Studying Data Transformation in INSPIRE

Sven Schade[1], Carsten Keßler[2], Anand Nandipati[2]
[1] *Institute for Environment and Sustainability, Joined Research Centre, Ispra, Italy*
sven.schade@jrc.ec.europa.eu

[2] *Institute for Geoinformatics, University of Münster, Münster, Germany*
{carsten.kessler, a_nand01}@uni-muenster.de

The INSPIRE directive (EC 2007) obligates all EU member states to support and contribute to a common Spatial Data Infrastructure (SDI) (Nebert 2004) for Europe that provides integrated geospatial information services. These services should rely mostly on existing geospatial data sources, without significant collection of new data and re-engineering of already established databases (INSPIRE 2004). These sources have heterogeneous data models (schemas), which makes them hard to integrate. The schema transformation across various systems is hampered by differences in terminology and conceptualization, particularly when multiple information communities are involved (Schade 2010). The INSPIRE directive stresses the need to address this lack of harmonization between various geospatial datasets and the difficulty in using such datasets. INSPIRE suggests common schemas to be followed, but the major problem is how to make the existing data sources compatible with INSPIRE Data Specifications (INSPIRE 2009). Manual mapping from national data models to INSPIRE schemas is tedious and error prone. To automate this task, services are required that can transform datasets from one schema to another to be able to talk to each other (INSPIRE 2008).

The transformation process involves matching of source and target schema elements, generation of schema transformation rules, execution of the rules on source data sets, and delivery of the desired data set to the user (Chunyuan 2010, forthcoming). Dealing with schema transformation requires knowledge of involved schemas, underlying standards (such as the ISO 19100 series (ISO 2001)), used conceptual schema languages (such as UML (Rumbaugh 1999)), basic SDI concepts, and much more. In addition, novel approaches, such as the use of ontologies, start providing successive approaches in semantic interoperability and semantic matchmaking (Schade 2010). In summary, it requires well educated experts. With INSPIRE becoming reality, education and training on schema transformation and related notions is a major challenge today.

During the last years, we completed two study projects on schema transformation in INSPIRE at the Institute for Geoinformatics (IFGI), Münster. Diploma, Bachelor, and Master students in Geoinformatics were addressed. In a first phase, the participants studied existing schema transformation solutions. Our survey covered Feature Manipulation Engine (FME) (Safe Software 2008), GoPublisher (Snowflake Software 2008), Spatial Data Integrator (Camp to Camp 2008), and GeoXSLT (Klausen 2006). A brief summary of the results is given below (Table 1).

	FME	GeoPublisher	Spatial Data Int	GeoXSLT
GML support	in-built	in-built	in-built	Possible, but must be user defined
Service	WFS	WFS	not included	not included
GUI	sophisticated, graph-based	table-based	sophisticated, graph-based	not included
Rule support	more then 600 operators in-built, user may add self-defined operators	simple filters in-built, user-defined rules can be added using XSLT or Java code	many operators in-built, user may add self-defined operators (using Java)	standard XSLT operations with extensions to include spatial processing from the GeoTools library
Software	commercial	commercial	open source	open source
Comment	use of the GUI has steep learning curve	table-based GUI has quite steep learning curve, hard to include own operators	already not performing well with small data sets	generation of XSLT code inconvenient in general

Table 1: Comparison of existing schema mapping tools (Chunyuan 2010, forthcoming).

In a second phase, the students proposed an approach to perform schema transformation on expert-level and non-expert level. Focus groups were built in order to work on the architecture, concrete schema mapping rules, metadata issues, and use of ontologies. A summary of the results has been published (Beckmann 2009). In the third phase, we iterated over the architectural developments and the implementation.

We implemented the suggested approach to perform schema transformation in the context of two use cases. We translated data from a common German data model for roads (part of ATKIS (AdV 2002)) and from a data model for forest roads (NavLog 2005) to the INSPIRE Data Specification for Transport Networks (INSPIRE 2009b).

Both projects, the first covering phases one and two, and the second basically iterating over the suggested approach, were evaluated using the common schema of the University of Münster. In general, the projects were very well acknowledged by all participants. The practical relevance of the topic and qualifications for possible future employments motivated high commitment. Apart from example INSPIRE data sets and two publications (Beckmann 2009, Chunyuan 2010, forthcoming); the project also improved the participants' technical and soft skills. Seven theses (one diploma and six master theses) followed after the projects were completed.

Skills in schema transformation are highly required in the INSPIRE context, both on practical and theoretical level. With the organization of these two projects we contributed to capacity building and we hope that more projects will follow.

References

AdV (2002). ATKIS - Objektartenkatalog Basis-DLM – Version 3.1 (German). Hannover, Germany.

Beckmann. O., Blut, C., Deelmann, T., Michels, H., Osmanov, A., Roth, M., Weerasinghe, H., Wilden, M., Schade, S. (2008). Getting Inspired? Geoinformatik 2009. Osnarbrück, Germany.

Camp to Camp (2008). http://www.spatialdataintegrator.com, last accessed 01.02.2010.

EC (2007). Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE). European Commission, Official Journal of the European Union.

INSPIRE. (2004). INSPIRE scoping paper. http://www.ec-gis.org/inspire/reports/ inspire_ scoping 24 mar04.pdf, last accessed 01.02.2010.

INSPIRE (2008). Network Services Architecture – Version 3.0. 2008. INSPIRE Networking Services Drafting Team.

INSPIRE (2009). D2.5: Generic Conceptual Model, Version 3.2. INSPIRE Data Specifications Drafting Team.

INSPIRE (2009b). INSPIRE Data Specification Transport Networks – Guidelines (Version 3.0.1). INSPIRE Thematic Working Group on Transport Networks.

ISO (2001). 19101 Geographic information - Reference model. ISO/TC211 Standard.

Klausen F.M. (2006). GeoXSLT: GML processing with XSLT and spatial Extensions. Department of informatics. University of Oslo.

Lehto, L. (2007). Schema translations in a Web service based SDI. 10th AGILE International Conference on Geographic Information Science, The European Information Society: Leading the way with geo-information, Aalborg, Denmark.

NavLog (2005). Formatbeschreibung zur Erstellung einer forstspezifischen Navigationsdatenbasis - Pragmatisches Shape Forst. http://www.kwf-online.de /deutsch/ arbeit/geodat/Spezifikation_pragmatischesshapeForst _3_0.pdf (German), last accessed 01.02.2010.

Nebert, D. (2004). Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0. Global Spatial Data Infrastructure.

Rumbaugh, J., I. Jacobson, and G. Booch (1999). UML Reference Manual. Boston, MA, Addison-Wesley.

SafeSoftware (2008). Spatial ETL: Making Spatial Data Accessible. White Paper. http://www.safe.com/technology/whitepapers/spatial-etl.php, last accessed 01.02.2010.

Schade, S (2010). Ontology-Driven Translation of Geospatial Data. ISO Press, Amsterdam, Netherlands.

Snowflake Software (2008). Schema Translation in Practice. White Paper. http://83.138.131.106/eurosdr/workshops/models_2006/discussion/Schema %20Translation%20in%20Practice.pdf, last accessed 01.02.2010.