stereotype. See figure 4 for an example.

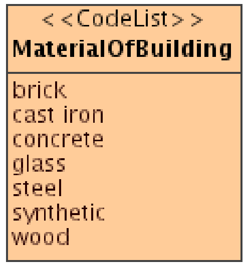

Figure 4 Example of CodeList in the NEN3610 base model

MAKING A SECTOR MODEL INSIDE THE BASE MODEL

The base model NEN3610 is only a template model. This means that no data can exist in the model (only abstract classes); the model is a collection of definitions that need to be extended by a sector to become a model in which data can be exchanged. Such a sector model is an extension of the base model. In a sector model all definitions and classes as defined in NEN3610 can be extended, refined or constricted in any way as long as the definitions in the sector model do not contradict with the base model. This means that new attributes can be added and existing attributes can be restricted to a more restrictive type or cardinality. Making an obligatory attribute optional cannot be done because an instance of that class without the attribute would not contain enough data to make a valid instance of the parent class.

A sector that wishes to model a specific phenomenon chooses the class in the base UML model that matches the phenomenon best and creates a subclass. In the case of NEN3610 this means that either one of the predefined classes should be used or if there is no applicable class, the base class GeoFeature can be extended. Within the subclass the following things can be changed in relation to its parent classes. See Figure 5 for an example of TOP10NL:

- Attributes and relationships are be added (e.g. heightClass, height, geometry, …)
- The location attribute (that was optional in NEN3610) will never be used in TOP10NL, so it is declared with [0] cardinality.
- Some attributes that were optional are made obligatory (featureStartTime)
- When there is inheritance in the base model (e.g. Road from GeoFeature) and in the domain model, then result is diamond inheritance.

Throughout the TOP10NL model constraints are used to express semantics that cannot be expressed directly in UML:

- When a road is classified as a national road, the road number must match a specific pattern.
- If the width of a road is less than two metres, then the geometry type is line, otherwise the geometry type is area.
- Only at roads that are classified as intersections the geometries of two roads may overlap.

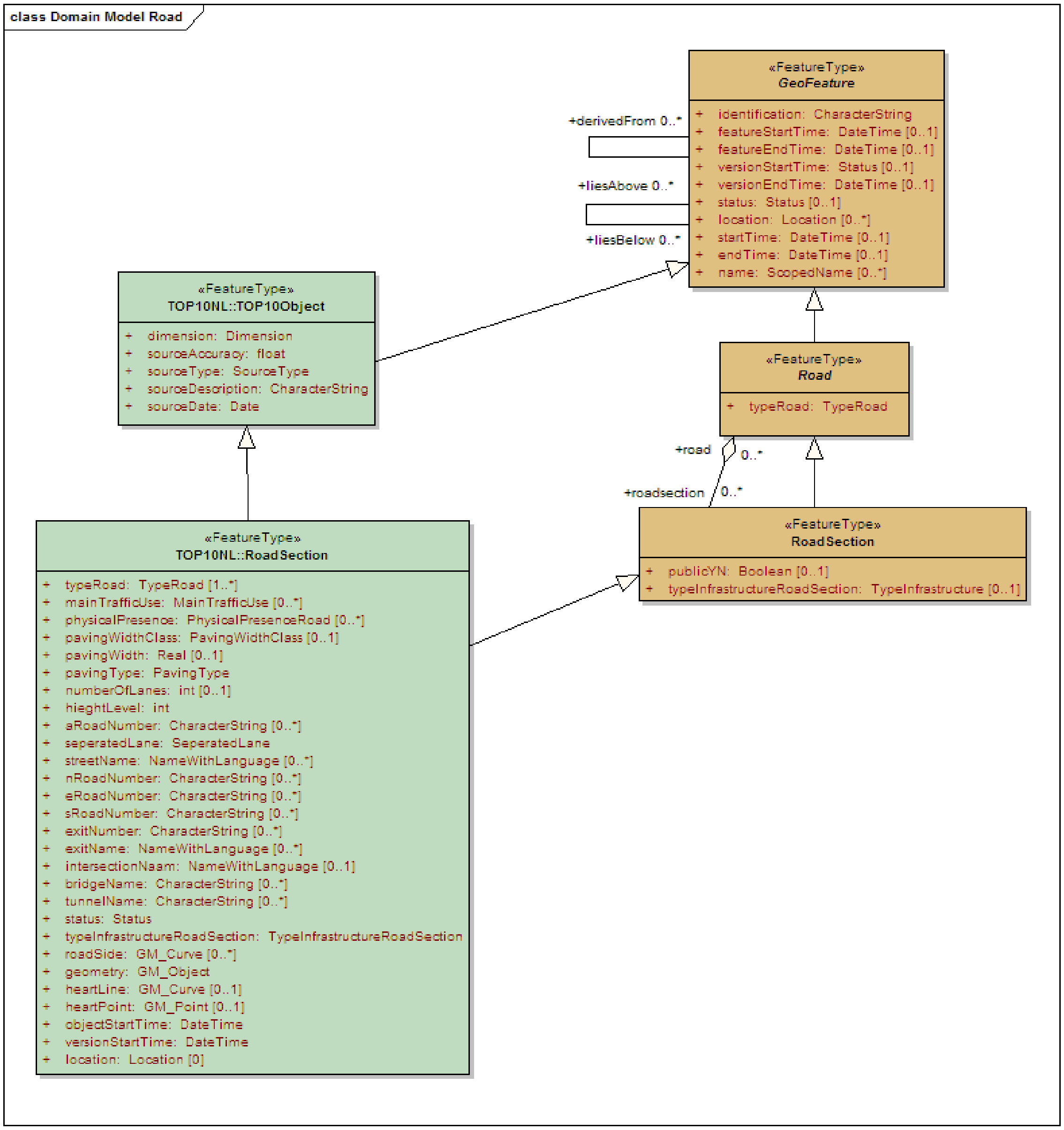

Figure 5: Extension of the NEN3610 Road base class by extension in TOP10NL (multiple/inheritance)

VALIDATION AND CONFORMANCE TESTING

To create more interoperability the quality of the data exchange should be validated. In the Netherlands this is done by validating the GML-data against the application schema (sector model.xsd). This can be done with an Internet based XML Schema parser (figure 6).

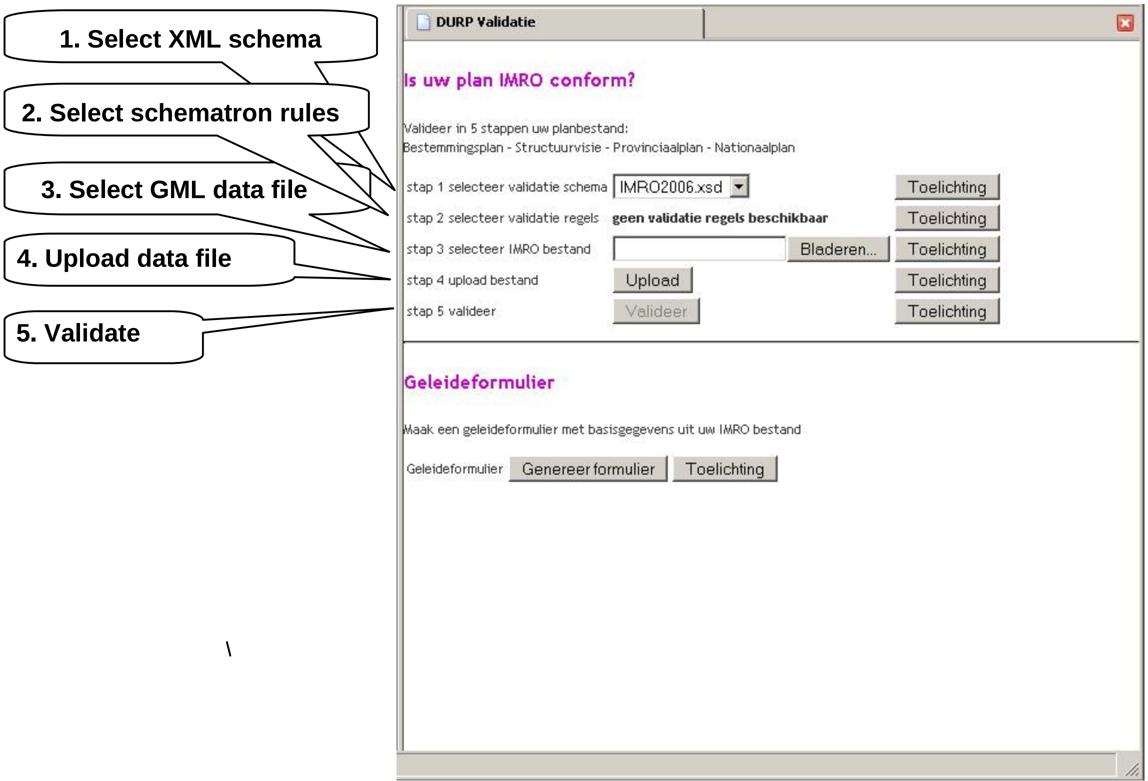

Figure 6: Validation of GML-data.via a web interface

More sector models in NEN3610 to come

TOP10NL is the first model developed as a sector model in NEN3610 since its publication in December 2005. Currently about a dozen sectors are working on their sector model and more are on their way to get started. The information modeling community grows quickly. Therefore, since 2007 multi-sector model meetings are started to align all the developments and improve interoperability between the sectors. These multi-sector meetings are the base ground to make decisions on future changes on NEN3610.

INSPIRE

Within Europe the INSPIRE 'Generic Conceptual Model' (Deliverable 2.5 of the Data Specifications Drafting Team) has similar goals as the ones behind the Netherlands NEN3610 developments. However, if one takes a look at the huge difference in the scope of the different themes (from reference systems to hydrography and from cadastral parcels to atmospheric conditions), several questions do arise: what does harmonization of the geographic information mean? what is the purpose of harmonization? And is it possible to harmonize such a wide range of themes? These were also the questions faced by the INSPIRE Data Specification Drafting Team and probably a main contribution of the Drafting Team is the identification of a set of so called harmonization components, which make the concept of harmonization more tangible. Examples of harmonization components addressed in document D2.5 are: rules for application schemas, coordinate referencing and units model, identifier management, multi-lingual text and cultural adaptability, object referencing modeling, multiple representations (levels of detail) and consistency, and more. All these components do apply to (nearly) all themes identified within INSPIRE and there is common approach to these shared components. The result is the so-called Generic Conceptual Model. Using the Generic Conceptual Model within the different themes will therefore result in a first level of harmonization. This document represents the 'Generic Conceptual Model'. See figure 7 for a possible approach for an INSPIRE base model.

Similar to the Netherlands example, the next step is that the individual themes (as defined in the Annexes I, II and III of the draft Directive) have to be modeled based on this Generic Conceptual Model. In a separate deliverable of the Drafting Team Data Specification (D2.6 'Methodology for Data Specification') the process will be specified how this should be done. The result will then be data product specifications for the individual themes, i.e. conceptual information models that describe the relevant classes, their attributes, relationships, constraints, and possibly also operations as well as other as information as appropriate like data capturing information or data quality requirements. Care has to be taken that common or shared spatial object types relevant in multiple themes are identified and modeled in a consistent manner. This could then be considered a second level of harmonization: agreement on the shared (formal) semantics between the different INSPIRE themes.

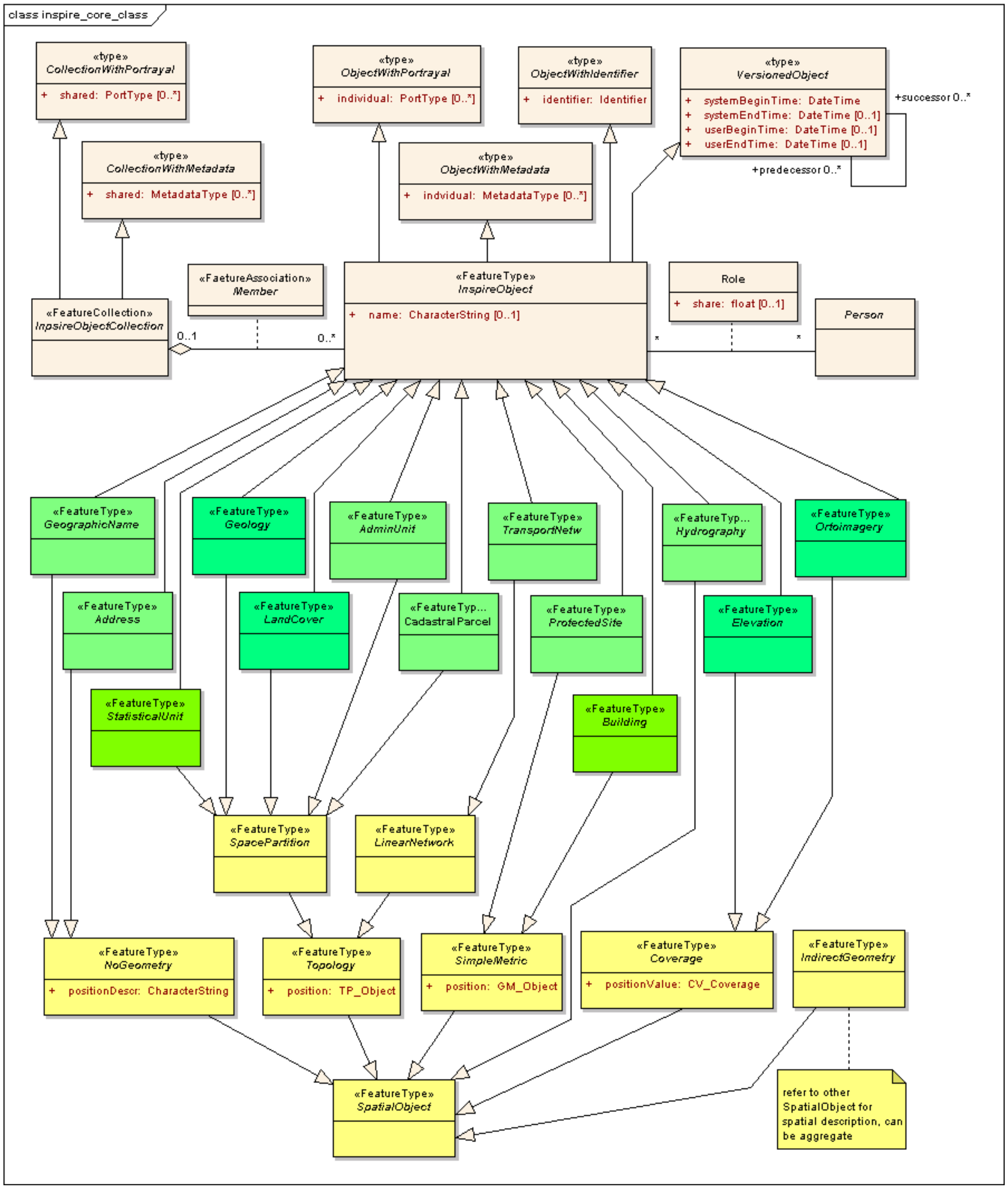

Figure 7:A possible aproach for an INSPIRE base model incliding thematic semantic

Concluding remark

The process of making a base model for geo-information and the first sector models for the Netherlands has been an interesting process that has given many new insights in modelling spatial nationwide with the use of base models and sector models. A base model to which many parties have contributed and feel responsible for is very valuable. In the Netherlands the publication of the base model has lead to more than a dozen initiatives for making sector models and a strong cooperation between the sectors for harmonization.