

Harmonising Dutch National Geodata Conform To INSPIRE Using Combined Transformation*

Ebrahim Hemmatnia¹, Just van den Broecke², Ko van Raamsdonk³

^{1,3}Cadastre, Land Registry and Mapping Agency, Apeldoorn, The Netherlands

²Just Objects, Amstelveen, The Netherlands

Abstract

This paper addresses schema and coordinate transformations together with INSPIRE-compliant data delivery 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 transformation. 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

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

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 transboundary cooperation projects and the need for transboundary geospatial analyses in cross-border regions (Garnter and Ortag, 2009).

However, large scale cross-border geodata were 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 actualised 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. This is not possible without harmonisation of geodata across Europe.

“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 initiative.

This paper describes recently gained knowledge and skills for the realisation of harmonising Dutch national geodata conform 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 (version 3.0) were examined: a) Addresses (AD) and b) Cadastral Parcels (CP), using Free and Open Source Software (FOSS) 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..

With the Directive 2007/2/EC (INSPIRE, 2010), 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 (Asch, 2009). In fact INSPIRE plays a key role for geodata harmonisation in Europe.

The INSPIRE Directive covers five areas: metadata (creation and publishing), data and services (interoperability), network services specifications, monitoring and reporting, and data sharing (Geoconnexion.com, 2009). Each of these areas should be implemented for all INSPIRE thematic geodata. INSPIRE addresses 34 thematic (geo) data fields and metadata (Annex I, II, and III of the Directive); the data are to be prepared according to the INSPIRE Implementing Rules (IR) in a consistent and comparable way (Asch, 2009).

In order to comply with INSPIRE data specifications, national geodata providers have to transform their local data models to INSPIRE data specifications. This includes coordinate transformation from local Coordinate Reference Systems (CRS) to the European Terrestrial Reference System 1989 (ETRS89) mandated by INSPIRE.

According to the INSPIRE Network Services Drafting Team (2009) 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 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.

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 run by EuroGeographics. ESDIN aims at helping national mapping & 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:

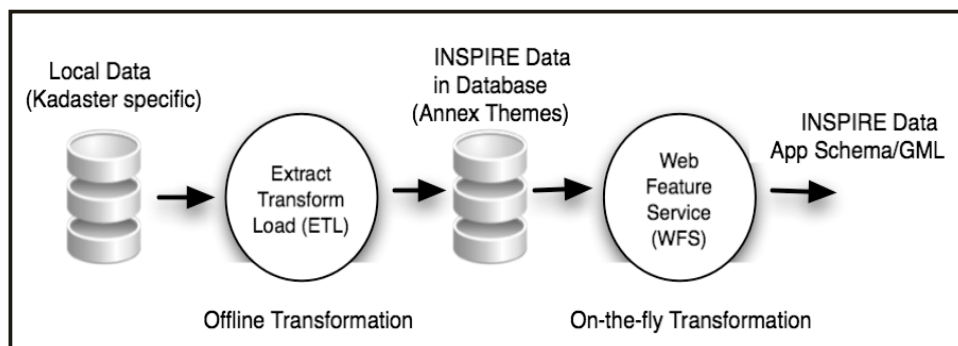
- **Off-line Transformation:** the transformation is realised through an entirely off-line pre-processing approach using a copy of a data set from the original database through a service database.
- **On-the-fly Transformation:** the transformation is arranged through an online process during the interactive request-response communication through web services.
- **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 a service database. Subsequently data in the service database is transformed during the execution of Web services.

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. 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). 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 above described steps 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: Software Architecture

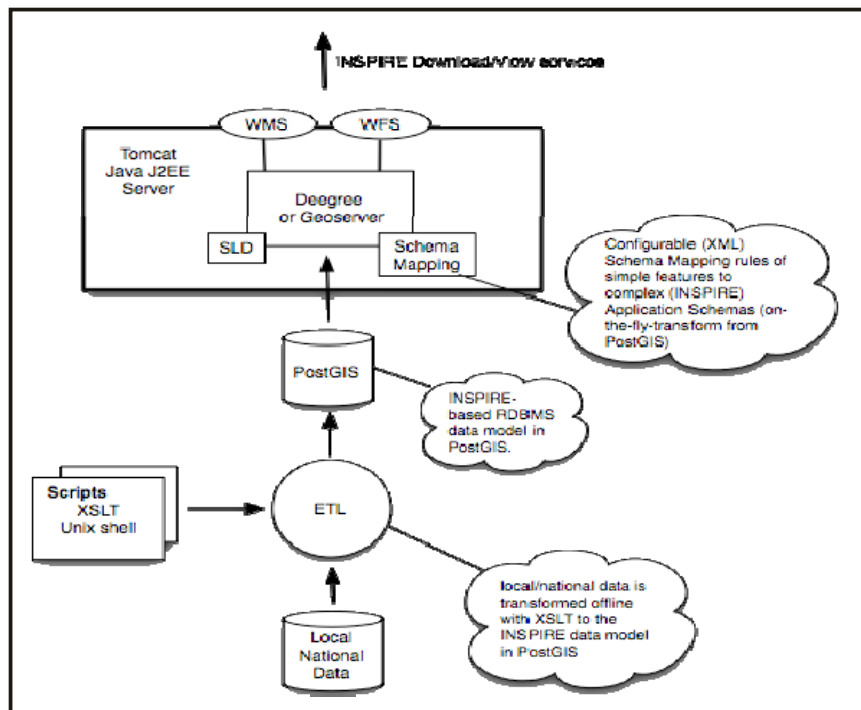


Figure 2 illustrates the Combined Transformation process and the various FOSS-tools and components used within the implementation of the overall architecture. The key FOSS tools are PostGIS, a spatial database based on PostgreSQL and Deegree, as a WFS/WMS framework. In a future research also GeoServer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

Figure 3: Processing chain for Combined Transformation

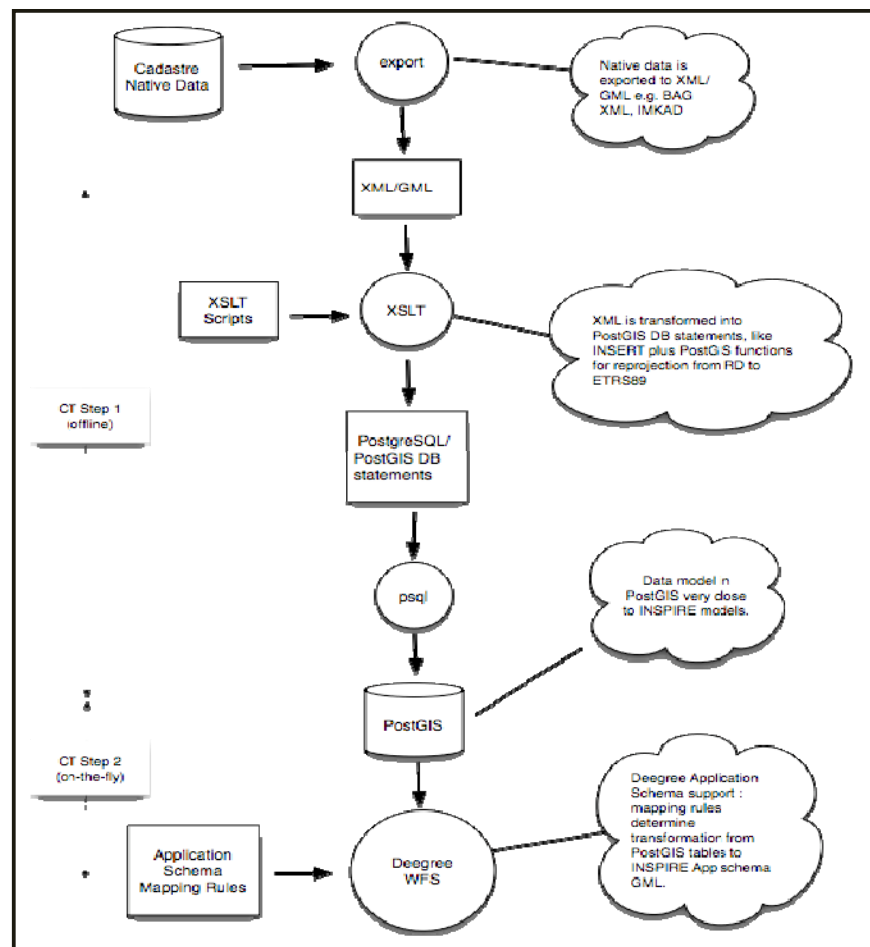


Figure 3 shows details from all steps of the implementation of Combined Transformation. The entire process-chain starts at the top of the figure. The arrows denote the flow of data; the circles depict processing steps and the rectangles various files for intermediate storage or scripting.

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.

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 scripts¹.

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

```
<bag_LVC:Verblijfsobject>
  <bag_LVC:gerelateerdeAdressen>
    <bag_LVC:hoofdadres>
      <bag_LVC:identificatie>0532200000002281</bag_LVC:identificatie>
    </bag_LVC:hoofdadres>
  </bag_LVC:gerelateerdeAdressen>
  <bag_LVC:identificatie>0532010000000001</bag_LVC:identificatie>
  <bag_LVC:aanduidingRecordInactief>N</bag_LVC:aanduidingRecordInactief>
  <bag_LVC:aanduidingRecordCorrectie>0</bag_LVC:aanduidingRecordCorrectie>
  <bag_LVC:officieel>N</bag_LVC:officieel>
  <bag_LVC:verblijfsobjectGeometrie>
    <gml:Point srsName="urn:ogc:def:crs:EPSG::28992">
      <gml:pos>144741.0 523376.0 0.0</gml:pos>
    </gml:Point>
  </bag_LVC:verblijfsobjectGeometrie>
  <bag_LVC:gebruiksdoelVerblijfsobject>woonfunctie</bag_LVC:gebruiksdoelVerblijfsobject>
```

¹ A UNIX shell is a program that makes the facilities of the operating system available to interactive users (cs.princeton, 2010).

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

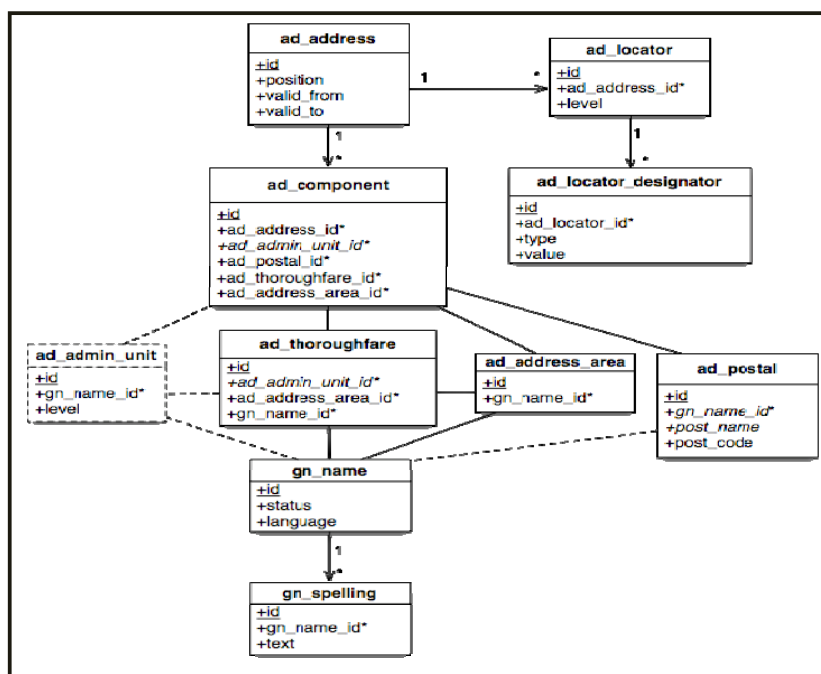
```
INSERT INTO ad_address (id,valid from,position)
SELECT '0532010000000001', '2004-10-19 10:23:54+02',
ST_Transform(ST_GeomFromEWKT('SRID=28992;POINT(144741 523376)'), 4258)
WHERE NOT EXISTS (SELECT id FROM ad_address WHERE id='0532010000000001');
```

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 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.

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

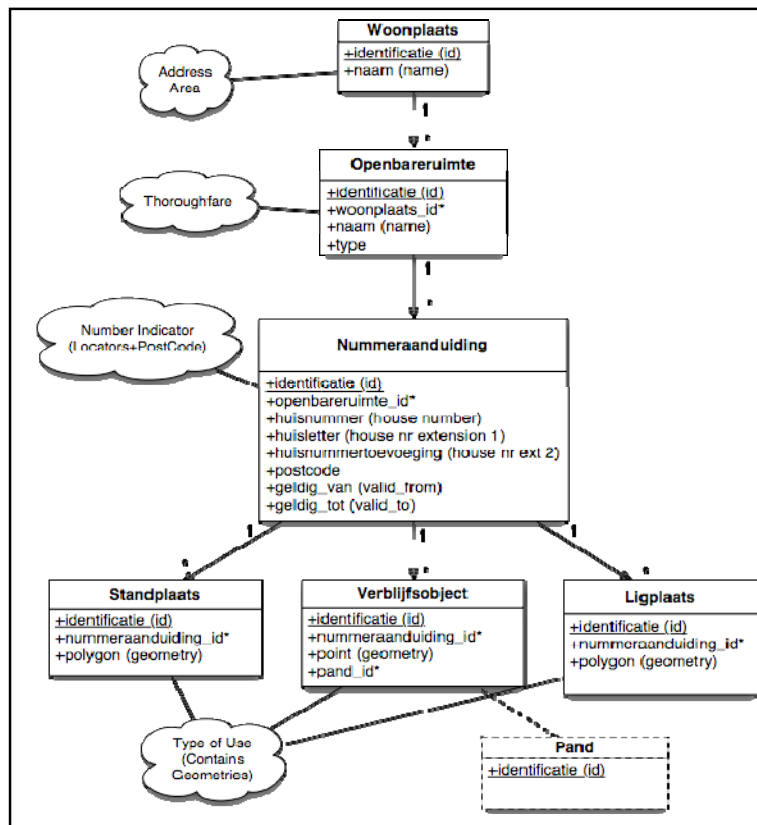


In Figure 6, each table name is prefixed with the corresponding abbreviation of the INSPIRE Annex I data theme, thus “ad_” denotes Addresses (AD) and “gn_” denotes Geographical Names (GN) as this Annex I data theme is used within the Addresses theme. Entities and relations depicted with dashed lines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

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.

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



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 below.

Table 1: Mapping Dutch national addresses and buildings (BAG) to INSPIRE

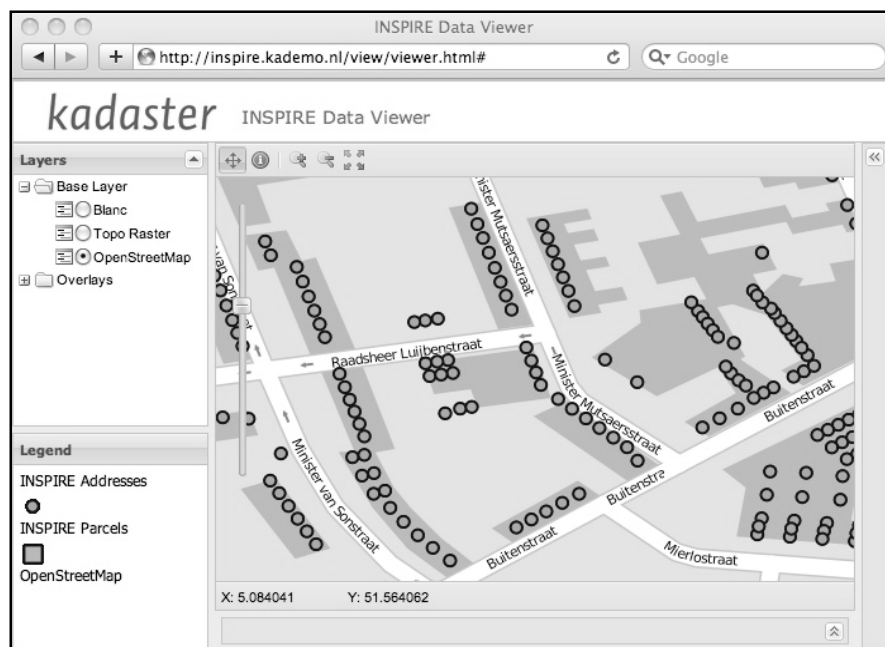
Dutch Address Table	Attribute	INSPIRE AD Table	Attribute
Nummeraanduiding	identificatie	ad_address	id
Nummeraanduiding	huisnummer huisletter huisnummertoevoeging	ad_locator ad_locator_designator	type value
Nummeraanduiding	postcode	ad_postal	post_code
Ligplaats	polygon	ad_address	position
Standplaats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

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).

5. RESULTS AND DISCUSSION

This paper presented a methodology through which Dutch national geodata conform 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 Tradeoffs

On analysis of this examination one can discern advantages and disadvantages. Firstly the advantages will be named.

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.

As opposed to 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. For example, the database can provide re-use 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 for establishing 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 mutate. 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 mutations within the source or a regular refresh of the intermediate database from the source. This depends greatly on the required actuality of the data and/or the amount of processing needed for the offline transform. Overall, the advantages outweigh the disadvantages, at least in the context of this examination 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 own 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 conform to INSPIRE specifications.

Content transformation was the key issue addressed by ESDIN WP 11. In this context and during this research Combined Transformation was examined.

Combined Transformation 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 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

- Asch K. (2009). International viewpoint and news; New EC Directive to "INSPIRE" the European Geoscience Community, at: <http://www.springerlink.com/content/51w47r3v27563248/fulltext.pdf?page=1>, [accessed 30 January 2010].
- CS.Princeton.(2010).at:<http://www.cs.princeton.edu/courses/archive/spr02/cs217/asgts/ish/ish.htm>, [accessed 5 June 2010].
- 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).

- Geoconnexion (2009). Update on INSPIRE Directive Implementation, at http://www.geoconnexion.com/geo_online_article/Update-on-INSPIRE-Directive-Implementation/343, [accessed 30 January 2010].
- INSPIRE (2009). Draft Implementing Rules for Transformation Services, at [http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE_Draft_Implementing_Rules_Transformation_Services_\(version_3.0\).pdf](http://inspire.jrc.ec.europa.eu/documents/Network_Services/INSPIRE_Draft_Implementing_Rules_Transformation_Services_(version_3.0).pdf), [accessed 29 January 2010].
- INSPIRE (2010). INSPIRE Data Specifications on Addresses - Guidelines v 3.0.1, at: http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_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, [accessed 30 January 2010].