Harmonising Dutch National Geodata Conformant to INSPIRE Using Combined Transformation*

Ebrahim Hemmatnia¹, Just van den Broecke², Ko van Raamsdonk³ Cadastre, Land Registry and Mapping Agency, Apeldoorn, The Netherlands^{1, 3} Just Objects, Amstelveen, The Netherlands²

Abstract

This paper addresses schema and coordinate transformations together with INSPIRE-compliant data delivery services of Dutch national geodata using Combined Transformation (CT). This transformation approach, originating from the ESDIN Best Practices project (Lehto et al., 2009), combines both offline and on-the-fly transformations. Dutch national data sets for addresses and cadastral parcels were successfully transformed into INSPIRE data models. A Web Feature Service (WFS) and a Web Mapping Service (WMS) were developed as data delivery services. This research was performed using Free and Open Source Software (FOSS) within the Netherlands Cadastre, Land Registry and Mapping Agency.

Keywords: INSPIRE, ESDIN, Geodata, FOSS, Combined Transformation, OGC, Web services

1 INTRODUCTION

The use of public geodata in environmental matters is continuously increasing worldwide. Geodata is of great importance in many sectors of society and has many applications including forestry planning, environmental analysis, transport planning, directing and managing emergency and rescue services and total defence (Jönsson, 2008). Within commercial, governmental and research institutions many geo-applications are being developed in order to make it easier to carry out daily activities. Current EU developments have led to a growth in both the number of trans-boundary cooperation projects and the need for transboundary geospatial analyses in cross-border regions (Garnter and Ortag, 2009).

^{*} This work is licensed under the Creative Commons Attribution-Noncommercial-Nonderivative Works 3.0 License. To view a copy of this license, visit <u>http://creativecommons.org/licenses/by-nc-nd/3.0/</u> or send a letter to Creative Commons, 543 Howard Street, 5th Floor, San Francisco, California, 94105, USA.

DOI: 10.2902/1725-0463.2010.05.art15

However, large scale cross-border geodata are not yet accessible on a European level. Availability and usability of heterogeneous geodata have a substantial impact on success and efficiency of cross-border mapping (Witschas, 2005). Full realisation of societal benefits can only be achieved if shared geodata are fully interoperable, usable and understandable by the global, interdisciplinary community. Having access to shared and up-to-date geodata can assist decision makers, such as those in public health and security, to solve cross-border issues which have an impact on the European environment. Combined Transformation (CT) may also aid in merging data as different members may even publish to the same database. "Geodata harmonisation" means standardised access to geodata that originate from various sources. However, most, if not all, member states of the European Union have local standards, which make sharing geodata across borders tremendously difficult. This issue is addressed by the INSPIRE Directive (2007/2/EC).

This paper describes recently gained knowledge and skills for the realisation of harmonising Dutch national geodata conformant to INSPIRE specifications and making these data available through web-services as standardised by the Open Geospatial Consortium (OGC), namely the Web Map Service (WMS) and Web Feature Service (WFS). Within this approach, two specific themes from INSPIRE Annex I (EC, 2007) were examined: a) Addresses (AD) and b) Cadastral Parcels (CP), using Free and Open Source Software (FOSS) for the implementation of INSPIRE within the Netherlands Cadastre, Land Registry and Mapping Agency.

The following six sections are included in this paper: Section 1 gives an introduction to this research. Section 2 includes a background of INSPIRE and also describes INSPIRE transformation approaches. Section 3 contains an introduction to the European Spatial Data Infrastructure Network (ESDIN) and introduces the approaches for transformation within ESDIN. Section 4 describes Combined Transformation applied to Dutch geodata. Section 5 presents the results and discussions. Section 6 presents the conclusions and recommendations for further research.

2 INSPIRE AND TRANSFORMATION APPROACHES

This section describes the INSPIRE initiative and its transformation approaches. According to the Directive 2007/2/EC (EC, 2007), the European Community is determined to create an "INfrastructure for SPatial InfoRmation in Europe" (INSPIRE) to provide a legislative framework that will enhance the accessibility of environmentally relevant thematic data for EC-politicians, economists, scientists, and citizens. In fact INSPIRE plays a key role for geodata harmonisation in

Europe. INSPIRE foresees the establishment of more detailed provisions, known as Implementing Rules (IR). Implementing Rules need to be established for:

- Metadata
- Network services
- Data specifications
- Data and service sharing
- Monitoring and reporting.

The INSPIRE Directive addresses 34 spatial data themes needed for environmental applications. These themes are subdivided in the three annexes of the directive: Annex I, II and III (EC, 2007). In order to comply with INSPIRE geodata specifications, national geodata providers have to transform their local data models to the models provided by the INSPIRE data specifications through the so called schema transformation. Through this approach geodata in local schemas can be transformed to geodata in INSPIRE schemas. This includes coordinate transformation from local Coordinate Reference Systems (CRS) to the European Terrestrial Reference System 1989 (ETRS89) mandated by the INSPIRE View and Download services required by the INSPIRE Directive². According to the INSPIRE Network Services Drafting Team three general alternatives exist to execute the necessary transformations:

- Offline transformation of the spatial data and provision of the transformed data in the View and Download service.
- On-the-fly transformation of the spatial data during request processing in the View and Download service.
- Transformation through a separate Transformation Service as defined in Draft Implementing Rules for INSPIRE Transformation (INSPIRE, 2009).

The scope of this research was to explore the first two transformation alternatives. The third alternative, an optional separate Transformation Service, was not considered at this time since the focus of this study was to explore standardised View and Download services.

3 ESDIN AND APPROACHES FOR TRANSFORMATION

This section provides background on the ESDIN project. In addition, the INSPIRE transformation approaches that were further developed within ESDIN are presented.

² Services required by the Directive (discovery services, view services, download services, transformation services, services allowing spatial data services to be invoked) conforming to the INSPIRE network services implementing rules (INSPIRE, 2008).

3.1 ESDIN Background

To support the implementation of INSPIRE, the European Commission has funded several projects. One of these is ESDIN – European Spatial Data Infrastructure Network, an eContentPlus project managed by EuroGeographics. ESDIN aims at helping national mapping and cadastral agencies to comply with INSPIRE specifications. ESDIN bridges the gap between INSPIRE specifications and practical implementation of those specifications.

3.2 ESDIN Transformation Approaches

The research presented in this paper was initially carried out in the context of ESDIN WP11 and in particular deliverable D11.1, which addresses Best Practices on content transformation in the context of INSPIRE network services (Lehto et al., 2009).

According to (Lehto et al., 2009) three major methods can be utilised to organize content transformation within a web-based data delivery workflow:

- Off-line Transformation: the transformation is realised through an entirely off-line pre-processing approach using a copy of a data set from the original database through an intermediate database. Through this method database storage would be challenging as one would need to store blobs of GML. While this may be doable, performing spatial queries and supporting any WFS filters would be challenging and greatly impacting performance if done in а linear fashion. Still. if these storage/query/indexing issues could be resolved this could be a viable direction, deserving future research.
- On-the-fly Transformation: the transformation is executed through an online process during the interactive request-response communication through web services. In fact through this approach theoretically this would be the most ideal form of transformation, as no data copies are required, working directly from source data. For simple data themes as for example Cadastral Parcels this could be workable. In practice however, several issues arise. The first is performance: on each WFS request at least two transformations need to be performed in both directions: coordinate transformation and model transformation. In more complex cases where source data may be spread over multiple tables this could have considerable performance implications. The second is complexity: a WFS can receive potentially any kind of query (spatial, non-spatial). These queries need to be mapped to the source data model, which is often very different. This will render very complex mappings.
- Combined Transformation: the transformation is applied using a combination of the two earlier introduced methods. Firstly, data from the original database is transformed and stored in an intermediate database.

Subsequently data in the intermediate database is transformed during the execution of Web services. Most NMCA's (e.g. the Dutch Kadaster) have separate datasets mapping to INSPIRE themes. At the same time most INSPIRE-themes define cross-theme relationships, for example an Address may refer to a Cadastral Parcel. Such relationships may not exist within the local source datasets, making this kind of integration challenging. In Combined Transformation the database model has these cross-theme relationships built-in, such that for example on transformation at least spatial relationships can be derived, like in the above example an Address could be linked to a Cadastral Parcel.

The first two above-mentioned approaches were already specified by the INSPIRE Network Services Drafting Team and were elaborated further within ESDIN.

4 COMBINED TRANSFORMATION APPLIED TO DUTCH GEODATA

This section describes the Combined Transformation approach that was applied to Dutch geodata. Technically the only NMAC's-specific step when using Combined Transformation is step 1 of the transformation from the local model to the intermediate database. During this research Combined Transformation (CT) was applied to two Dutch national data sets, namely: addresses and cadastral parcel data. The general scheme for these transformations is shown in Figure 1.

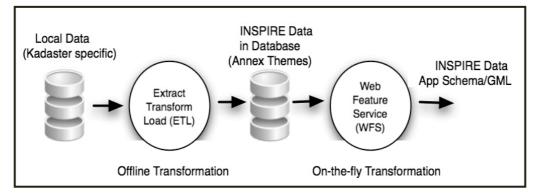


Figure 1: Combined Transformation Applied to Dutch Geodata

In Figure 1, the two sequential processing steps of Combined Transformation (offline and on-the-fly transformation) are depicted as circles. The arrows denote the flow of data from local data sources towards the end-client. The scenario implied by Figure 1 is as follows: the starting points are the Dutch national data

sets shown as "Local Data" on the left. These data sets reside in a relational database management system (RDBMS) using Oracle Spatial. The first step, offline transformation, can be regarded as ETL (Extract, Transform, Load) processing and has three sub steps:

- 1. Extract: data is exported from Oracle Spatial in a local (GML) format
- 2. Transform: exported data is transformed into a target data (INSPIRE) model
- 3. Load: the transformed data is loaded into a database

The intermediate database ("INSPIRE Data") plays a key role in the entire process. This spatial database applies a data model that reflects solely INSPIRE Annex data themes, in this case Addresses (AD) and Cadastral Parcels (CP). This spatial database is realized using a custom relational/spatial datamodel that corresponds to INSPIRE data themes and their cross-theme relationships. Basically this realizes an object-relational mapping where the "objects" are the XML object definitions within the INSPIRE data theme application schemas. To facilitate the object-relational mapping, the relational model within the database is defined to be as closely as possible to the INSPIRE data theme definitions. In addition, data is stored in ETRS89, since coordinate transformation is part of the ETL-processing.

The second, on-the-fly, transformation is performed by a Web Feature Service. For each WFS-request, data from the intermediate database is read and transformed into GML-data compliant with the respective Application Schema for each data theme. The steps described above may make this approach to Combined Transformation generic and possibly applicable to other national data sets, since only the first offline (ETL) step is specific to national data models.

4.1 Implementation of Combined Transformation

For the technical implementation of Combined Transformation described in the previous section, a software architecture was established and implemented. A high-level view of this architecture is depicted in Figure 2.

Figure 2 illustrates the Combined Transformation process and the various FOSStools and components used within the implementation of the overall architecture. The key FOSS tools are PostGIS, a spatial database based on PostgreSQL and Deegree, as a WFS/WMS framework. In a future research also GeoServer will be applied to publish geodata as WMS/WFS. The chain of Combined Transformation is described in further detail in Figure 3.

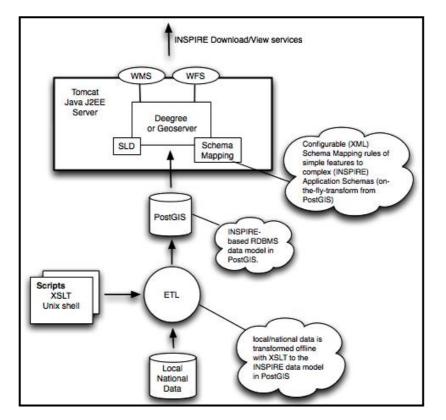


Figure 2: Software Architecture

Figure 3 shows details from all steps of the implementation of Combined Transformation. The entire process-chain starts at the top of the figure. The arrows denote the flow of data; the circles depict processing steps and the rectangles various files for intermediate storage or scripting.

Firstly, data is exported (extracted) from local databases ("Cadastre Native Data") into GML according to the local Dutch geodata models, such as IMBAG (Information model for key registers addresses and buildings) for Addresses. Through this approach the following commands were executed:

- Export: tools like FME and GDAL/OGR were used to export data from the local production Oracle Spatial database into flat feature GML or Dutch data models such as **Dutch national addresses and buildings (BAG)**.
- XSLT: this step was executed using Unix/DOS command line tools that call an XSLT processor. the input was XML (GML) but the output was not XML but rather lines of database INSERT statements according to the

relational model that reflects INSPIRE object models of the intermediate database

 PSQL: this step was executed using the standard PostgreSQL commandline tool called "psql".

All above commands are tied together within a single (UNIX) script that can be executed as a form of batch processing.

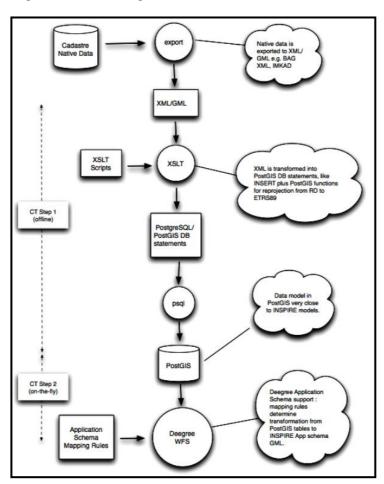


Figure 3: Processing chain for Combined Transformation

Figure 4 shows a fragment of a IMBAG GML address element, Subsequently, Extensible Stylesheet Language Transformations (XSLT) scripts are executed. Normally, XSLT is used to transform a source XML tree into a target XML tree, but in this case the XSLT scripts were used to transform the local data from GML into SQL (INSERT/UPDATE) statements for PostgreSQL/PostGIS. The

generated SQL statements comply with the data model of the intermediate "INSPIRE Database" discussed above. The PostGIS spatial function ST_Transform is generated as part of the XSLT-processing to perform coordinate transformation to ETRS89 during INSERT. The PostgreSQL tool (psql) can simply be invoked to perform the actual database inserts. All these steps were invoked using a pipeline of UNIX shell scripts³.

Figure 4: Example GML Fragment of a Dutch (IMBAG) address

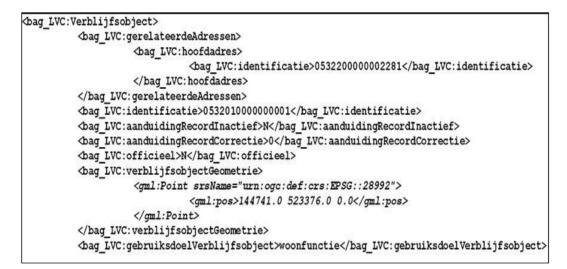


Figure 5 shows an example of an SQL statement generated through XSLT from the local source GML for Dutch addresses (IMBAG).

Figure 5: Example SQL Statement generated with XSLT

```
INSERT INTO ad_address (id,valid_from,position)
SELECT '053201000000001','2004-10-19 10:23:54+02',
ST_Transform(ST_GeomFromEWKT('SRID=28992:POINT(144741 523376)'), 4258)
WHERE NOT EXISTS (SELECT id FROM ad_address WHERE id='053201000000001');
```

The second, on-the-fly, step for transforming data in the "INSPIRE Database" to GML was performed by a Deegree WFS. This WFS was configured in order to map data from a PostGIS database to Application Schema GML. These mappings were specified through annotations within each GML Application

³ A UNIX shell is a program that makes the facilities of the operating system available to interactive users

Schema. This configuration maps the database schema in PostGIS to INSPIRE Annex I data themes for Addresses (AD) and Cadastral Parcels (CP).

4.2 Model Transformation for Addresses

This section exposes some details for the model transformation of Dutch addresses (IMBAG) to its respective INSPIRE Annex I theme for Addresses abbreviated as "AD". The AD specification (INSPIRE Addresses, 2010) describes a generic, object-oriented address model in both UML and as a GML Application Schema. Taking the mandatory elements from this model, a model for a RDBMS was created. This is depicted in the Entity/Relationship (E/R) diagram in Figure 6.

In the simplest approach "non-mandatory" were handled as follows: voidable GML object properties are predefined within the mapping template, i.e. they are not explicitly mapped. Properties that have 0...N occurrences are simply left out. In a more advanced approach properties were defined within the intermediate database. When a property has a null value within a database record, the mapping have either generated a "Nill" (void) element or left the element out for properties that had 0...N occurrences. In general all mappings are 1:1 on both the table and attribute level. The data model had been established in order to have the simplest mapping possible.

In Figure 6, each table name is prefixed with the corresponding abbreviation of the INSPIRE Annex I data theme, thus "ad_" denotes Addresses (AD) and "gn_" denotes Geographical Names (GN) as this Annex I data theme is used within the Addresses theme. Entities and relations depicted with dashed lines are part of the model but not used in the Dutch addressing scheme. For simplicity, most optional elements and attributes had been omitted.

The table "ad_address" is the key table, containing the geographic location ("position") of a single address and is the parent for one or more locators, such as house numbers within the table "ad_locator". In addition, a single row from "ad_address" is the parent of one or more address components from the table "ad_component". An address component can be an instance of an administrative unit (ad_admin_unit), a thoroughfare (ad_thoroughfare), an address area (ad_address_area) or postal code/postal name (ad_postal).

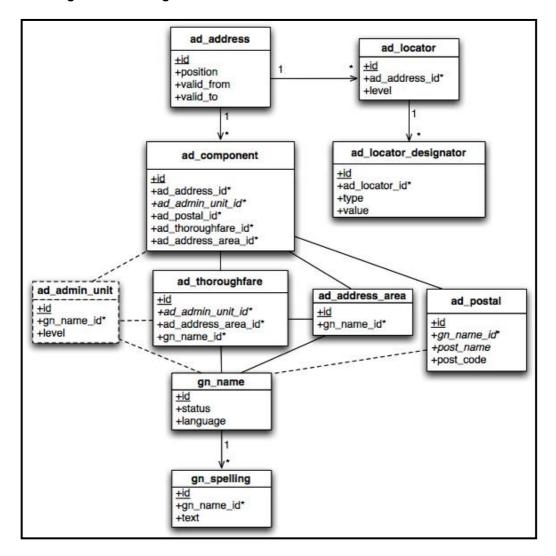


Figure 6: E/R Diagram derived from INSPIRE Annex I Addresses theme

The Dutch national data model for addresses (IMBAG) is object-based and uses a different hierarchy. A simplified E/R model is depicted in Figure 7.

Although the Dutch data model for addresses is very different from the INSPIRE Addresses data theme, a mapping could be established between these two models since most elements/attributes have similar counterparts within the INSPIRE model. Without the intention to be complete, some mapping examples are provided in Table 1.

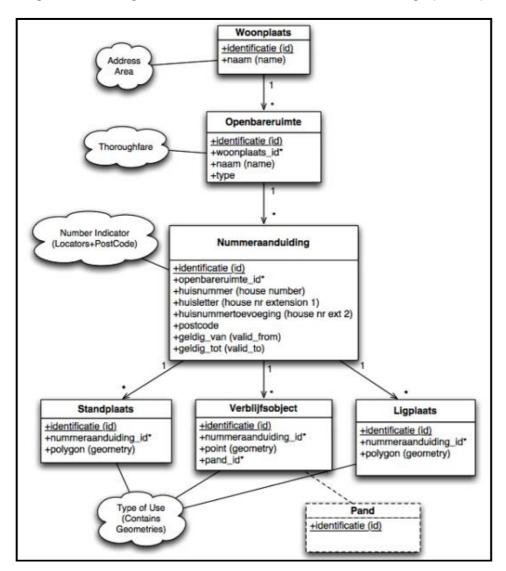


Figure 7: E/R Diagram - Dutch national addresses and buildings (IMBAG)

The two columns on the left in Table 1 denote names of the tables/attributes for the Dutch addresses (IMBAG). The two columns on the right denote the INSPIRE Addresses (AD) theme tables and attributes to which these are mapped (using XSLT). In most cases these mappings were easily encoded with XSLT. Some PostGIS spatial functions were generated to deal with coordinate transformation (using ST_Transform) and for example mapping a polygon to a point (using the PostGIS ST Centroid function).

The detailed exercise for cadastral parcels was not mentioned in this article because Addresses (AD) have a more complex model and would illustrate the

advantage of the approach better, and to limit the length of the paper with "duplicate" information.

Dutch Address Table	Attribute	INSPIRE AD Table	Attribute
Nummeraanduiding	identificatie	ad_address	id
Nummeraanduiding	huisnummer huisletter huisnummertoevoegi ng	ad_locator ad_locator_designator	type value
Nummeraanduiding	postcode	ad_postal	post_code
Ligplaats	polygon	ad_address	position
Standplaats	polygon	ad_address	position
Verblijfsobject	point	ad_address	position
Openbareruimte	identificatie	ad_thoroughfare gn_name gn_spelling	id
	naam		text
Woonplaats	identificatie	ad_address_area gn_name	id
		gn_spelling	text

Table 1: Mapping BAG to INSPIRE

5. RESULTS AND DISCUSSION

This paper presented a methodology through which Dutch national geodata conformant to INSPIRE specifications were harmonised and published within the ESDIN project. The methodology was tested by applying it to test data sets. The test data consisted of about 200 000 Dutch addresses and about 130 000 cadastral parcels which were transformed in the tests. The data resulting from this transformation was stored in PostGIS and made available as GML for download using a WFS based on Deegree. The GML (version 3.2.1) data from this WFS was successfully validated against the original INSPIRE Annex I (version 3) GML application schemas. For viewing data, a WMS was configured. A browser-based viewer was developed using GeoExt, a powerful FOSS JavaScript framework based on both OpenLayers and the ExtJS framework. A screenshot showing INSPIRE addresses on OpenStreetMap (OSM) is depicted in Figure 8.

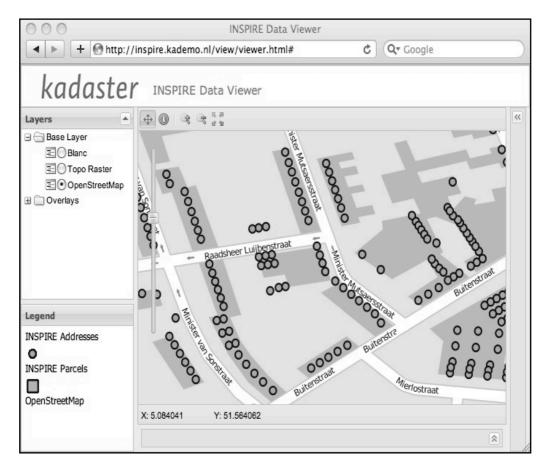


Figure 8: The INSPIRE data viewer showing Addresses

5.1 Combined Transformation Trade-offs

From the analysis of this experience, we can see both advantages and disadvantages. The advantages are that, in general, transforming local data to INSPIRE data themes can be a complex problem when performed as a single transformation. Combined Transformation applies a well established software design rule called "divide and conquer" by subdividing the transformation problem into two manageable sub-problems.

Since only the (offline) first step of Combined Transformation is specific to a national data model, National Mapping Agencies may re-use the same setup. In addition, by having a single database model to which the locally specific transformation is performed, cross-border projects can be realised more easily.

Rather than treating each data theme and local dataset as a separate entity or project, a common integrated INSPIRE-theme database model can give rise to several integrations, and can be re-used for common themes. For example the Annex I theme Geographical Names (GN) is used in Addresses (AD) but may also be used in Administrative Units and other Annex I, II, III themes.

This approach also allows to establish cross-theme relationships, such as for example Cadastral Parcels and Addresses. In some cases NMCA's may use the data model to integrate otherwise locally separate data sets. Since the data within the database is close to the final INSPIRE model and already in the target coordinate system (ETRS89) performance is optimised as opposed to on-the-fly transformations and re-projections.

There were also some disadvantages of the approach of Combined Transformation. Most importantly, the offline transformation step may render the intermediate database out of sync with the source data when the latter changes. For example, cadastral parcels may be split; new addresses may be added etc. To overcome this issue, several approaches are possible, such as setting a trigger on changes within the source or a regular refresh of the intermediate database from the source. This depends greatly on the how up-to-date the data needs to be and/or the amount of processing needed for the offline transformation. Overall, the advantages outweigh the disadvantages, at least in the context of this study and the data themes that were used.

6. CONCLUSIONS AND RECOMMENDATIONS FOR FUTHER RESEARCH

Geodata is becoming increasingly used in many application areas in wider contexts. While cross-border cooperation among the EU's member states is growing, the need to access geodata is also increasing. Almost every state member has its local geodata standards and specifications that make data sharing a challenging issue when it comes to data harmonisation. INSPIRE has facilitated this matter through defining pan-European specifications. Besides, ESDIN attempts to help all European NMCA's practising with cross-border data harmonisation and transformation conformant to INSPIRE specifications.

Content transformation was the key issue addressed by ESDIN WP 11. In this context and during this research Combined Transformation (CT) was examined. CT is a viable approach for the complex problem of transforming national data sets to INSPIRE data themes. The CT approach could also facilitate joint cross-border projects, using a single database where each partner applies its specific offline transformation to publish to this database. However, being dependent on an extra database can be seen as the main disadvantage of this approach.

Since this approach has been limited to two INSPIRE Data Themes (AD and CP) and Dutch national data sets, it would be interesting to see how this approach performs with other INSPIRE Data themes and other local datasets within other NMCA's. Further, the enhancements within the existing intermediate database model, making it more complete with for example more optional attributes, and possibly structural changes to comply with local data models can be foreseen.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Mr. Frank Arntsen from the Norwegian Mapping Authority (NMA) who made the first setup for a WFS for the INSPIRE Addresses theme with PostGIS and Deegree and making this setup available.

REFERENCES

- European Commission (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) <u>http://eurlex.europa.eu/JOHtml.do?uri=OJ:L:2007:108:SOM:EN:HTML</u>
- Garnter, G. and F. Ortag (2009). Cartography in Central and Eastern Europe; Chapter 10: Cross-Border Mapping: *Geodata, Geonames, Multilinguality and More*, page: 163, Springer Berlin Heidelberg, Haberstrasse 7, D-69126 Heidelberg, Germany. Craglia, M. and A. Annoni (2003).
- INSPIRE (2008). Network Services Architecture, at: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/network/D3_5_I NSPIRE_NS_Architecture_v3-0.pdf, [accessed 15 January 2010].
- INSPIRE (2008). Drafting Team "Data Specifications, definition of Annex Themes and Scope, at: http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecificatio ns/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pdf, [accessed 29 January 2010].
- INSPIRE (2009). Draft Implementing Rules for Transformation Services, at http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE_Dr aft_Implementing_Rules_Transformation_Services_(version_3.0).pdf, [accessed 29 January 2010].

INSPIRE Status Report (2009). at:http://inspire.jrc.ec.europa.eu/documents/INSPIRE_/INSPIRE_Status% 20_Report_OCT_09.pdf, [accessed 15 December 2010].

- INSPIRE (2010). INSPIRE Data Specifications on Addresses *Guidelines v* 3.0.1,at:http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INS PIRE_DataSpecification_AD_v3.0.1.pdf, [accessed 15 January 2010].
- Jönsson. S. (2008). A National Geodata Strategy, at: http://www.gsdi.org/gsdiconf/gsdi11/papers/pdf/202.pdf, [accessed 29 January 2010].
- Lehto L. et al. (2009). Best Practice for Content Transformations Enabling INSPIRE-Compliant Data Delivery, *ESDIN Deliverable D11.1, ECP-2007-GEO-317008*, October 9, 2009.
- Witschas S. (2005). CROSS-BORDER MAPPING EXPERIENCES WITH GEODATA AND GEONAMES, Leibniz Institute of Ecological and Regional Development (IOER), Weberplatz 1, 01217 Dresden, Germany,at:http://www2.ioer.de/recherche/pdf/2005_witschas_icc.pdf,[a ccessed 30 January 2010].