GeoTest: A Testing Environment for Swedish Geodata

Anders Östman[1], Imad Abugessaisa[2], Solgerd Tanzilli[3], Xin He[4], Mohamed El-Mekawy[5]

[1]University of Gävle, SE-80176 Gävle, Sweden. Anders.Ostman @hig.se
[2]Future Position X, SE-801 33 Gävle, Sweden. Imad@fpx.se
[3]National Land Survey of Sweden, SE-801 82, Gävle, Sweden. Solgerd.Tanzilli@lm.se
[4]University of Gävle, SE-80176 Gävle, Sweden. Xin.He@hig.se
[5]Future Position X, SE-801 33 Gävle, Sweden. Mohamed.El-Mekawy@fpx.se

Abstract

GeoTest is a project initiated by Future Position X (a GIS cluster organization in Gävle, Sweden), the National Land Survey of Sweden (NLS) and the University of Gävle. The project aims to test Swedish geodata and make sure they comply with the INSPIRE specifications in Annex I–III. The purpose of this paper is to present the development of and experiences from GeoTest in developing the required infrastructure for testing Swedish geodata.

Due to the high demand for geodata in modern society, and anticipated requirements by INSPIRE, the issue of testing a national SDI is very important. An adequate testing environment is required. In this context, two aspects are discussed:

1. Technical requirements for GeoTest.
2. Organizational and coordination aspects.

The technical requirements for GeoTest are to provide an environment to test current specifications in the context of INSPIRE Annex I. The methodology is crucial to govern the testing process. The developed methodology mainly relates to the transformation testing of the themes specified in Annex I–III. The objective of the tests is to ensure that the specifications are balanced in terms of costs and that they contribute to address user needs. At this stage, GeoTest focuses on transformation testing. The aim is to test that the transformations from local schemas to INSPIRE schemas are technically feasible.

The methodology is based on four stages; the main strategy used in this testing was the ETL (extract-transform-load) approach. The testing process started with a preliminary desk study, with the objective to collect basic information about the themes and identify the availability and sources of schemas. The extraction process helped to identify the costly procedures when generating GML data that conform to the source schemas; in this stage we encountered some problems in the extraction process, as the data are loosely coupled to the source schema. The transformation procedures of the GML data from the extraction process to the INSPIRE GML schema performed in three sub-stages, mainly related to schema mapping, matching and transformation.

The testing process was coordinated with NLS, via one contact person per theme. The extraction for the sample data was based on random selection of 5% of the sheet index. The paper presents the results of testing five themes addresses, cadastral parcels, geographical names, hydrography, and transport networks. The result shows that in total, 13 INSPIRE schemas has been investigated. Of these, data may be delivered for 10 schemas.

Keywords: INSPIRE, Transformation Testing, ETL, Schema Mapping, GML, FME, Triple Helix model

1. INTRODUCTION

INSPIRE is a European directive aiming to establish an infrastructure for geospatial information in the European Community (European Union, 2007). It entered into force on May 15th 2007 and according to the implementation plan, the directive shall be implemented into national legislations within two years. The European Commission will during the coming years publish specifications for the implementation of the INSPIRE directive. This paper mainly relates to the data specifications for the themes specified in Annex I. It is a request that the member states through their legally mandated organizations (LMO) and the user communities through their spatial data interest communities (SDIC), test these specifications. The objective of these tests is to assure that the specifications are balanced in terms of cost and that they contribute to solving user needs. The draft data specification for the Annex I themes was published in December 2008. The testing was finished in late February 2009.

The data specification testing consists of two parts, one transformation testing and one application testing. The aim of the transformation testing is to study how to transform data in local schemas to data in INSPIRE schemas. When INSPIRE is being implemented, this type of transformations is to be provided by the data provider. The aim of the application testing is to study if the specifications adhere to user needs and to test the integration of several themes. These tests are to be driven by use cases. Since the time plan is so compressed and since it requires quite a long calendar time to establish use cases, the application testing is for the time being not considered in this paper.

According to the INSPIRE directive, it has to be implemented in the national laws by May 2009. In December 2008, a Swedish Environmental Information Act and Regulation was drafted. In the regulation, responsible agencies for each INSPIRE data theme are proposed. But due to uncertainties about the detailed content of the corresponding implementation rules, a shared responsibility is often proposed. For many data themes in Annex I of the INSPIRE directive, the National Land Survey of Sweden (NLS) is proposed to have such a shared responsibility. As a consequence, this paper focuses on the transformation of schemas and data provided by the NLS. In particular, geographical names, cadastral parcels, hydrography, transportation networks and addresses are considered.

A data transformation requires here that a number of software tools are used in order to transform data in local schemas to the INSPIRE schemas. Software testing is here defined as the process of executing software and comparing the observed behavior to the desired behavior (Watson and McCabe, 1996). As a consequence, we can then consider the data transformation tests as a kind of software testing.

Software testing methodologies are traditionally of two different types, namely black-box testing where no knowledge about the design of the software modules is known to the tester and white-box testing, where internal data structures and algorithms are known. White-box testing are mainly used as an integrated part of software development projects. Since commercial off-the shelf (COTS) software is to be used in this study, the black-box testing methods are more suitable.

There is a variety of black-box testing methods (Kaner and Bach, 2006) such as domain testing, scenario testing, function testing etc. Scenario testing is based on meaningful combinations of functions and variables, in contrast to domain tests and combinatorial tests where more artificial combinations are used. Since the purpose here is to estimate costs and identify real problems, scenario testing is most appropriate.

The objectives of this paper are to describe the method for the transformation tests that has been developed and to identify limitations in NLS ability to provide data according to INSPIRE draft implementation rules.

2. METHODOLOGY

The aim of the transformation testing is to study how to transform data in local schemas to INSPIRE schemas. The test results shall also be used for input to cost considerations.

The following items are to be tested

I. Data content and structure by generic comparison of schemas
II. Delivery (can local data be mapped to target)
III. Reference systems (coordinate system transformation might be required)
IV. Data quality and metadata (are metadata available)

The testing is here proposed to be made in 4 stages, each one explained below. The main strategy in this testing is to use the ETL (extract-transform-load) approach. However, when the INSPIRE services are to be implemented (in 2 - 7 years time), other approaches might be used.

2.1. Preliminary desk study

The objective of the desk study is to gather some basic information about the data theme being tested. The desk study shall among other things include the following actions.

Identification of source schema
The goal here is to identify the schemas that have been used for the source data sets. Even non-formal descriptions of schemas are of great value. In conjunction with this work, the question whether the data follows the schema shall also be investigated, for instance through interviews.

Preliminary schema matching
Schema matching is the process of identifying corresponding concepts in the source schema (national data sets) and the target schema (INSPIRE specification). The proposed procedure is here to first, for each INSPIRE feature type identify the corresponding feature type(s) in the national data sets. Then secondly, for each attribute in the INSPIRE feature types, identify the corresponding attribute(s) in the national data sets.

Metadata survey
The goal here is to investigate if the metadata as specified in the INSPIRE specifications are present. The goal is also to specify the spatial reference systems being used. In case data is expected to be inhomogeneous, this should also be specified here (Hillmann *et al* 2004).

2.2. Generation of Source GML data

The objective of the extraction process is to identify costly procedures when generating GML (Geography Markup Language) data that conforms to the source schema. This problem is expected to increase in cases when the source schema is not specified or when data are loosely coupled to the source schema.

Specify source schema in XML and constraints

In cases when the source schema is specified, it has to be converted to XML. In cases where such a schema is lacking, a schema has to be created. In addition to the schemas, some geometric constraints might also need to be specified, such as edge matching constraints, topological constraints etc, if not specified in the schema.

Extract sample data sets

The goal here is to extract sample data sets from the entire source database. The sample data sets shall be representative for the quantification of problems that may occur. The extraction should be based on random sampling.

Convert sample data sets to GML/XML

In case the extraction above is not based on XML/GML, a transformation from the export format to GML/XML is required.

Quantification of data inconsistencies

When the source data is loaded in a XML parser, the consistency with the schema can be studied. For the time being, procedures for automatic correction have not yet been developed. Instead, one of the two following approaches can be applied. The first is based on data reduction, or simply deleting the features that do not fit the schema. The second option is to gradually relax the constraints in the schema. It is here proposed that both methods are used and reported.

2.3. Transformation to INSPIRE GML schema

The objective of the transformation test is to estimate the occurrence of costly procedures in a future schema translation process.

Schema matching and mapping

The goal here is to specify the transformation rules when transforming from source schema to INSPIRE schema. Based on the preliminary desk study, a final matching and mapping is performed. The transformation rules are then used for a XSLT (Extensible Stylesheet Language Transformations) based transformation.

Schema transformation

The objective of this task is not to do the actual transformation, only to estimate the occurrence of costly procedure. It is expected that the transformation tools are implemented as WPS (Web Processing Service) or manual processes.

3. Results and analysis

This part of the paper will illustrate the results obtained by the steps discussed in the methodology part. Each section represents one stage in the testing methodology. It will focus on the problematic issues relating to transformation and quantify them.

3.1. The preliminary desk study

3.1.1. Identification of NLS application schemas

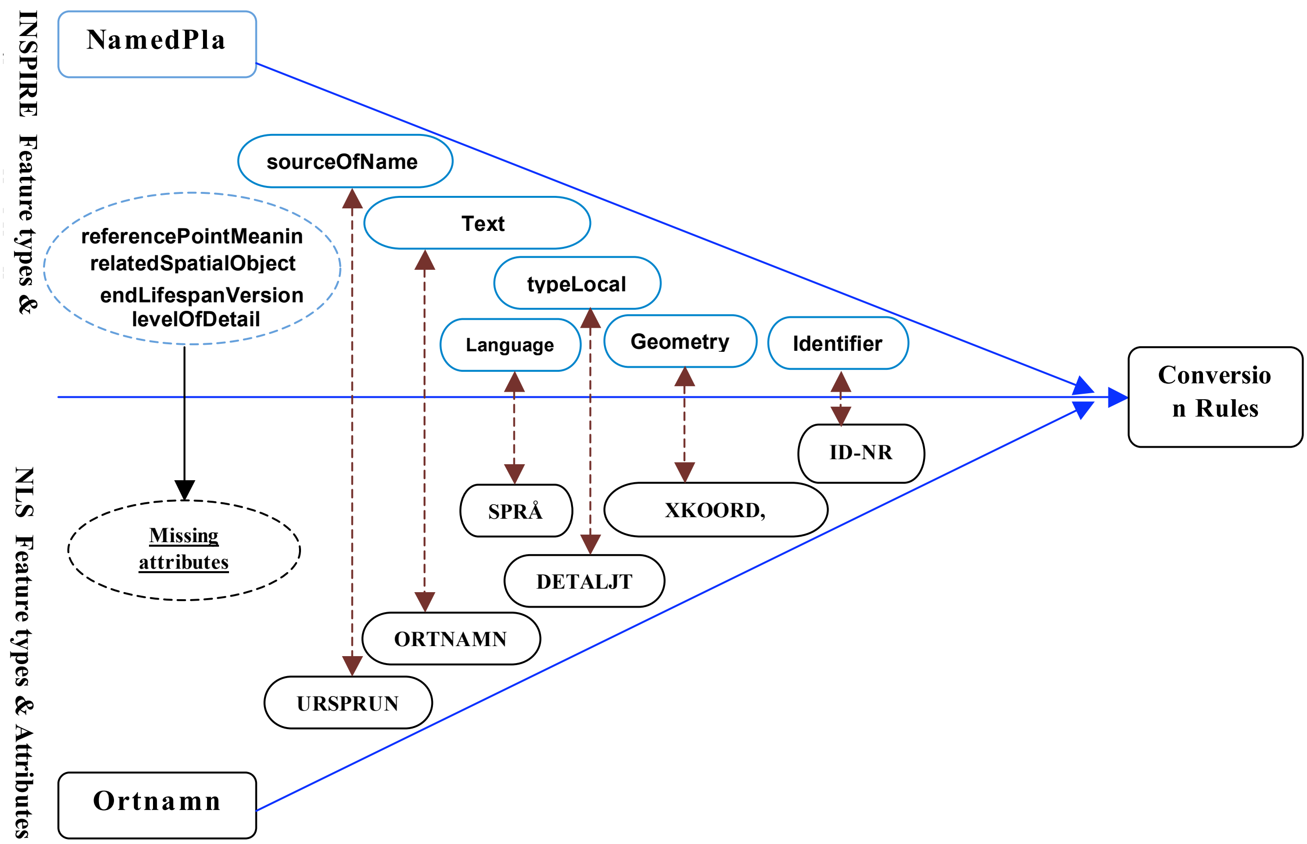
For each INSPIRE schema, the corresponding data sources and schemas at NLS were identified. In some cases, UML class diagrams were available, while in other cases Entity-Relationship diagrams were used. The result is summarized in table 1.

Table 1: Availability of models and schemas from NLS.

Theme	NLS	
Address (AD)	UML in PDF	Yes
	Source schema	No
Hydrography (HY)	UML in PDF	Yes
	Source schema	No
Geographical Names (GN)	UML in PDF	Yes
	Source schema	Yes
Transport Networks (TN)	UML	No
	ER diagram	Yes
	Source schema	Yes
Cadastral Parcels (CP)	UML	No
	ER diagram	Yes
	Source schema	Yes

3.1.2. Schema matching and mapping

Figure 1: An illustration of schema matching process, corresponding elements are joined by dot arrows.



The main objective of the schema matching is to identify corresponding concepts in the source schema (Rahm *et al*, 2001), which are here national data sets, and the target schema, which is here the INSPIRE specifications. The matching process considered both the Swedish-to-English language issues as well as the semantic differences in both schemas for this process (Bohannon *et al*, 2006 and Madhavan *et al* 2001). An example of schema matching is show in figure 1. As shown here, most elements of the feature type "NamedPlace“ can be matched directly to a corresponding element in the NLS schema.

The preliminary schema matching and mapping was executed manually as a desk study. In those cases where correspondence was found, the required data transformation procedures were identified. Some transformations were considered to be problematic if they required substantial manual work (McCann *et al* 2005 and Milo *et al* 1998).

One such example is that the INSPIRE specification differentiates between a slip road and a motorway. However, the NLS datasets do not make such differentiation. For the time being, there are no automatic procedures for extracting slip roads from the motorway features.. The following section presents an example of the way the schema for geographical names has been matched.

The resulting transformation rules specify the source data type (NLS), the target data type (INSPIRE), the conversion rules and its meaning (table 2).

Table 2: Transformation rules for Geographical Names

Source data type	Target data type	Conversion Type	Meaning
Code list	Character String	CodelistToText	Codelist value converted to character string
number*2	GML object	CoordinateToPoint	Coordinate pair converted to GML Point
Text or missing value	Text	Assign(Value)	Target value in brackets used instead of source text
Integer	Character String	IntegerToText	Integer value converted to character string
Char	Text	Equal	No conversion required
CodeList	CodeList	Assign	Target value used instead of source value

Table 3 summarizes the result of schema mapping and matching. In total, 13 INSPIRE schemas has been investigated. Of these, data may be delivered for 10 schemas.

In table 3 we can for instance observe that for the road transport network schema, only 33% of the INSPIRE defined feature types had a correspondence in the source database. For these feature types 67% of the mandatory elements can be transformed and 23% of the optional elements

Table 3: Results from desk study Number of feature types that can be provided and elements that may be transformed from the Swedish data sets to the INSPIRE data specifications

INSPIRE Schema	Spatial Feature Types			Mandatory Elements			Optional Elements		
	INSPIRE	*Provided*	*Perce-ntage*	*INSPIRE*	*Transformable*	*Perce-ntage*	*INSPIRE*	*Transf-ormable*	*Perce-ntage*
GN	1	1	100%	3	2	67%	16	8	50%
AD	6	6	100%	7	6	86%	91	25	27%
CP	3	3	100%	11	9	82%	30	14	47%
Road TN	6	2	33%	3	2	67%	53	12	23%
Railway TN	8	2	25%	4	4	100%	27	5	19%
Water TN	5	2	40%	4	4	100%	13	3	23%
Phy. Waters	10	7	70%	12	5	42%	55	12	22%
HY. Facilities	4	1	25%	1	1	100%	5	2	40%
HY. POI	4	2	50%	3	2	67%	28	14	50%
HY Man Made Objects	8	4	50%	4	4	100%	24	7	29%

3.1.3. **Metadata survey**

A metadata survey was also conducted at NLS from December 2008 through February 2009 as part of the Annex I INSPIRE testing process. The survey aimed to find the availability and characteristics of metadata according to the INSPIRE metadata implementation rules that the commission adopted on 3 December 2008[1]. It checked the availability of the metadata elements for the five themes of addresses, cadastral parcels, geographical names, hydrography, and transportation networks the.

The thematic groups at NLS answered and provided information on 90 metadata elements, both mandatory and optional. The survey's response rate was 100%. The thematic groups also provided answers to questions concerning

a) if the metadata being collected is based on any standard
b) the languages used for documenting the metadata
c) the metadata's temporal coverage
d) the discovery mechanism (search, browse and other features)
e) the metadata's accessibility (online URL, intranet or offline)
f) whether any application schema may be used for metadata transformation.

The survey revealed that in regard to the standards followed, only the geographical names' metadata followed the basic geographic database's (Grundläggande Geografiska Databas, or GGD) rules, which comprise an internal system that NLS uses. The rest of the themes follow no standard in documenting their metadata.

The survey also revealed that extracted data sets for the schema transformation tests were missing the metadata when delivered for testing. The stored but not delivered metadata were of an operational nature and a limited number of elements were available. No metadata were available for response to customers' requests.

With respect to discovery mechanisms, the survey revealed that metadata could be discovered by searching certain map sheets and then browsing information related to their temporal extent and resolution via the GeoLex system[2.] The survey also found that the

[1] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:326:0012:01:EN:HTML
[2] http://www.geolex.lm.se/

delivery format is unspecified and that metadata elements were unavailable for transformation to INSPIRE target schemas.

Table 4 summarize the availability of the mandatory metadata elements, as specified by the INSPIRE implementation rules for each theme.

Table 4: Availability of mandatory metadata elements

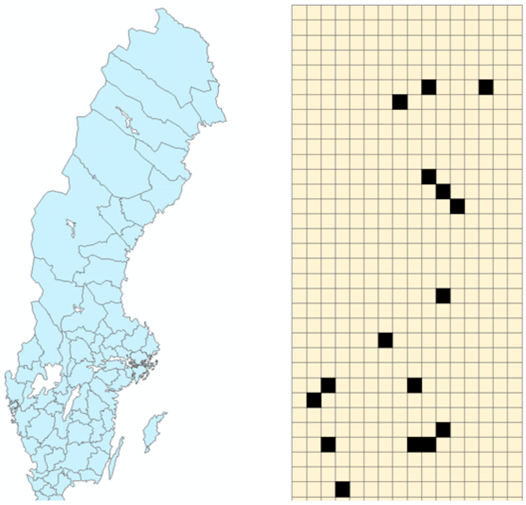
Theme	A available elements	Not available	Lack of Information	Not Sufficient Information
AD	28,0%	64,0%	0,0%	7,1%
CP	75,0%	21,4%	3,5%	0,0%
HY	35,7%	64,2%	0,0%	0,0%
GN	53,7%	14,2%	3,5%	28,5%
TN	75,0%	25,0%	0,0%	0,0%

3.2. Generation of Source GML data

3.2.1. **Extracting sample datasets**

The datasets used for testing was based on different scales for different themes. As an example, the maps for geographical names are scaled at 1:10,000 For testing the transformations on real data, a random sample of 5% of the 318 map sheets was used see figure 2. The sampling was based on an existing map sheet index. The datasets were delivered in ESRI shapefile format.

Figure 2: Boundary map of Sweden (left) and the map index with selected sheets in black (right)



3.2.2. **Specification of source schemas and constraints**

The main purpose of this process was to specify a source XML schema for each theme and to perform generic comparisons with the target schemas. Generic comparisons of UML models were based on

a) identifying classes and their corresponding spatial feature types in the target schemas

b) listing all attributes in the UMLs that describe a class's properties and listing each's corresponding attributes, which are simple elements, in the target schema
c) identifying the types of relationships and integrity constraints within the UMLs in order to identify foreign keys and relationships and the corresponding constraints or restrictions in the target schema.

The datasets for geographic names, hydrography, and addresses lack proper source schemas, while the datasets for cadastral parcels and transport networks have source schemas expressed as entity relationship (ER) diagrams. Such differences in the availability of source schemas is due to both cadastral-parcel registry's and road networks' databases being implemented in different platforms with different capabilities. For the other three themes, the source schemas ware derived from the extracted datasets.

The results of the generic comparisons show that

a) in the case of geographic names, addresses, and hydrography, the source schemas did not conform to the UML diagrams provided by NLS. These diagrams missed for instance many feature types such as complex elements
b) the database tables for geographic names, addresses, and hydrography lack such integrity constraints (ICs) as foreign keys and relationships. This because data have been delivered as ESRI shapefile, where each feature is being represented in a flat file (single table)
c) both hydrography and transport networks' themes fully conformed to their ER diagrams and all integrity constraints were present.

3.2.3. **Converting sample datasets to GML/XML**

According to the selected method, the datasets should be transformed to GML format after extraction. A spatial ETL was here used for this conversion process. In particular, the Feature Manipulation Engine (FME) server from Safe Software and the ArcGIS 9.3 Interoperability extension was used to build the workbench. This process read and transformed all input features into GML, representing each shapefile in GML as a more concrete generic feature collection. The result of this process was that all features in the shape files were transformed to GML format.

The second measure was a data-type consistency check. The resulting GML files were loaded in an XML parser (Liquid Studio) in order to study the consistency of the schema for each file. Also here it was found that the converted data conformed to the GML schema.

3.3. Transformation to INSPIRE GML schema

The final stage of testing methodology is to transform the source data in GML into the target schema, a process called transformation. This transformation included several steps for each theme with the objective to estimate the procedures' costs..

The problems that occurred during the transformation of real data, were quantified by specified several countermeasures and used as an evaluation matrix. These included the identification of target elements that were un-translated. Here only those elements for which corresponding elements in the target schema could be identified were included. Those elements where source data were missing were excluded

.

Another countermeasure was the identification of problematic translations. Elements that were mapped during the desk study were supposed to be transformed to the target schema. Some of those un-translated elements was considered to be problematic transformations and identified. A problematic transformation is here defined as transformations that require manual processing in order to be accurate.

The problematic transformations mainly consist of data representation problems, such as incompatible classification systems and geometry types. One example is from the road transportation network. Here, in the source schema the width of the road is represented by a domain of widths (4-7 meter, 7-10 meters etc). The corresponding element in the target schema is a RoadTN:width type (gml:measure type), which uses the reference system as unit of measure (UoM)(Lake *et al* 2004). This problem applies to all roads in the database. Another problem relates to the cadastral parcels, where the target schema specifies that the geometry should be represented as a simple polygon. However, for around 8 % of the Swedish parcels, the geometry is represented as a line or point. In addition, several parcels are also represented as complex polygons. A generic geometry element is therefore proposed as a proper geometric representation of cadastral parcels.

4. Organizational and coordination aspects

Future Position X (FPX) is a non-profit association aiming to build a closer cooperation in research, development, internationalization and marketing in the GI field. The cluster was formed in 2004 by public and privat actors and most of FPX members are working in the Gävle region with Geodata and GIS. FPX is the owner of the GeoTest project. NLS and the University of Gävle are stakeholders.

The testing environment is installed at FPX, located in the science park in Gävle, Sweden. At that location the testing team has been coordinating and operating the testing procedures. The idea of locating the project to this address was to benefit from FPX[3] location and GIS cluster organization as well as from cooperation with the many different actors hosted at the science park.

Figure 3: Triple Helix model (Etzkowitz, 2008)

A testing process based on the Triple Helix Model (Etzkowitz, 2008) has been established (see figure 3). Here active participation and cooperation between owners and interest parties from university, municipality, authority, industry and other organizations and networks promoting Swedish spatial data infrastructure, SDI has been cooperating. The application of the Triple Helix model has so far, according to INSPIRE testing, been fruitful.

The process has been influenced and driven forward by different interest parties and networks connected to the INSPIRE testing task. Daily management routines coupled to a

[3] FPX GIS cluster program is described at www.fpx.se

Macronet time plan with defined milestones outlined assured the testing of the five themes in a timely manner.

The testing team is composed of three specialists from the university, one specialist from FPX and one project leader from NLS. To help identify the source datasets, models and schemas an expert team was established at NLS. The team consisted of one expert for each theme and one for metadata. These experts helped defining all datasets and executing the orders for all extractions of datasets.

In Sweden two other LMO's and SDIC's has been testing for INSPIRE. Cooperation with these two has been of a special interest according to the ongoing INSPIRE coordination and strategy work conducted from the geodata secretariat[4] at NLS. The Geodata secretariat at NLS is coordinating the strategy work for implementing the INSPIRE directive in Sweden. GeoTest is part of the Geodata strategy.

Swedish Standards Institute, SIS is an important organization and interest party for the GeoTest testing. SIS is conducting several member forums where standard issues are being discussed. GeoTest is taking part as a member in a forum where Geodata services are discussed.

5. DISCUSSION

This paper has contributed to the testing process of INSPIRE specifications. Its main contribution was the testing methodology that has been developed and verified. The methodology offered the needed technical infrastructure required for testing spatial data.

The result from the transformation test shows that the black-box testing procedures are reliable and suitable in transformation testing. The scenario testing helped to identify countermeasures and cost related parameters and pinpoint the real problems.

The main question was to what extent NLS's data can be mapped to the INSPIRE schema. Thirteen INSPIRE schemas have been investigated during schema matching and mapping, and when doing the real transformation it was found that data for ten schemas might be delivered according to the implementation rules. Also the result from metadata survey answered to the question of the availability and characteristics of the dataset-level metadata as specified by INSPIRE.

It was also found that for each theme, feature types specified by INSPIRE is either a) covered by the NLS database (transformable), b) covered by the NLS database (but it need special transformation rules (problematic transformation), and c) not covered by the NLS database.

During transformation from Shape format to GML, the file size has been increased and the data types have been converted to text format to conform to XML encoding. However, consistency checking has shown that the actual values of the data kept without change.

Current XML editors are limited in their capabilities to handle normal size GML file and perform simple operation on such file. More efforts in the future should be dedicated to address these problems.

Quite a lot of organizations and parties have been involved in INSPIRE Directive, and they are not limited to LMO and SDIC, but also to the research and business communities.

[4] www.geodata.se and www.lantmateriet.se

As GeoTest organization based on Triple Helix model, the model makes it possible to all groups to participate and cooperate between each other. This model make it easy not only to share the resources but also to share knowledge and experience towards INSPIRE implementation.

6. CONCLUSIONS

The paper demonstrates an environment for testing Swedish Geodata. The methodology is based on an ETL approach. The organization and coordination with external parties is based on Triple Helix model. The results obtained so far showing that the method used was very successful and can be used for testing other type of themes in different levels.

A Swedish draft Environment Information Regulation has recently been published. Here, the responsibilities of the governmental agencies for providing data and services according to the INSPIRE directive are specified. This study shows that NLS may be a suitable data provider for the themes Geographical Names, Cadastral Parcels and Addresses. For the other two themes being studied, other agencies should probably act as responsible data provider.

It is also shown that NLS may provide some information related to the Transportation Networks and Hydrography themes. Although not shown here, it can be assumed that the information at NLS may be of value for the agencies being assigned as responsible data provider for these themes and schemas. A stronger cooperation among the agencies therefore seems to be necessary.

A number of costly transformation procedures have been identified. To what extent these costs justifies a change in the INSPIRE implementation rules remains to be seen.

The current metadata at the NLS do not conform to the current INSPIRE implementation rules. One reason for this is the quite frequent changes in the metadata standards that have occurred during recent years. This problem is also related to the lack of maintenance routines for metadata. However, one conclusion is that NLS should pay more attention to the metadata questions, in order to have them fully implemented by 2010 as the INSPIRE directive require.

Future work will focus on developing methods for the testing of application and services that are required by INSPIRE directive in the near future.

7. ACKNOWLEDGEMENT

This paper has been prepared as a part of the Future Position X project NYSTA 39686, financially supported by the European Union structural funds, objective 2. Additional support has been provided by the National Land Survey of Sweden and the University of Gävle.

REFERENCES

Bohannon, P. Elnahrawy, E. Fan, W. and Flaster M. (2006),"Putting Context into Schema Matching" *Proceedings VLDB '06, September 12-15 Seoul, Korea* pp 307-3118.

European Union, (2007) Directive 2007/2/EC of the European Parliament and of the Council. At: http://www.ec-gis.org/inspire/directive/l_10820070425en00010014.pdf, [accessed 17.09.2008].

E. Rahm and P. A. Bernstein. (2001) *A survey of approaches to automatic schema matching, The VLDB Journal*, 10(4): 334-350.

Etzkowitz, H. (2008) The Triple Helix: University-Industry-Government Innovation In Action 2008, London, Routledge. Hillmann, D. and Elaine L.(2004), Metadata in Practice, American Library Association
Published by ALA Editions.

Kaner C, Bach J, (2006) Black Box Software Testing. At http://www.testingeducation.org/BBST/ [accessed 16,02,2009].

Lake, A. Burggraf, D. Trninic,M. and Rae, L (2004) Geography Mark-up Language (GML), Published by John Wiley and Sons.

Madhavan, J. Bernstein, P. A. and Rahm, E. (2001) ''Generic schema matching with Cupid'' *Proceedings of the 27th International Conference on Very Large Data Bases San Francisco, CA*, pp 49-58.

McCann,R. AlShebli, B, Le, Q. Nguyen, H. Vu, L. and Doan. A. (2005) ''Mapping maintenance for data integration systems''. *Proceedings of the 31st VLDB Conference, Trondheim, pp1118-1129* Norway.

Milo, T. Zohar, S. (1998)''Using schema matching to simplify heterogeneous data translation'' *Proceedings of the 24th VLDB Conference New York, USA.*

WATSON, A. H., McCabe, T. J. (1996) Structured Testing: A Testing Methodology Using the Cyclomatic Complexity Metric. Computer Systems Laboratory, National Institute of Standards and Technology Special Publication 500-235, Gaithersburg, MD 20899-0001.

Wick, M. Rohanimanesh, K. Schultz,K. McCallum, A.,(2008). ''A Unified Approach for Schema Matching, Coreference and Canonicalization'', *Proceedings KDD'08, August 24–27, 2008, Las Vegas, Nevada, USA.*