APPLICATION OF THE GEOUML TOOLS FOR THE PRODUCTION AND VALIDATION OF INSPIRE DATASETS

A. Belussi[1], J. Marca[2], M. Negri[2], G. Pelagatti[2]
(1) University of Verona, Verona, Italy
(2) Politecnico di Milano, Milan, Italy

Premises. This project has several premises. First, INSPIRE Data Specifications contain information models describing "attributes, relationships, constraints, and possibly also operations as well as other appropriate information like data capturing information or data quality requirements" ("Data Specifications", 2009, INSPIRE Generic Conceptual Model). The structure of INSPIRE datasets is oriented to the exchange of data, not to its storage and manipulation in a database. Therefore data transformation is required. The second premise is the existence of tools developed by Politecnico di Milano SpatialDBGroup together with CISIS (http://www.centrointerregionale-gis.it/), a National Association of local authorities, namely Regions and Provinces, in order to create and validate spatial databases:

- GeoUML Catalogue supports an operator to model in UML syntax using simple form interfaces, and produces an automatic implementation of SQL structures for PostGIS and/or Oracle spatial databases.
- GeoUML Validator allows to check the compliance of data against the specification defined with the Catalogue.

A Data Specification is stored in the Catalogue database and can be easily exported or imported as XML formatted file. The Catalogue has been used for defining the so called "Italian National Core", a Content Standard for Geo-topographic Databases (DM 10 novembre 2011 Regole tecniche per la definizione delle specifiche di contenuto dei database geotopografici. Gazzetta Ufficiale n. 48 del 27/02/2012 - Supplemento ordinario n. 37). There is therefore a great interest in extending the tools in order to produce automatically the INSPIRE datasets from the NC databases and in validating the produced datasets against constraints defined in the INSPIRE specification.

Scenario. One of the possible scenarios (but very usual in many contextes) is the following one:

- an organisation (data provider) is willing to provide WFS and GML conformant to INSPIRE specifications (services and data);
- this organisation is hosting geodata related to one or more INSPIRE themes on a spatial relational database, called here Source Database
- in order to facilitate the implementation of INSPIRE compliant GML data, the organisation implements a new "INSPIRE-structured" spatial database, called here INSPIRE Database
- a Transformation Procedure is created which extracts the data from the Source Database and loads it into the INSPIRE Database
- the INSPIRE Database is "validated" also using topological operators, in order to identify also topological constraints gaps.

We assume that both the Source Database and the INSPIRE Database are SQL based and that their physical schemas have been generated by the GeoUML Catalogue tool from the corresponding conceptual schemas, called SC_{SOURCE} and $SC_{INSPIRE}$. In this scenario the availability of the conceptual schemas suggests different areas where the tools can provide a great benefit:

1. Creation of the GeoUML specification $SC_{INSPIRE}$, automatic generation of the corresponding physical SQL structure and Validation of the INSPIRE Database with respect to the specification
2. (Semi)automatic generation of the Transformation Procedure using a set of correspondence rules between elements of SC_{SOURCE} and $SC_{INSPIRE}$
3. Automatic generation of the WFS configuration from the $SC_{INSPIRE}$

In the sequel we describe the work which has already been done and the research directions which we are following in order to deal with the first two items.

1. Some experiments have been carried out using the INSPIRE Data Specification on Transport Networks (TN). The GeoUML Catalogue has been used to define the $SC_{INSPIRE}$ of TN. Each element of TN UML schemas has been translated in GeoUML constructs and spatial constraints described as INSPIRE requirements or recommendations in TN have been formalized using GeoUML spatial integrity constraints (e.g. from Requirement 10 of TN: "*In a Transport Networks data set which contains nodes, these nodes shall only be present where Transport Links connect or end*" a GeoUML topological constraint has been obtained). Starting from the $SC_{INSPIRE}$ of TN the GeoUML Catalogue produced the SQL scripts for the TN INSPIRE database creation by storing the mapping between conceptual and physical structures. Some data have been loaded into the TN INSPIRE database. Finally, the $SC_{INSPIRE}$ with mapping has been imported in the GeoUML Validator in order to validate the TN INSPIRE database with respect to the GeoUML schema, including spatial integrity constraints, like Requirement 10. The same approach could be easily applied to other INSPIRE Themes, and validate any INSPIRE dataset.
2. A critical aspect in the described scenario is constituted by the creation of the Transformation Procedure, because this implies solving the model harmonization problem. The simplest solution is to write it manually, but this solution is expensive and error-prone. In this case it would be very important to make the Transformation Procedure reusable by many institutions having similar Source Databases, for instance the databases which are conformant with the Italian National Core. Any automatic support which can be given to this operation is of course beneficial. The Catalogue can be extended to give this support, because it already allows to compare different specifications. More important is the fact that the separation of the conceptual from the physical schema allows to distinguish between those harmonization problems which are at conceptual level and those which are simply at physical level. We think that this approach will be more powerful than the direct XSD-to-XSD mapping, where conceptual aspects and structural aspects are intermixed.