Nordic Journal of Surveying and Real Estate Research 8:1 (2011) 54–64
submitted on 29 June, 2011
revised on 19 October, 2011
accepted on 26 October, 2011

Implementation of INSPIRE Schema Transformation Service

Pekka Latvala, Lassi Lehto
Finnish Geodetic Institute, Department of Geoinformatics and Cartography,
P.O. Box 15, FI-02431, Masala, Finland
pekka.latvala@fgi.fi, lassi.lehto@fgi.fi

Abstract. *This paper studies three different approaches for the implementation of INSPIRE Schema Transformation Service. The first approach uses WPS interface for serving a custom-built transformer application relying on developed schema mapping language for defining the transformations. It performs the schema transformations on-the-fly and is implemented as an individual service. The other two approaches are based on different versions of the deegree WFS application and they use custom XSLT scripts for executing the schema transformations. First of these two approaches is incorporated inside an INSPIRE Download Service and it performs the transformations on-the-fly, whereas the other approach relies on off-line transformations and works as a Download Service that serves pre-transformed data.*

Keywords: *INSPIRE, network services, transformation service, schema transformation*

1 Introduction

The European Union (EU) has a urgent need for European-wide spatial data from various themes that are related to the protection of the environment. The data is needed for supporting the EU's environmental legislation and for enhancing the monitoring of the implementation of the related laws across the whole area of the EU. Such data should be easily available in a common, harmonized and combined form that seamlessly covers the cross-border regions between the different countries (EC, 2007).

Currently the EU member states are storing their spatial data sets in their own database structures and coordinate systems that are usually not compatible with each other. The process of changing the currently existing National Spatial Data Infrastructures (NSDI) into a common European form would require a significant amount of work and the costs of such activities would be considerable. Moreover, the member states might not be willing to go through such a process, because they have their own interests in maintaining the current infrastructures (ESDIN, 2008).

In order to solve the data interoperability issues and to increase the use, sharing and utilization of spatial data and the related services in Europe, the EU has issued a directive in 2007, called Infrastructure for Spatial Information in Europe (INSPIRE). The INSPIRE directive aims at establishing a common European Spatial Data Infrastructure (ESDI). The INSPIRE directive doesn't require the EU member states to change their current spatial data infrastructures. Instead, the common ESDI will be created on top of the existing infrastructures and the required data interoperability will be achieved through data transformations that are executed between the member states' national data forms and the common INSPIRE data forms.

The INSPIRE directive lists a number of themes, from which the member states are required to produce harmonized data. The details of each of these common data forms are described in the INSPIRE Data Specifications that ensure that the different parties produce mutually compatible data. The INSPIRE directive also requires that the member states have to produce various network services that can be used for accessing (and manipulating) the spatial data. The network service types that are listed in the INSPIRE directive are: Discovery Service, View Service, Download Service, Transformation Service and Invoke Service (EC, 2007).

This paper focuses on schema transformations that are executed on XML-based Geography Markup Language (GML) data in the context of the INSPIRE Transformation Service. The main objective of the work is to test in practice various approaches that can be taken for the implementation of the INSPIRE Schema Transformation Service and to assess their suitability and practicality for that purpose. The studied approaches use different open-source applications as platforms for the execution of the schema transformations, apply open standards for the communication between the client and the server, and are implemented with different architectural approaches. The paper addresses three scenarios. The first scenario is based on the 52North's WPS application (52North, 2011) and it uses the Web Processing Service (WPS) (OGC, 2011a) standard for handling the communications between the client and the server. In the first scenario the schema transformations are executed with a custom Java application that uses a custom-built schema transformation language for defining the transformations. The other two scenarios utilize the Web Feature Service (WFS) (OGC, 2011b) standard for handling the communications and they are based on different versions of the deegree WFS application (deegree, 2011). In the first WFS-based scenario, the schema transformations are executed on-the-fly, whereas in the second scenario the transformations are performed as an off-line process. In both cases the transformations are executed through custom-built Extensible Stylesheet Language Transformations (XSLT) (W3C, 2007) scripts.

This work was carried out as a part of the EU-project: European Spatial Data Infrastructure Network (ESDIN) (ESDIN, 2011) that helps the National Mapping and Cadastral Agencies (NMCAs) to reach the requirements of the INSPIRE directive. The actions that the ESDIN project takes include the building of various INSPIRE Network Services, helping the member states to transform the spatial data into the selected INSPIRE Annex 1 data themes, and creating Best Practice documents that recommend tools and methods for the INSPIRE implementation process.

The structure of the paper is following: the chapter 2 introduces the technical background of the topic, together with the related work. The chapter 3 presents the three studied approaches for implementing the INSPIRE Transformation Service. The paper ends with the conclusions of the work.

2 Background and related work

The execution of data transformations is one of the key elements in the process of creating the pan-European, harmonized data sets that are required by the INSPIRE directive. In general, the data transformations can be considered to take place at three conceptual levels: the syntactic level, the schematic level and the semantic level (Lehto, 2007a). At syntactic level, the transformations consist of changing the representation language of the data. At schematic level, the transformations are focused on changing data's structure and schema vocabulary. At semantic level, the transformations cannot always be executed in a straightforward manner, because the source and the target domain often contain non-matching elements. Semantic transformations are often handled with ontologies that can be used for modeling and defining the concepts of some particular domain and the relationships between them.

In general, the methods for executing the XML transformations can be divided into two categories. The first one uses Application Programming Interfaces (API) together with custom-built transformation code and the other one uses XML transformation languages. The APIs can be further divided into two main categories: event-based APIs and tree-based APIs. The event-based APIs are usually faster and thus more suitable for situations where efficiency and performance of the transformation are the key issues. The tree-based APIs are somewhat more flexible to program, but they require a lot more computer resources and therefore might not be suitable for performing on-the-fly transformations. XML transformation languages are generally a suitable approach for executing the transformations but they often yield low performance levels (Harold, 2002).

The conceptual levels of GML documents and the components of schema transformations have been identified by Lehto (Lehto, 2007a). The identified transformation components are: Filtering, Renaming, Reclassification, Merging and Splitting, Reordering, Value conversions, Morphing and Augmentation. Each of the components can be applied on different sets of conceptual levels in the GML transformations.

In the Web Service environment, the execution of schematic transformations can be generally categorized as off-line and on-the-fly transformation approaches (Lehto et al., 2011).

Off-line transformations are usually executed entirely as a pre-processing phase and the transformed data is served afterwards to the users through Web Services. The off-line transformation approach is suitable for situations where the differences between the source and the target schemas are considerable, which usually leads to the execution of heavy calculations and to long processing times. Off-line transformations are also suitable for situations where the transformation process requires some manual tuning during its execution.

The on-the-fly transformation approach is based on real-time on-line process where the client request the transformation from the server and the server executes the transformation and returns the results to clients right after the process. The on-the-fly transformation approach is suitable for situations where the results are needed immediately and where the transformations are fast enough to be performed in real-time. The limitation of this approach is that the executed calculations should be relatively simple so that they can be carried out within a reasonable time span.

The two approaches can also be combined together so that complex calculations can be executed as an off-line process and the remaining fine-tuning can be performed on-the-fly. The benefit of the combined approach is that although the overall transformation process might be heavy, some parts of it can be left customizable and the Transformation Service is still able to provide the transformed data rapidly to the client (Lehto et al., 2009).

The execution of data transformations for spatial data in the Web Service environment have been previously studied for schematic transformations by Lehto (Lehto, 2007b) and by Foerster et al. (Foerster, et al., 2010) and for semantic transformations by Donaubauer et al. (Donaubauer et al., 2007). Foerster et al. present an approach that combines map generalization and schema transformation through chained Web Services. Donaubauer et al. present a Web Service for performing semantic transformations that applies the model driven approach for executing the transformations. The chaining of geospatial Web Services has also been studied by Kiehle et al. (Kiehle et al., 2006) where a Web Service chaining approach is presented that uses Web Service orchestration to chain WPS-based processes together.

The use of ontologies in the geospatial context have been studied by Friis-Christensen et al. (Friis-Christensen et al., 2005) who utilize ontologies for automating the schema mapping process, by Fonseca et al. (Fonseca et al., 2006) who present a method for measuring the interoperability of geographical ontologies and by Gnägi et al. (Gnägi et al., 2006) who present a 4-step procedure for the execution of data transformations.

2.1 Inspire Transformation Service

The INSPIRE Transformation Service is one of the Network Service types that is recognized in the INSPIRE directive. Its purpose is to transform spatial data from the INSPIRE data providers' national data forms into the common INSPIRE data forms that are specified as XML Schemas. The INSPIRE Transformation Service can be implemented for instance as a Schema Transformation Service or as a Coordinate Transformation Service. The work presented in this paper focuses on the Schema Transformation Service. A technical guidance document has been created for the Schema Transformation Service (Howard et al., 2010) that contains information about its architecture and presents the relevant use cases for its implementation.

Data providers are not necessarily required to provide the Transformation Service, if they are otherwise able to serve INSPIRE-compliant versions of their relevant data sets. When the Transformation Service is provided, it can be created either as its own individual service, or alternatively, it can be combined with Discovery, Download or View Service. A typical approach is to combine the Transformation Service with the Download Service. This kind of service combination can be implemented with various architectural approaches. One approach is to integrate the Transformation Service inside the Download Service, so that it is not visible to the calling application. This approach removes the need for standardizing interfaces between the two services. Another approach is to build the Transformation Service on top of the Download Service as its own, individual service. This approach requires the specification of the interfaces that will be used for the communication between the Download Service and the Transformation Service. The different architectural approaches for the implementation of the INSPIRE Transformation Service are described in more detail in (INSPIRE, 2009) and in (Lehto, 2009).

2.2 Related projects

The Humboldt project (Humboldt, 2011a) is closely related to the INSPIRE implementation process and one of its goals is to produce various software tools that are useful in the process of creating the harmonized INSPIRE data sets. One of tools that has been created in the Humboldt project is the Conceptual Schema Transformer library (Humboldt, 2011b) that can be used for performing schema transformations for spatial data.

The European Commission's Joint Research Centre's Monitoring Agricultural Resources (MARS) Unit is creating an agricultural administration tool, called the "Integrated Administration and Control System" (IACS) that contains a Land Parcel Information System (LPIS) component. The LPIS component will be used for monitoring the agricultural areas in the EU. Currently, the LPIS Core Model that is a GML application schema, is being developed (Sagris and Devos, 2009). The project is planning to implement various Web Services and they have identified the need for the execution of on-the-fly schema transformations.

3 The studied implementation approaches

The implementation of the INSPIRE Schema Transformation Service was studied with three implementation scenarios. These studied scenarios are presented in this chapter.

3.1 The WPS-based approach

The first studied approach for the implementation of the INSPIRE Schema Transformation Service uses the 52North's WPS application (52North, 2011) as a platform for the custom-built, Java-based schema transformer application that performs on-the-fly schema transformations. Architecturally, the approach works as a individual Transformation Service. The schema transformations are executed in the transformer application that is placed inside the WPS application as a WPS-process. The WPS application uses the standardized WPS interface for handling the communications between the client and the server. The WPS interface specifies three operations: *GetCapabilities, DescribeProcess* and *Execute*. Generally, the *GetCapabilities* operation is executed first in order to get the general information about the WPS service and a list of the offered calculation processes. After that, the *DescribeProcess* operation is executed in order to get detailed metadata about specific process and the specific details of its execution. Finally, the WPS process can be executed with the *Execute* operation.

The custom-built transformer application performs schema transformations for GML data. The transformations are based on mapping documents that are defined with a custom-built XML-based schema mapping language. The reading of the XML documents is done in the transformer application with the Xerces parser library, with the event-based Simple API for XML (SAX) interface. The SAX is generally fast and memory-efficient API, which should make it a suitable option for executing on-the-fly schema transformations. The custom schema mapping language supports one-to-one and one-to-many transformation scenarios and it is able to perform to a certain degree all the components of the schematic GML transformations presented by Lehto (Lehto, 2007a). Figure 1 presents the architecture of the WPS-based implementation approach and the general workflow in a basic use case of the *Execute* operation where the client requests transformation from the server by providing the input data set and the mapping file.

The WPS-based transformation approach was tested with two transformation scenarios, in which GML data sets were transformed into INSPIRE schemas. The first transformation scenario was simple one-to-one mapping scenario that was executed on the Finnish place names data set of the National Land Survey of Finland between the national schema and the INSPIRE Geographical Names application schema (INSPIRE, 2010a). The transformation steps consisted mostly of straightforward mappings, string concatenations and constant value mappings. Overall the first transformation scenario was executed successfully.

The second transformation scenario was more complicated one-to-many mapping scenario. Here the transformation was executed on the Finnish road network data set, Digiroad*,* from the national schema to the INSPIRE Road Transport Networks application schema (INSPIRE, 2010b). The second transformation scenario revealed that the transformer application and the schema mapping language would have needed wider functionalities, as the source data couldn't be transformed into full compliance with the INSPIRE schema. The main reason was that the mapping language and the transformer application couldn't handle the creation of references between different output features. They would need more functionalities in order to handle these kind of references that are actually used in several INSPIRE data specifications.

Overall, the WPS-based implementation approach offers clients wide possibility to customize the results of the transformations, because they can provide the mapping files for the transformation process. However, this kind of functionality is not very relevant to be offered to the end-users because the INSPIRE application schemas are fixed and the end-users would need a thorough understanding of the source data and of the INSPIRE schemas in order to create appropriate mapping files.

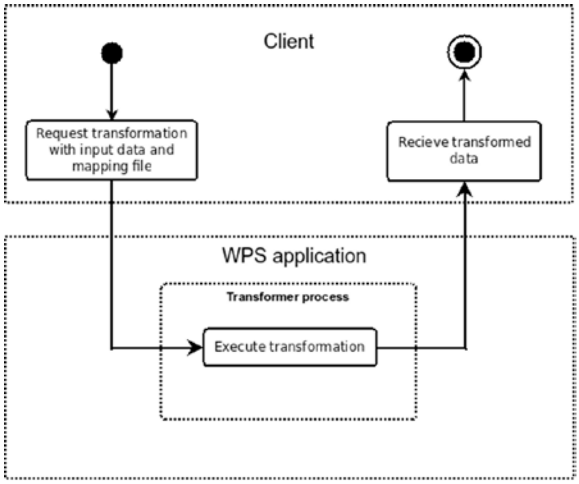

Figure 1. *Workflow of the Execute request in the WPS approach.*

3.2 The WFS-based approaches

The two WFS-based approaches for the implementation of the INSPIRE Schema Transformation Service were based on the versions 2.3 and 3.0.2 of the deegree WFS application (deegree, 2011). Both applications were installed on top of the Apache Tomcat Application Server and they both use the PostgreSQL database with the PostGIS extension to store the data. In both scenarios the schema transformations were executed with custom XSLT scripts. Both applications were tested with data that were taken from Finland and Romania and that corresponded to the INSPIRE Administrative Units (INSPIRE, 2010c) and Hydrography (INSPIRE, 2010d) Data Specifications. The main difference between the two approaches is that with deegree 2.3 the transformations are executed on-the-fly and with deegree 3.0.2 they are executed as an off-line process.

3.2.1 The deegree WFS 2.3 approach

In the deegree WFS 2.3-based approach, the source data were first loaded from Shapefiles into the database. The data were then configured to be served from the database in a simple GML 3.1.1 based encoding. The configuration was done with deegree's scripts that create .xsd files that define the appropriate mappings between the database and the GML elements. After the initial GML encoding had been created, the data was transformed on-the-fly into the INSPIRE schemas with custom-made XLST scripts. The XSLT scripts had to be written for two purposes: for transforming the incoming user queries from the INSPIRE schema into the local schema and for transforming the server responses from the local schema into the INSPIRE schema. In order to support filtering in the user queries, the GML elements were also mapped between the output data form and the inner data form. Architecturally this approach corresponds to the Transformation Service that is incorporated inside the Download Service. As a result, the end-users are not aware of any transformations and they can only see the INSPIRE-compliant data. Some problems were encountered that were related to WFS queries that used the *maxFeatures* parameter. In these cases, when the data was combined in the XSLT scripts from several database tables into the same output feature, some of the needed data was sometimes left unfetched because the feature limit was met.

The general workflow of the basic use case, where the clients request data from the server, is presented in Figure 2. First the client makes the data request in the INSPIRE form. Then the query is transformed to match the internal data form, the data is fetched from the database, transformed into the INSPIRE form and returned to the client.

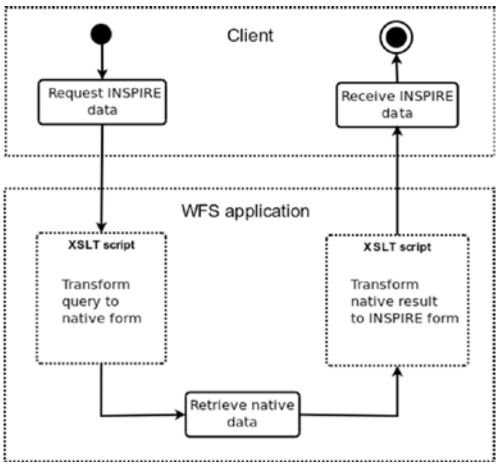

Figure 2. *Workflow of the data download requests in the WFS 2.3-based approach.*

Overall, the deegree WFS 2.3-based approach provides a working solution for the problem of providing INSPIRE-compliant data through the use of on-the-fly schema transformations. For the end-users, it is easier to use than the WPS-based approach, but it offers no control of the transformation process. The main benefit of the approach is that the original data can be updated in the database and the update process has no effect on the transformation process as long as the database structure doesn't change. Another benefit is that the same database may be used for many kinds of transformations by creating different sets of transformation files.

3.2.2 The deegree WFS 3.0.2 approach

The third implementation approach was based on the deegree WFS 3.0.2 inspireNode package. It contains ready-to-use implementations of the INSPIRE View and Download Services and it supports directly all of the INSPIRE Annex 1 data themes. In this approach, the schema transformations were carried out as an off-line process. Architecturally the deegree WFS 3.0.2 application works as a INSPIRE Download Service that serves pre-transformed data through the

WFS interface. In the transformation process, the source data were first converted with the FME Universal Translator application from Shapefiles into initial GML encoding and then transformed into the INSPIRE schemas with custom-built XSLT scripts. The transformed data was loaded into the database and deegree was configured to serve the data through its WFS functionality.

The basic workflow of data download queries is described in Figure 3, in which the clients request the pre-transformed data in the INSPIRE form and the data is served directly from the database back to the clients.

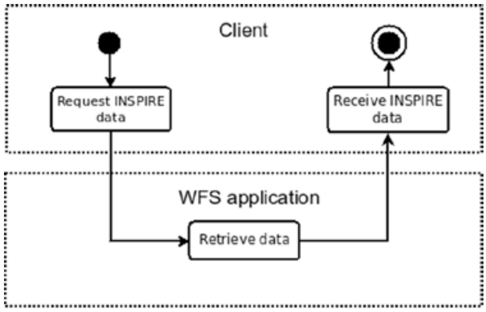

Figure 3. *Workflow of the data download requests in the WFS 3.0.2-based approach.*

The deegree WFS 3.0.2-based approach provides a working solution for serving pre-transformed INSPIRE-compliant data. The main benefit of using the off-line transformation approach is that the queries are fast to perform, because the transformations don't have to be executed in real-time. The main problem with this approach is that the data has to be re-transformed and re-loaded into the database whenever it is updated or when the transformations are changed. Therefore this approach is not very suitable for serving data that is updated often. Like the deegree WFS 2.3-based approach, also this approach doesn't give any control of the transformations to clients.

4 Conclusions

The three presented approaches provide different viewpoints to the implementation of the INSPIRE Schema Transformation Service. The studied approaches cover off-line and on-the-fly transformation scenarios and they rely on different architectural strategies in their implementation.

All of the approaches have their own pros and cons and not any single solution could be recommended as best for all kinds of situations. The main usage scenarios of the Transformation Service, as well as the general amount and the update rate of the data, should be taken into account when the implementation strategy of the INSPIRE Transformation Service is evaluated. If the source data is updated often, a Download Service that performs on-the-fly transformations offers the most flexible solution. If the updating happens rarely, off-line transformations might be a better solution. For supporting service interoperability and in order to make the services easy to use, it is important that the produced services support standardized interfaces in the communications between the client and the server.

Acknowledgements. This work was carried out as a part of EU-funded project: European Spatial Data Infrastructure Network (ESDIN).

References

52North. (2011). 52n WPS. Available at: http://52north.org/maven/project-sites/wps/52n-wps-webapp/

Deegree. (2011). deegree project web page. Available at: http://www.deegree.org/

Donaubauer, A., Straub, F., Schilcher, M. (2007). mdWFS: A Concept of Web-enabling Semantic Transformation. Proceedings of the 10th Agile International Conference on Geographic Information Science. Aalborg, Denmark.

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). Available at:http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2007:108:0001:0014:EN:PDF

ESDIN. (2008). Annex 1, Description of Work, ESDIN, Underpinning the European Spatial Data Infrastructure with a Best Practice Network. Available at: http://www.esdin.eu/project/esdin-project-annex-1-description-work

ESDIN. (2011). ESDIN, European Spatial Data Infrastructure Network. Project web page. Available at: http://www.esdin.eu/

Foerster, T., Lehto, L., Sarjakoski, T., Sarjakoski, L. T., Stoter, J. (2010). Map generalization and schema transformation of geospatial data combined in a Web Service context. Computers, Environment and Urban Systems, 34(1), 79–88.

Fonseca, F., Câmara, G., Monteiro, A. M. (2006), A Framework for Measuring the Interoperability of Geo-Ontologies. Spatial Cognition & Computation, Vol. 6, No. 4, 2006.

Friis-Christensen, A., Schade, S., Peedell, S. (2005). Approaches to solve schema heterogeneity at the European level. Proceedings of the EC-GI & GIS Workshop, Alghero, Sardinia, Jun 29 – Jul 1, 2005.

Gnägi, H. R., Morf, A., Staub, P. (2006). Semantic Interoperability through the Definition of Conceptual Model Transformations. Proceedings of the AGILE Conference on Geographic Information Science, 2006, Visegrád.

Harold E. R. (2002). Processing XML with Java. Available at: http://www.cafeconleche.org/books/xmljava/

Howard, M., Payne, S., Sunderland, R. (2010). Technical Guidance for the INSPIRE Schema Transformation Network Service. Available at: http://inspire.jrc.ec.europa.eu/documents/Network_Services/JRC_INSPIRE-TransformService_TG_v3-0.pdf

Humboldt. (2011a). Humboldt project web page. Available at: http://www.esdi-humboldt.eu/home.html

Humboldt. (2011b). Conceptual Schema Transformer. Available at: http://community.esdi-humboldt.eu/projects/show/cst

INSPIRE. (2009). Draft Implementing Rules for INSPIRE Transformation Services.

Available at: http://www.esmis.government.bg/dox/inspire/4.2/INSPIRE_Draft_Implementing_Rules_Transformation_Services_(version_3.0).pdf

INSPIRE. (2010a). INSPIRE Data Specification on Geographical Names – Guidelines. Available at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_GN_v3.0.1.pdf

INSPIRE. (2010b). INSPIRE Data Specification on Transport Networks – Guidelines. Available at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_TN_v3.1.pdf

INSPIRE. (2010c). INSPIRE Data Specification on Administrative units – Guidelines. Available at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_AU_v3.0.1.pdf

INSPIRE. (2010d). INSPIRE Data Specification on Hydrography – Guidelines. Available at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_HY_v3.0.1.pdf

Kiehle, C., Greve, K., Heier, C. (2006). Standardized Geoprocesisng – Taking Spatial Data Infrastructures one Step Further. Proceedings of the 9th AGILE International Conference on Geographic Information Science (pp. 273–282).

Lehto, L. (2007a). Schema Translations in a Web Service Based SDI. Proceedings of the 10th AGILE Conference on Geographic Information Science, 'The European Information Society: Leading the way with geo-information', May 8–11, 2007 Aalborg, Denmark.

Lehto, L. (2007b). Real-Time Content Transformations in a Web Service-Based Delivery Architecture for Geographic Information. Doctoral dissertation, Helsinki University of Technology. Publications of the Finnish Geodetic Institute, N:o 138, 2007, Kirkkonummi, 51 p. + 99 p.

Lehto, L. (2009). Real-time Content Transformations in the European Spatial Data Infrastructure. Proceedings of the 24th International Cartographic Conference, Nov 15–21, 2009, Santiago, Chile.

Lehto, L. et al. (2009). Best Practices for Content Transformations Enabling INSPIRE-Compliant Data Delivery. Deliverable 11.1. Public report. ESDIN, European Spatial Data Infrastructure Network, ECP-2007-GEO-317008, 44 p.

Lehto, L. et al. (2011). Recommendations for Operational Deployment of Services. Deliverable 11.5. Public report. ESDIN, European Spatial Data Infrastructure Network, ECP-2007-GEO-317008, Feb 28, 2011, 67 p.

OGC. (2011a). Web Processing Service. Available at: http://www.opengeospatial.org/standards/wps

OGC. (2011b). Web Feature Service. Available at: http://www.opengeospatial.org/standards/wfs

Sagris, V., Devos, W. (2009). Core Conceptual Model for Land Parcel Identification System (LCM). GeoCAP technical specification, version 1.1. Available at: http://mars.jrc.ec.europa.eu/mars/content/download/1676/9118/file/10272pubsy.pdf

W3C. (2007). XSL Transformations (XSLT) Version 2.0. Available at: http://www.w3.org/TR/xslt20/