

Design of the Data Transformation Architecture for the INSPIRE Data Model Browser

A. Belussi, S. Migliorini
Department of Computer Science,
University of Verona – Italy
alberto.belussi@univr.it

P. Cipriano
EC-JRC
Ispra – Italy
pg.cipriano@gmail.com

M. Negri, G. Pelagatti
Dep. of Electronics, Information and
Bioengineering, Politecnico of Milan – Italy
giuseppe.pelagatti@polimi.it

Abstract

The INSPIRE directive requires that inside a Spatial Data Infrastructure (SDI) data are provided using a model compliant with the INSPIRE Data Model (IDM). Therefore, one of the main issues during the implementation of an SDI is the transformation of existing source databases in the way defined by the IDM. In literature, many aspects of the INSPIRE transformation problem have been studied and classified, this paper deals with the definition of the transformation architecture and the schema transformation levels.

1 Introduction

One of the main problems facing the implementation of the European Spatial Data Infrastructure foreseen by the INSPIRE directive is the transformation of data stored in existing source databases into the model defined by the INSPIRE data specifications. These specifications are formally defined by a set of application schemas, called *INSPIRE schemas*; which are based on a model, called *INSPIRE Data Model* (IDM). The IDM definition is mostly contained in the *INSPIRE generic conceptual model* [1], which in turn refers to many EN ISO 19100 Standards [2]. The aforementioned transformation may be called *INSPIRE transformation*.

The *INSPIRE Data Model Browser* (IDMB) is a tool that allows one to represent an INSPIRE schema in a form which is simpler to read by non-UML-experts. It is derived from an operational tool, the GeoUML catalogue, which has been funded by Italian local governments, produced by Politecnico of Milan with CISIS (Centro Interregionale per i Sistemi Informatici geografici e Statistici), and used for the specification of the Italian National Core content, and by Italian Regions and other public bodies for particular content specification. IDMB will be distributed freely since 2014, also through the European Commission Joinup Platform (<https://joinup.ec.europa.eu/community/are3na/home>).

This paper analyzes the data transformation problem in order to derive the requirements for extending IDMB with data transformation capabilities. This paper assumes without loss of generality that the source databases are Spatial SQL databases.

The INSPIRE transformation approaches have been extensively studied and analysed [3-6] and some tools that perform to some extent this kind of transformation are available. However, a completely satisfactory solution has not been found yet, and data producers facing with the difficult problem of choosing both methodology and tools to satisfy the INSPIRE requirements.

2 Main Transformation Approaches

Two basic types of *schema transformations* from a source schema, defined with a Source Data Model (SDM), into a target schema, defined in a Target Data Model (TDM), can be classified:

- *Content transformation*: the source and target models are the same (SDM=TDM).
- *Model transformation*: the source and target models are different (SDM≠TDM), but no content transformation is performed: the performed schema transformations do not depend on the schemas, but only on the models.

A schema transformation which is neither a content nor a model transformation is called *mixed transformation*. In general, an INSPIRE transformation requires to perform a mixed transformation where SDM=SQL, TDM=GML, source content= “Content of the source database”, and target content= “Content of INSPIRE specification”.

Since in a mixed transformation the most difficult part is due to the content aspects, it is useful to perform this part in a common model situation; i.e., to perform it as a content transformation such that SDM=TDM. Therefore, it is convenient to decompose a mixed transformation into a content transformation and model transformation. Two basic approaches that can be define by considering the possible orders of these two transformations:

- A. The *model-content approach*: first apply a model transformation of the source database to GML and then apply a content transformation using GML.
- B. The *content-model approach*: first apply a content transformation from the source database to a new database, often called the INSPIRE database, using SQL, and then apply a model transformation from the INSPIRE database to GML. This is the approach considered in this paper and its details are reported in Table 1.

The fundamental differences between the two approaches are:

- The environment where the data transformation is executed: GML in the first approach and an SQL database in the second one.
 - The model transformation (WFS configuration) must be performed for each different source database in the first approach, while it is defined only once for the INSPIRE database in the second one.
- [6] T. Reitz, “Classifying Schema Transformation Approaches and Tools”, ibidem

Table 1: The content-model transformation approach.

	Level	SDM	TDM
Content transformation	model	SQL	SQL
	schema	SQL/DDL	SQL transform.
	instance	Source DB	derived SQL/DML scripts
Model transf.	model	SQL	GML
	schema	SQL/DDL	WFS conf. (only once)
	instance	Inspire DB	GML XSD (Inspire) GML data

3 Conclusion

In the *content-model approach*, the implementation of the target classes as relational tables allows one to perform the transformations by SQL queries. The use of an SQL transformation allows one to guarantee the “stability” of new objects and identifiers. Moreover, in the SQL queries spatial indices can be used in order to improve the performance of transformations based on geometric properties. In order to automate the transformation process, it is possible to define some template, which can be used to automatically generate all transformations sharing the same basic structure. The definition of such templates has highlighted the necessity to save some partial results and perform the transformation in several steps; hence performing it in a database environment is convenient (if not necessary).

References

- [1] European Commission. (2013). *INSPIRE Generic Conceptual Model*. [Online]. Available: http://inspire.jrc.ec.europa.eu/documents/Data_Specification_s/D2.5_v3.4rc3.pdf [Accessed: 27-Nov-2013]
- [2] European Committee for Standardization. (2013). CEN/TC 287 published standards. [Online]. Available: <http://www.cen.eu/cen/Sectors/TechnicalCommitteesWorksshops/CENTechnicalCommittees/Pages/Standards.aspx?param=6268&title=CEN/TC%20287> [Accessed: 27-Nov-2013]
- [3] F. Arntsen and M. Borrebaek. (2013). “From production data base to INSPIRE data using WFS: potential methods” presented at the INSPIRE KEN workshop. [Online]. Available: <http://www.eurogeographics.org/content/inspire-ken-eurosdr-workshop> [Accessed: 27-Nov-2013]
- [4] M.L. Vautier, “Schema Transformation Concepts”, ibidem
- [5] S. Balley, S.Mustière and N. Abadie, “Involving Semantics in schema transformations”, ibidem