GIMA

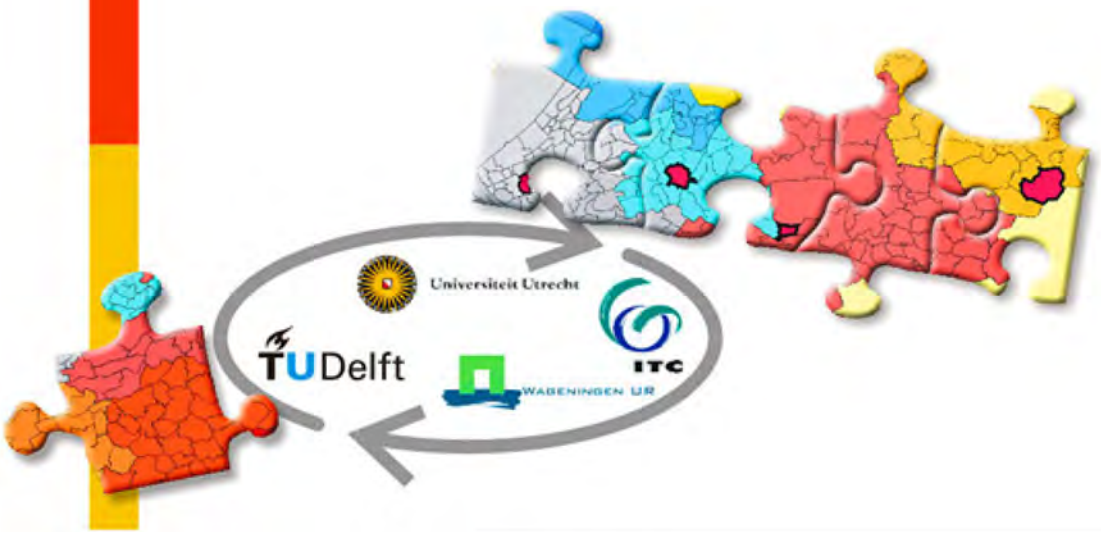
Geographical Information Management and Applications

INSPIRE Compliant Datasets
Transformation & Conformance Testing

MSc Thesis – July 2012

Author: Nart Tamash
E-mail: nart.tamash@gmail.com

Supervisor: Drs. Marian de Vries (TUD)
Professor: Prof. Dr. Ir. Peter van Oosterom (TUD)
Reviewer: Dr. Ir. Rob Lemmens (ITC)



Geographical Information Management & Applications (GIMA)

Master of Science Thesis – July 2012

INSPIRE Compliant Datasets – Transformation & Conformance Testing

by Nart Tamash
nart.tamash@gmail.com

Supervisor:
Drs. Marian de Vries (Delft University of Technology)

Professor:
Prof. Dr. Ir. Peter van Oosterom (Delft University of Technology)

Reviewer:
Dr. Ir. Rob Lemmens (Faculty of Geo-Information Science and Earth Observation of the University of Twente – ITC)

© 2012
Nart Tamash
All Rights Reserved

Acknowledgements

This thesis is the result of a very intense and steep-curved study period, but that has proved very inspiring in the end. My interest for the subject was quite high even before knowing much about it, and now, with the completion of this thesis, I can say that this interest has increased even more and I hope that I will have another chance in the future to be involved in INSPIRE related work.

Therefore, I would first and foremost like to thank Marian de Vries, who was my thesis supervisor, for her support and feedback during the entire period I have been working on this thesis. Also, I am very grateful to Peter van Oosterom, the responsible professor, whose wide experience on the subject guided me in the right direction and provided valuable input.

I also owe my acknowledgements to Richard Rombouts and Debbie Wilson from Snowflake Software, by showing interest in my research topic and by answering the many questions that I had, and ultimately, by providing access to GO Publisher, one of the main software tools used in the project.

I would also like to send my warmest greetings to all my colleagues, who participated with me in the GIMA programme (full and part timers from 2010, and part timers from 2009), and also to all the GIMA staff for making this programme what it is. I am really honoured to have been part of it.

Last but not least, I want to thank my family for the faith they have in me, my friends for encouragement, and especially my girlfriend who has been very patient and supportive throughout the entire programme.

Abstract

The INSPIRE initiative sets up a framework for the creation of an European Spatial Data Infrastructure (ESDI), which will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access in general to spatial information across Europe. To do so, several common specifications have been developed in a wide range of areas including data, metadata, and network services. The most challenging aspect of INSPIRE will probably be harmonising the actual data models across Europe to the common INSPIRE ones, giving the amount of time allocated for this process. This task is not only challenging because of the amount of data that will be involved in the process, but also because of the very varied source data models and amount of data providers that will be involved at various stages and having to cooperate at European level under a unique framework.

Therefore, the main question that arises, and that is on the mind of many data providers across Europe, is how is that really achieved and what does it involve? This thesis aims to clarify that aspect by focusing on data transformation and conformance testing. The research follows a stepped approach, first of all by putting into the INSPIRE context, concepts like interoperability and data harmonisation, extending to the importance of geographic information standards in this sense, as well as the ultimate goal of a spatial data infrastructure.

A case study is considered, where at first, source data from the UK mapping agency, Ordnance Survey, that falls into the scope of one of the INSPIRE thematic themes, namely Administrative Units, is analysed and compared to the target data model proposed by INSPIRE, trying to identify similarities and differences that may pose problems. The process continues with identifying software tools that are capable to perform data transformation based on INSPIRE requirements, and eventually using one of them to transform the data. After the transformation, encountered bottlenecks are discussed, both from the source data side, but also from the target data model side.

The last step is to formally test the produced datasets as required by the standards that INSPIRE rely upon, by means of an Abstract Test Suite (ATS) and Executable Test Suite (ETS). This is maybe one of the crucial aspects of INSPIRE data harmonisation process, as there is still some ambiguity between legally binding and not legally binding requirements, an aspect that will directly influence testing and its interpretation, hence the obligations of each data provider.

The thesis will conclude with observations that are relevant not only for the Administrative Units theme, but also for the wider scope of INSPIRE data transformation and conformance testing. Main bottlenecks are discussed, but also recommendations are given that would definitely be relevant for further research, as well as for the INSPIRE community.

Keywords: INSPIRE, data harmonisation, transformation, interoperability, standards, spatial data infrastructure, conformance, compliance, abstract test suite, executable test suite

Table of Contents

1. INTRODUCTION ..... - 1 -

1.1 PROBLEM STATEMENT ..... - 1 -
1.2 RESEARCH OBJECTIVES ..... - 3 -
1.2.1 Research limitations ..... - 3 -
1.3 METHODOLOGY ..... - 4 -
1.4 SUMMARY OF CHAPTERS ..... - 6 -

2. LITERATURE REVIEW ..... - 7 -

2.1 INTEROPERABILITY & DATA HARMONISATION ..... - 7 -
2.1.1 Interoperability concepts ..... - 7 -
2.1.2 Data harmonisation ..... - 10 -
2.2 STANDARDS IN GEO-INFORMATION ..... - 14 -
2.2.1 International Organisation for Standardisation ..... - 15 -
2.2.2 European Committee for Standardisation ..... - 15 -
2.2.3 Open Geospatial Consortium ..... - 16 -
2.3 SPATIAL DATA INFRASTRUCTURES ..... - 16 -
2.4 INSPIRE DIRECTIVE ..... - 19 -
2.4.1 Data specifications ..... - 24 -
2.4.2 Conformance testing ..... - 27 -
2.5 CONCLUSIONS ..... - 29 -

3. CASE STUDY: INSPIRE ADMINISTRATIVE UNITS – GAP ANALYSIS ..... - 31 -

3.1 SOURCE DATA ..... - 31 -
3.2 ADMINISTRATIVE UNITS DATA SPECIFICATIONS ..... - 36 -
3.3 CONCLUSIONS ..... - 42 -

4. DATA TRANSFORMATION ..... - 45 -

4.1 TOOLS OVERVIEW ..... - 45 -
4.1.1 Safe Software FME ..... - 45 -
4.1.2 Snowflake Software GO Publisher / GO Loader ..... - 46 -
4.1.3 Esri ArcGIS for INSPIRE ..... - 47 -
4.1.4 The HUMBOLDT Framework ..... - 48 -
4.1.5 Other FOSS solutions ..... - 49 -
4.2 TRANSFORMATION APPROACH ..... - 50 -
4.3 DISCUSSION ..... - 51 -

5. PRODUCING INSPIRE COMPLIANT DATA ..... - 53 -

5.1 DATA PREPARATION ..... - 53 -
5.2 DATA TRANSFORMATION ..... - 56 -
5.2.1 Administrative Units ..... - 59 -
5.2.2 Administrative Boundaries ..... - 71 -
5.2.3 Condominiums ..... - 74 -
5.2.4 NUTS Regions ..... - 74 -
5.3 EVALUATION AND CONCLUSIONS ..... - 75 -
5.3.1 GO Publisher ..... - 75 -
5.3.2 Transformation issues ..... - 76 -

6. CONFORMANCE TESTING ..... - 79 -

6.1 TESTING PROCEDURE ..... - 81 -
6.2 ATS/ETS ..... - 83 -
6.2.1 Application schema ..... - 83 -
6.2.2 Geometry/Topology ..... - 93 -
6.2.3 Reference system ..... - 96 -
6.2.4 Data quality ..... - 100 -
6.2.5 Delivery ..... - 101 -
6.3 EVALUATION AND CONCLUSIONS ..... - 102 -
6.3.1 Application schema ..... - 103 -
6.3.2 Geometry/Topology ..... - 103 -
6.3.3 Reference systems ..... - 104 -
6.3.4 Data quality ..... - 104 -
6.3.5 Delivery ..... - 104 -
6.3.6 Conformance and compliance ..... - 104 -

7. SUMMARY AND CONCLUSIONS ..... - 105 -

7.1 INSPIRE AND DATA INTEROPERABILITY ..... - 105 -
7.2 GAP ANALYSIS AND DATA TRANSFORMATION ..... - 106 -
7.3 CONFORMANCE TESTING ..... - 107 -
7.4 RESEARCH QUESTIONS ..... - 108 -
7.5 RECOMMENDATIONS ..... - 109 -

REFERENCES ..... - 111 -

ANNEX A – INSPIRE FEATURE CATALOGUE (ADMINISTRATIVE UNITS) ..... - 115 -

ANNEX B – ADMINISTRATIVE UNITS REQUIREMENTS / RECOMMENDATIONS ..... - 129 -

ANNEX C – INSPIRE GML SAMPLES ..... - 133 -

ANNEX D – ABSTRACT TEST SUITE FOR THE ADMINISTRATIVE UNITS THEME ..... - 137 -

Abbreviations

AU – Administrative Units
ATS – Abstract Test Suite
BL – Boundary-Line™
DT DS – Drafting Team Data Specifications
DT MD – Drafting Team Metadata
DT NS – Drafting Team Network Services
EU – European Union
EC – European Commission
ESDI – European Spatial Data Infrastructure
ETL – Extract-Transform-Load
ETS – Executable Test Suite
IDE – Integrated Development Environment
GCM – Generic Conceptual Model
GI – Geographic(al) Information
GIS – Geographic(al) Information System(s)
GML – Geography Markup Language
INSPIRE – INfrastructure for SPatial InfoRmation in Europe
ICS – Implementation Conformance Statement
IR – Implementing Rule(s)
IXIT – Implementation eXtra Information for Testing
ISO – International Organisation for Standardisation
JDK – Java Development Kit
LAU – Local Administrative Units
NSDI – National Spatial Data Infrastructure
NUTS – Nomenclature of Territorial Units for Statistical Units
OGC – Open Geospatial Consortium
ONS – Office for National Statistics
OS – Ordnance Survey
SDI – Spatial Data Infrastructure
SRID – Spatial System Reference Identifier
SUT – System Under Test
TG – Technical Guidelines
TWG – Thematic Working Group(s)
UML – Unified Modelling Language
URI – Unique Resource Identifier
XML – Extensible Markup Language

1. Introduction

This chapter will provide a general overview of the research setup. The main aspects that will be presented will relate to the problem background, the objectives and limitations of this research, and the methodology proposed to reach those objectives. The problem statement in Section 1.1 will provide a brief introduction describing the overall setting of the research and the problem that is going to be tackled. The research objectives in Section 1.2 will present a crisp formulation of the research goals, represented by a main research question and several sub-questions, but also what the research doesn't necessarily propose to achieve. Finally, the methodology in Section 1.3 will provide a brief overview of the necessary steps to take in order to reach the proposed objectives and answer the research questions, and Section 1.4 will summarise the content of the chapters.

1.1 Problem statement

The importance of spatial information in assisting the definition and monitoring policies of the European Union has grown rapidly in the last years. Application areas like agriculture, regional development, environmental management, transport and energy are the most important ones (Bernard *et al.*, 2005). Furthermore, governmental, commercial as well as research institutions are developing an increasing number of geo-applications in order to make it easier to carry out daily activities. At the European Union level there is a permanent increase in both the number of trans-boundary cooperation projects and the need for trans-boundary geographic analyses in cross-border regions. However, the success and efficiency of such cooperation projects heavily relies on the availability and usability of heterogeneous geo-data (Witschas, 2010). Therefore, in such cases, the full benefits, and eventually the success of these projects, will only be achieved if the shared geo-data is *"completely interoperable, usable and understandable by the global, interdisciplinary community"* (Hemmatnia *et al.*, 2010). This literally means standardised access to data that originates from various sources, and this can only be achieved through geo-data harmonisation. Unfortunately, in the European context, most, if not all EU member states have local standards which make sharing geo-data across borders tremendously difficult, sometimes impossible.

In response to the above mentioned issues, back in 2001 the INSPIRE (INfrastrcuture for SPatial InfoRmation in Europe) programme was initiated and in May 2007 the INSPIRE Directive came into force as a European legislation to which all EU member countries are bound. INSPIRE is supposed to be implemented in various stages with full implementation required by 2020. The Directive lays down a general framework for a European Spatial Data Infrastructure (ESDI). Bernard *et al.* (2005) state that in the context of INSPIRE, the term Spatial Data Infrastructure (SDI) follows the definition of the SDI Cookbook, meaning conceptually that it *"encompasses the policies, organisational remits, data, technologies, standard delivery mechanisms and financial and human resources necessary to ensure that those working with spatial data, whether at the global or the local scale, are not impeded in meeting their objectives"*. INSPIRE follows a subsidiary approach to build up the ESDI by setting specific requirements that national SDIs will be able to comply to. National SDIs in turn will be based on a number of regional and/or local SDIs that will comprise the same components (i.e. metadata, reference and thematic data, etc.). Therefore, INSPIRE will enable the sharing of

environmental spatial information among public sector organisations, assist policy makers, but will also give the wider public a better access to spatial information across Europe.

To ensure that the spatial data infrastructures of the Member States are compatible and usable in a Community and trans-boundary context, the Directive requires that common Implementing Rules (IR) are adopted in a number of specific areas across all data themes. These areas, or components, are: Metadata, Data Specifications, Network Services, Data and Service Sharing and Monitoring and Reporting (European Commission, 2007). This research will mainly be centred on the Data Specifications component with the objective of assessing necessary procedures to harmonise data according to the INSPIRE standards. Data Specifications can be considered the core element of the INSPIRE initiative since it provides the guidelines for the harmonisation of the various data models across European countries. As mentioned earlier, in this case, interoperability needs to take the various cross-community information needs into account. When this is added to the huge difference in the scope of the 34 different data themes proposed by INSPIRE, it becomes a real challenge to establish the specific requirements of, and for interoperability and harmonisation of the geographic information (Drafting Team 'Data Specifications', 2010a). The Data Specifications Drafting Team responsibility was to specifically overcome this challenge.

Fig. 1: INSPIRE logo (http://inspire.jrc.ec.europa.eu/)

In order to combine data from diversified sources into integrated, consistent, and unambiguous information products, various software tools have been developed to facilitate the transformation of a source data model to a target data model. However, these tools haven't been thoroughly and formally tested with the INSPIRE data specifications, which means they cannot be considered INSPIRE compliant products. What makes data harmonisation in the context of INSPIRE such a challenging process is the diversity in the source data models. 27 Member States means 27 different national data models, multiplied by the 34 data themes in INSPIRE, equals 918 different source data models that need to be transformed into the INSPIRE data models. The scale of this process is overwhelming. This research will look into the possibilities of transforming one of these data models into the INSPIRE data model by using existing software tools, and according to the published INSPIRE Data Specifications.

In addition, one of the requirements of the INSPIRE Data Specifications is the ability to test INSPIRE compliance. This is achieved by the means of an Abstract Test Suite (ATS) and its instantiation, an Executable Test Suite (ETS), which would test the harmonised datasets against the INSPIRE Implementing Rules and other technical recommendations. Although this process is based on an

existing international standard (i.e. ISO 19105 / Geographic Information – Conformance and testing), the ATS hasn't been developed yet for any of the INSPIRE data themes. Therefore, this research will also look to at developing an ATS for the data specification of the INSPIRE theme that will be used as case study.

1.2 Research objectives

The major objective of the thesis research will be to look at the different data harmonisation tools available and use them in the context of INSPIRE, in order to create compliant datasets. In the process of doing so, the limitation of these tools must be discussed in regard to the INSPIRE Data Specifications. A secondary objective, probably with an even greater interest and relevancy for the research process would also be to test the transformed datasets according to the standards in place and based on the INSPIRE requirements. This would stand in an Abstract Test Suite (ATS), developed for the specific data theme that will be picked as a case study, which will specify the necessary procedures to verify if all relevant requirements have been implemented and if a dataset can truly be considered INSPIRE conformant. Furthermore the ATS will lead to an Executable Test Suite (ETS) that will be able to perform the actual tests on the transformed dataset.

Therefore, the thesis should be able to answer the following central research question:

How can selected datasets be harmonised and declared INSPIRE compliant/conformant datasets according to the INSPIRE Data Specifications guidelines?

To answer this main research question the following sub-questions are posed:

- *What are the concepts of interoperability and data harmonisation and how do they fit in the context of INSPIRE?*
- *How does the source data relate to the relevant INSPIRE data model and what are the preliminary bottlenecks?*
- *What are the available data harmonisation tools and how can these be used to implement INSPIRE data requirements?*
- *What does INSPIRE compliant/conformant mean and how can harmonised datasets be tested against INSPIRE requirements by means of an ATS/ETS?*

1.2.1 Research limitations

This research is not aimed at software development, hence it is not about developing a new tool for data harmonisation but trying to apply existing tools and assess their suitability in the context of the INSPIRE Data Specifications. Also, the research doesn't necessarily propose to develop a complete testing framework for INSPIRE conformance, as some aspects that would need to be tested do not fall within the scope of Data Specifications only, but extend to other requirements of INSPIRE, like Metadata and Network Services. The object of the research will mainly be focused at the Data Specifications component, specifically at the data harmonisation process, therefore testing, will also be focused on these aspects only.

Another limitation of this research is the scale at which the implementation of the INSPIRE requirements is applied (i.e. the case study). For the scope of this research one data theme from the three Annexes will be selected based on data availability and Data Specifications completeness. Regarding the actual harmonisation process, it must be taken into consideration that the INSPIRE data themes obviously cover both types of geo-data, discrete (i.e. vector) and continuous (e.g. raster). The two different data types might require different encoding methods (i.e. discrete data in GML, and continuous, earth science data, in NetCDF) which automatically means completely different harmonisation techniques and tools. This project will focus on discrete data.

Another important point related to the data harmonisation process refers to the coordinate reference system (CRS) transformation. The INSPIRE data models use a unique coordinate reference system as it would have been expected in such an infrastructure. This mean that during the data harmonisation process a CRS transformation will also need to be performed, as all source data will probably be using the local/national coordinate reference system. CRS transformation is a very complex process, which is known to produce unstable results, and could be the object of another research project on its own. Therefore it is not within the scope of this research to perform a 'perfect' coordinate transformation during the data harmonisation process. As a consequence, during the development of the ATS the coordinate transformation will not be thoroughly tested.

1.3 Methodology

The research project will follow the regular phasing of a scientific study. The research will start with an in-depth literature and documentation study which will result in a conceptual process model and an analysis scheme but also the first written chapters of the thesis. The result of this step will carry the research forward to the empirical analysis which is followed by the written empirical chapters. Finally, the last chapters will deal with the interpretation of results, recommendations and conclusions.

Therefore, based on the above and in order to answer the proposed research questions and achieve the proposed objectives some clear defined methodology steps have to be established:

1. Literature review (Chapter 2):
    1.1. Interoperability and data harmonisation concepts – presenting general concepts of interoperability and data harmonisation, what are the benefits and where and how has it been applied before (Section 2.1).
    1.2. Standards in geo-information – in the context of data harmonisation standards play a key role as they facilitate data sharing. A general overview of spatial data standards will be presented (Section 2.2).
    1.3. Spatial data infrastructures – data interoperability and standards will inevitably take the discussion to SDIs, what they are, their purpose and benefits (Section 2.3).
    1.4. INSPIRE Directive – a general presentation of the initiative that proposes to build a European spatial data infrastructure, what has been done, what is the current state of play and what comes next, along with a higher emphasis on Data Specifications description and conformance testing (Section 2.4).
2. Case study presentation and analysis (Chapter 3):

2.1. Source data and selected theme – in order to apply the INSPIRE requirements one data theme from the Directive Annexes will be selected. The choice of the data theme and the source will be presented (Section 3.1).
2.2. Data specifications overview for the selected theme – the harmonisation guidelines for the selected themes will be discussed in detail (Section 3.2).

3. Implementation of the INSPIRE guidelines (Chapters 4 and 5):
   3.1. Presentation of harmonisation tools – overview of the different tools/software packages proposed to use to harmonise data according to the INSPIRE guidelines (Chapter 4).
   3.2. Methodology for applying INSPIRE guidelines – presentation of the steps taken to harmonise data with one of the presented tools (Sections 5.1 and 5.2).
   3.3. Presentation of results, limitations and suggestions – reflection on the transformation process and conclusions (Section 5.3).
4. INSPIRE compliance testing (Chapter 6):
   4.1. Presentation of conformance testing – description of the ATS standard for INSPIRE, what it is, and how should it be performed (Section 6.1).
   4.2. Development of test scenarios that would check the various requirements of INSPIRE towards the data harmonisation process (Section 6.2).
   4.3. Testing overview – discussion on the testing process, results achieved, limitations, and further development and recommendations (Section 6.3).
5. Summary and conclusions (Chapter 7):
   5.1. Summary – the wrap-up of the research will summarise the most important points and will provide specific discussions in regard to the main steps of the research. (Sections 7.1 to 7.3).
   5.2. Research Questions – this section will provide the main conclusions of the research by providing concise answers to the research questions (Section 7.2).
   5.3. Recommendations – based on the outcome of the research, recommendations will be given for future research, and why not for the INSPIRE Data Specifications Drafting Team and/or the Thematic Working Groups for the selected data theme (Section 7.5).

Figure 2 below illustrated the entire methodology of the research in a diagram, with the entire process being clearly split to emphasize which parts will answer which of the research sub-questions. The next and last section of this chapter will briefly summarize the chapters of the thesis and how these are split among the major methodology steps.

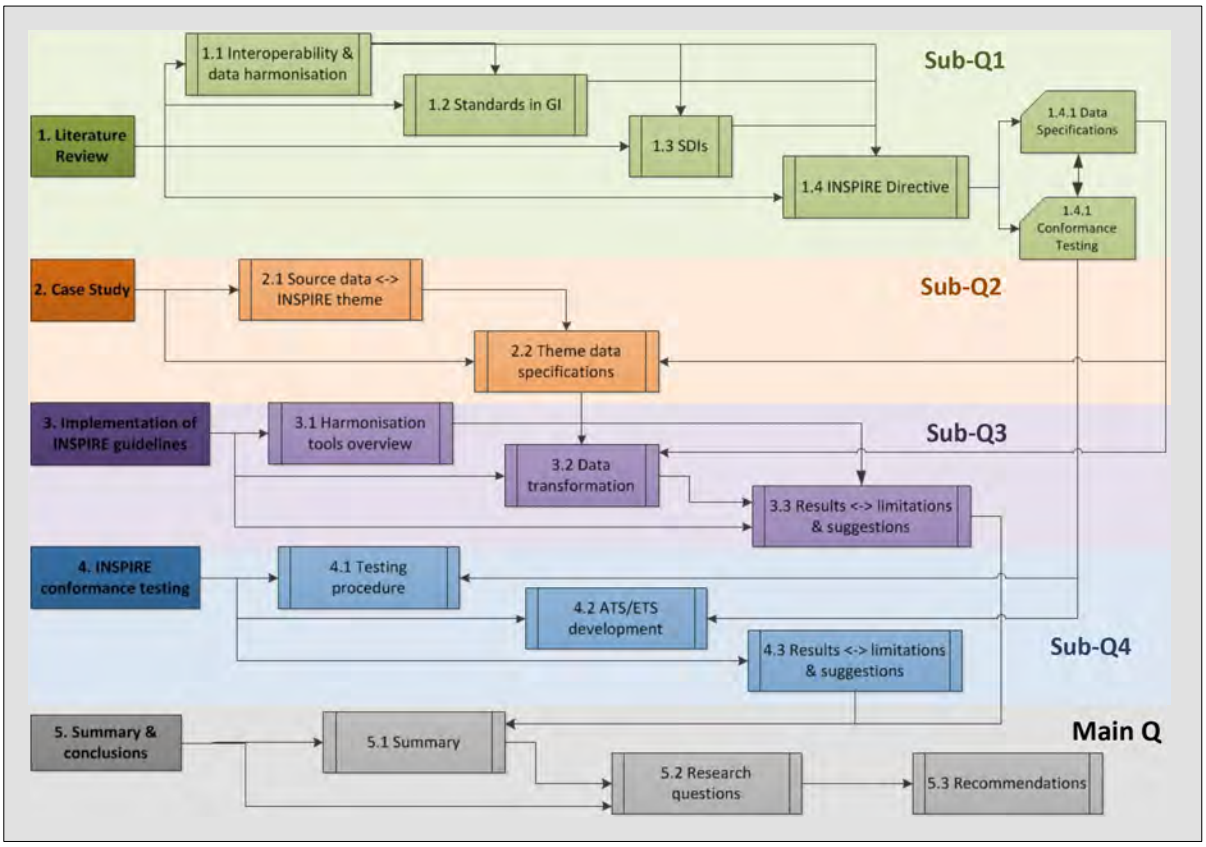


Fig. 2 Methodology diagram

1.4 Summary of chapters

The thesis chapters will follow the structure presented in the methodology section, each methodology step being linked to a specific chapter or section of the thesis. In addition, every thesis chapter (some grouped) will answer each of the research sub-questions, culminating with a summary of the whole process and providing an answer to the main research question in the last chapter of the thesis.

Therefore, Chapter 2 will answer the first research sub-question by means of a literature review. Chapter 3 will answer the second research sub-question by introducing the case study; it presents the source data to be used, and provides an in-depth analysis of the INSPIRE theme that the source data relates to, finalising in a gap analysis between the two data models. Chapter 4 will provide an overview of the available data transformation tools, and a brief approach to the transformation process, concluding with the selection of the tool to be used. Chapter 5 will describe the actual data transformation process, and along with Chapter 4 will answer the third research sub-question. In Chapter 6 the output of the data transformation will have to be tested against INSPIRE requirements by developing an Abstract Test Suite according to the standards, and executing the test cases as well, which will eventually answer the fourth research sub-question. Finally, Chapter 7 will reflect on the entire research process by providing several conclusions and recommendations, and, as already stated, providing an answer to the main research question.

2. Literature review

This chapter will present the background of the research problem, which will be achieved by reviewing numerous resources. A general overview of data interoperability and data harmonisation concepts will be given in Section 2.1, which will lead to a discussion about geo-information standards in Section 2.2. Furthermore, it will be discussed how interoperability and standards aid the implementation of spatial data infrastructures in Section 2.3. Everything will finally point to the fundamentals of the INSPIRE Directive and an in-depth presentation of the INSPIRE Data Specifications in Section 2.4.

2.1 Interoperability & data harmonisation

Interoperability and data harmonisation go hand in hand when it comes to delivering a standardised output, and this is not limited to spatial information only. Data harmonisation is 'a must' process to achieve interoperability, while interoperability can be considered the main keyword of a spatial data infrastructure. The following sub-sections will go into more detail about the two concepts.

2.1.1 Interoperability concepts

Interoperability has been and still is an issue in many areas of information systems. Computers are widely used nowadays and there is a growing need to share information and resources, such as data and services. Bishr (1998), states that interoperability could be described as a form of system's intelligence that enhances the cooperation between the components of information systems. That intelligence is used to find and provide access to services and resources, but also to perform operations across various information systems without knowing in advance what resources are available, or how to acquire them. In this sense, the actual concept of interoperability might rather seem very vague, so the question that is raised is when to call an information system interoperable?

Interoperability can be interpreted and understood in multiple ways by users. It can refer to anything from openness in the software industry, simplification of formats and standards and in the user-systems interaction, to transparency, and similarity between the 'vocabulary' of two datasets, software systems and/or organisations (Goodchild *et al.*, 1997). Therefore, in order to achieve full interoperable systems, it is first needed to assess the various levels of a system at which interoperability needs to occur. According to Manso *et al.* (2009), the most common interoperability levels identified in information systems in past research were related to the technological, syntactic and semantic aspects.

This theory also applies to Geographical Information Systems. Such information systems have been facing interoperability issues from several decades ago. Back in the 70s and 80s the majority of GIS applications were considered islands of information because they were independent systems where spatial data was captured, stored, analysed and displayed, internally only. The advance of technology and the increasing need of users to overcome the costs of data capture, leaded to the sharing of data by transferring it from one 'information island' to another, from one system to another. The transfer took place by a neutral format that could be understood by both source and target system, involving a lot of batch-oriented conversion, which would often lead to redundant

data. Since more than a decade ago, the inefficiency of such an approach was realised by users and the need of interoperable geographical information systems started to grow (Bishr, 1998). The first interoperability levels of geographic information systems to be identified were the information community and institution, enterprise, application, tools, middleware, data store, distributed computing environment and network (Manso *et al.*, 2009; Goodchild *et al.*, 1997). With further research, other interoperability levels were identified related to the following aspects: Semantic, Technical, Syntactic, Pragmatic, Organisational, Schematic/Structural, Dynamic, Legal, Conceptual, Social, Intercommunity, Political/Human, International, Empirical, and Physical. According to the analysis of Manso *et al.* (2009), the most cited interoperability level in literature is Semantic, followed by Technical, Syntactic and Pragmatic levels, whereas Empirical, International and Physical have been the least cited.

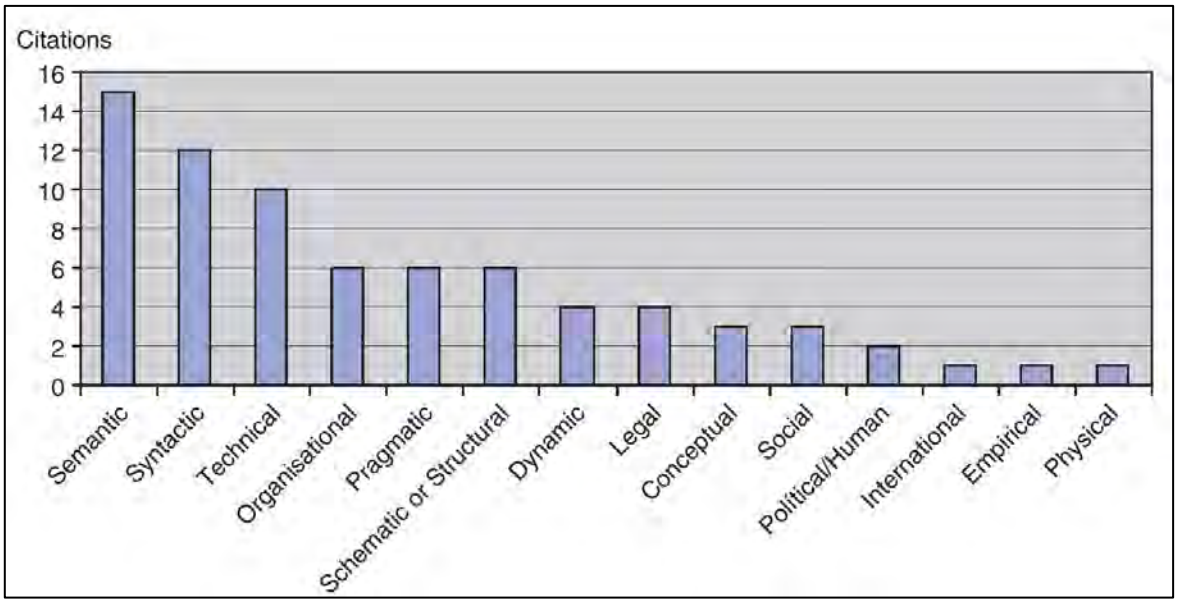


Fig. 3: Interoperability levels citation frequency (Manso et al., 2009)

Most of the theories regarding interoperability levels are structured in interoperability models, with different approaches, having specific advantages and disadvantages with respect to achieving interoperability in a particular context. Worth to mention are the Levels of Conceptual Interoperability Model (LCIM), defined by Tolk (2003) and refined by Turnitsa & Tolk (2006), as well as the Intermodel5 (Shanzhen *et al.*, 1999) that have been applied in the GIS domain (Manso & Wachowicz, 2009).

The selection of the interoperability levels and how they are connected to each other establishes the type of interoperability model. According to Manso *et al.* (2009), the integrated interoperability model fits best within the GIS context. With integrated models there is a common template where separate interoperability levels are associated to build up into a coherent whole. Each level would fulfil different functions, while the communication and data sharing would be performed by standardised procedures and using common databases. The integrated interoperability model is based most of the time on a hierarchical relation among the interoperability levels that reflects the degree of capability for interoperation. Figure 4 below, presents the integrated interoperability model by Turnitsa & Tolk (2006), based on such a hierarchical relation.

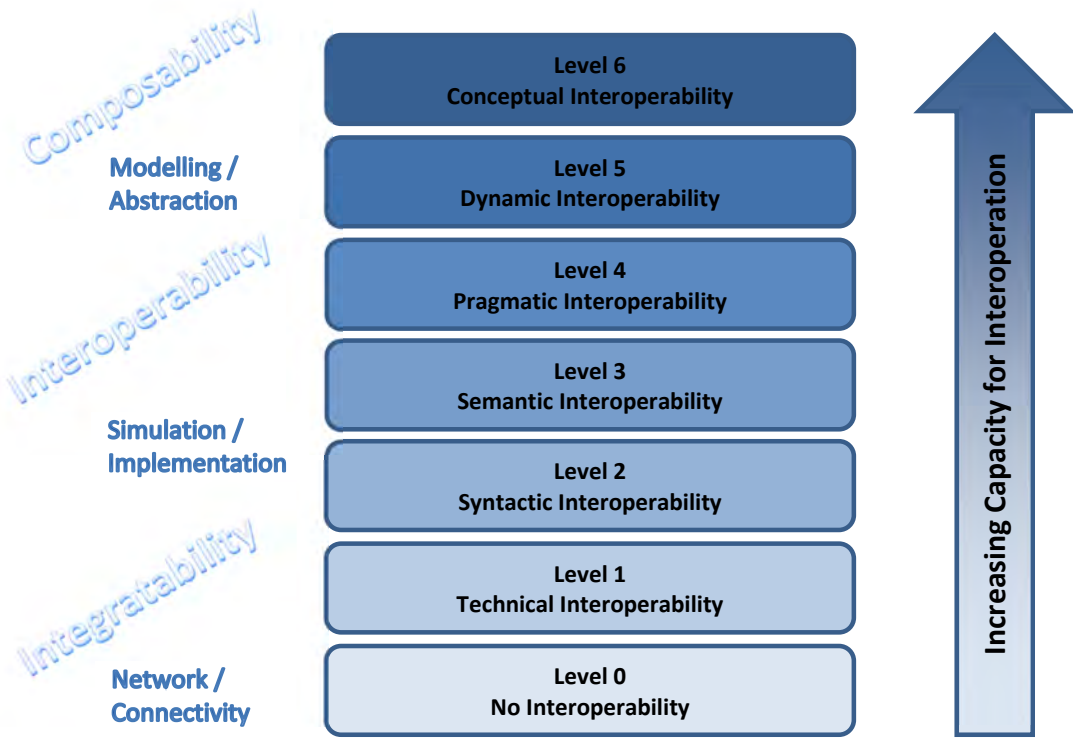


Fig. 4: The integrated interoperability model proposed by Turnitsa & Tolk (2006)

The challenge that remains today in researching GIS interoperability is related to the gap that exists between the different interoperability models. Closing that gap would provide a unified approach based on the strong and weak points of each interoperability model and their seamless integration. Another aspect to point out is concerning the existing relationships between the different interoperability levels. Manso *et al.* (2009), state that these relationships do not have to necessarily be of hierarchical nature. One example given is that, for instance, to reach conceptual interoperability, for which data models and application schemas are required, it is important to ensure syntactic and semantic interoperability although the pragmatic or dynamic interoperability levels do not seem necessary.

Taking a step back from interoperability levels and models, and trying to capture the bigger picture of how interoperability connects different systems, Lasshuyt & van Hekken (2001) identified three basic architectures for interoperable systems (see Figure 5).

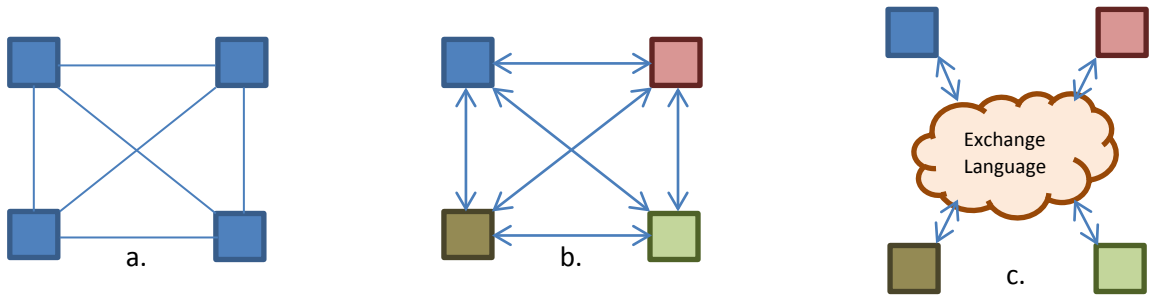


Fig. 5: Basic architectures for interoperable systems (adapted after Lasshuyt & van Hekken, 2001)

The architecture in Figure 5a depicts the standardisation of systems, where each system is identical. This situation often occurs when a distributed organisation sets up different systems throughout all its branches/offices and bases them on a single standard (i.e. a corporate information standard). However, in most cases this approach is prone to failure because the architecture is developed according to standards of a particular organisation or user community. Figure 5b illustrates a bilateral exchange architecture, where each system has its own internal architecture and dedicated interfaces are required between each pair of interconnected systems to exchange information. Finally, Figure 5c illustrates the standardisation of the exchange language architecture, which is considered to be the most practical solution for achieving true interoperability. The major advantage of this approach is that it is highly flexible, being able to adapt new systems without the need to alter the existing ones.

In the context of interoperability for a European Spatial Data Infrastructure (ESDI), the INSPIRE initiative seems to aim at being an 'exchange language' between the different systems (i.e. Member States / national SDIs) as depicted in Figure 5c.

2.1.2 Data harmonisation

The definition of interoperability is in most cases data-driven, especially in the case of geographic information systems where the main goal is to enable spatial data sharing. Therefore, spatial data interoperability can be defined as the ability to access, share and manipulate spatial data, stored in heterogeneous distributed repositories (Nowak *et al.*, 2005).

However, there are several aspects to consider when it comes to spatial data heterogeneity. Data format, coverage, scale, reference system, data model, ontologies and metadata schema are all contributing to the limitations of geo-data applications. These aspects can greatly vary between different organisations and especially in the case of trans-boundary cooperation projects. As an example, in the European Union where a large number of cross-border projects are initiated, the spatial data has traditionally been scattered and fragmented, even within single countries. There are 15 different tide gauge reference points across the EU countries, and the difference between these national vertical datums and the UELN 95/98 (United European Levelling Network) vary from -231 cm (for Belgium) to +22 cm (for Finland) (Villa *et al.*, 2007). Since this is only one of the many examples of discrepancies in spatial data heterogeneity, there is a strong need for international level harmonisation initiatives in the field of geographic information. The need for harmonised data is a fundamental point in building a Spatial Data Infrastructure which would bring together different data sources and different services and applications for retrieved spatial data.

According to Östman (2010) data harmonisation can be applied when data from different sources need to be merged, when there is an initiative to provide data according to a specific standard, or when non-standardised data needs to be imported to a target application. In this sense it is stated that there are also different definitions or views on data harmonisation:

- The act of specifying common characteristics of datasets.
- The act of making datasets compliant with specified characteristics.
- The act of removing tensions between two or several datasets.

Data harmonisation can be very complex comprising many different aspects, and there are also many different ways to define concepts related to data harmonisation, depending on which of these

aspects are in focus and which would better serve the needs of a specific initiative. One of the main challenges regarding the identification of harmonisation issues to be tackled (i.e. data specifications) is to find the right balance between a simple, easy to implement solution and a complex, powerful solution depending again on the needs, but also feasibility constraints (DT DS 2008b, Villa *et al.* 2007) (see Figure 6).

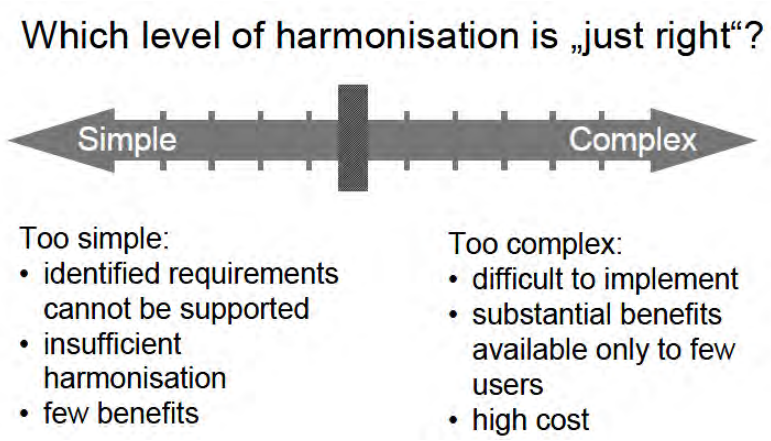


Fig. 6: Identification of harmonisation level (DT DS, 2008b)

Some of these aspects have already been identified earlier as limitations of data heterogeneity. To clearly emphasize the difference in scope when it comes to data harmonisation, it is enough to do a short review of different initiatives in the field. In the INSPIRE Directive data harmonisation is actually replaced with the term 'data interoperability', and it refers to *"the possibility for spatial datasets to be combined, and for services to interact, without repetitive manual intervention, in such a way that the result is coherent and the added value of the datasets and services is enhanced"* (European Commission, 2007). INSPIRE identifies several different aspects relevant for data harmonisation, also called interoperability components (see Figure 7).

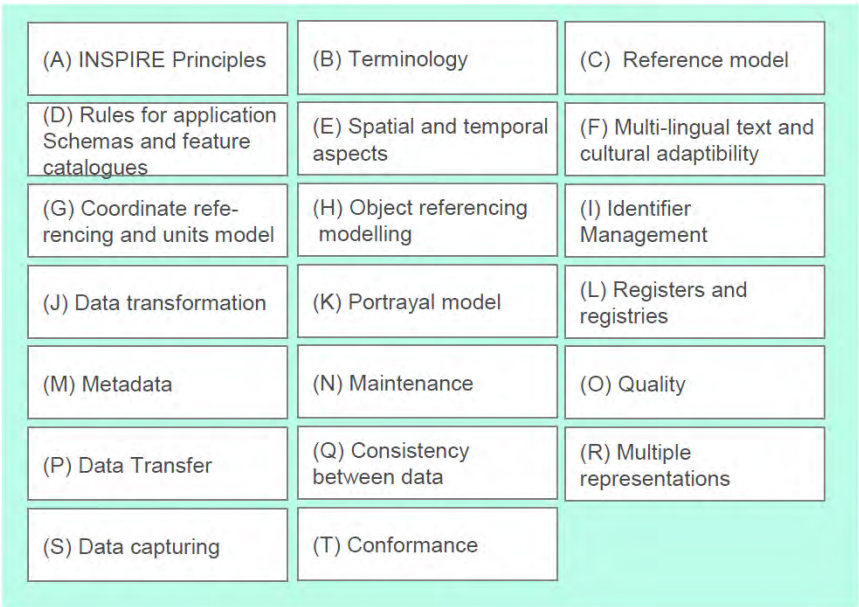


Fig. 7: INSPIRE data interoperability components (Drafting Team 'Data Specifications', 2010a)

On the other hand, another spatial data harmonisation initiative across Europe, called HUMBOLDT, has a slightly different view over these components. To be more concise, not all of the INSPIRE interoperability components are within the scope of HUMBOLDT, in which data harmonisation is defined as *"creating the possibility to combine data from heterogeneous sources into integrated, consistent and unambiguous information products, in a way that is of no concern to the end-user"* (de Vries *et al.*, 2007). Below are the most relevant harmonisation aspects that have been identified by Fichtinger *et al.* (2011) in HUMBOLDT, with the differences becoming apparent when comparing them to the INSPIRE interoperability components in Figure 7.

- Data formats and/or type of web service
- Coordinate reference systems
- Conceptual data model
- Classification Schemes
- Terminology
- Metadata
- Aggregation/Multiple representations
- Portrayal
- Multilingualism
- Temporal aspects
- Conformance to standards

All HUMBOLDT harmonisation aspects occur at general level, that of the data model or of the whole dataset. There may also be other aspects that occur at instance level, that of the object level, for individual objects or group of objects within a dataset. Some of these can be found in the remaining interoperability components of INSPIRE. They can relate to spatial consistency at country borders, like solving conflicts in case of spatially overlapping, detecting and possibly solving gaps, or merging geometry of spatial objects at both sides of a border, process also called edge-matching (Fichtinger *et al.* 2011, de Vries *et al.* 2010). Despite the slight difference of harmonisation issues that are dealt with, HUMBOLDT contributes to the implementation of a European Spatial Infrastructure (ESDI) that integrates the diversity of spatial data available for a multitude of European organisations, and it does support the INSPIRE Directive and its goals.

Data harmonisation has been a research issue for a long time, so no matter the complexity of the harmonisation level, there are three central issues that are usually prioritised and dealt with before other data harmonisation aspects are considered (Fichtinger *et al.* 2011, Villa *et al.* 2007):

- Syntax – related to different data formats.
- Structure – related to differences in conceptual schemas (e.g. differences in attributes of two schemas).
- Semantics – related to differences in the intended meaning of terms in specific contexts).

Syntax heterogeneity is not a very complex issue, giving the international standards that have been developed in the last few years. On the other hand, structural heterogeneity involves mapping of data models and in order to do so, it involves knowledge of the semantics behind. In this sense, semantic heterogeneity issues arise from a series of reasons like naming conflicts, scale conflicts, and precision or resolution conflicts (Friis-Christensen *et al.*, 2005).

Ultimately, the actual data harmonisation process mainly consists of two distinct work packages (Fichtinger *et al.*, 2011):

- Definition of the target schema – a common process and methodology of developing data specifications in order to achieve a harmonised conceptual schema. In this sense the RISE (Reference Specifications for Europe) project has proposed a repeatable methodology, and guidelines for developing, adopting and maintaining data product specifications. This guide is specifically addressed to issues concerning the harmonisation of heterogeneous data sources (Portele *et al.*, 2007).
- Processing – performing various processing steps to transform heterogeneous data from different sources to match the target data specifications. Processing steps can include, but are not limited to, transformation of data from source to target conceptual schema (i.e. Extract – Trransform – Load), coordinate reference system transformation, edge matching, language transformation, etc.

Since the objectives of this research are mainly centred on the second work package, there will be a stronger emphasize in reviewing the data transformation methods.

According to Östman (2010), the most important operation of data transformation is often considered to be schema translation from a source schema to a target schema. A schema in this context can be defined as a formal description of a model, and there are different schema types/levels. There is conceptual schema, where the model consists of data structures, code lists, etc. that can be expressed in UML (Unified Modelling Language). Then there is logical and/or physical schema which is related more to the physical structure of the datasets, and finally, there are the pure transfer files which are encoded in XML/GML formats. The schema translation is processed throughout three main steps:

- Schema matching – the process of finding semantically related objects (i.e. feature classes, attributes that correspond to each other). This is achieved through ontologies, thesaurus, dictionaries, etc.
- Schema mapping – the process of finding transformation rules (i.e. how to convert data from one schema or type of representation to another one). This can include different operations like reclassification, data type conversions, etc.
- Schema transformation – the process of extracting the data from the source database, transforming it according to the rules established in the schema mapping, and loading the data into the target database or application. This step is also called an Extract-Transform-Load (ETL) process.

Figure 8 below, presents the HUMBOLDT view on schema translation, which follows the approach of Staub (2007). As described, schema translation occurs at different levels, from Model A to Model B.

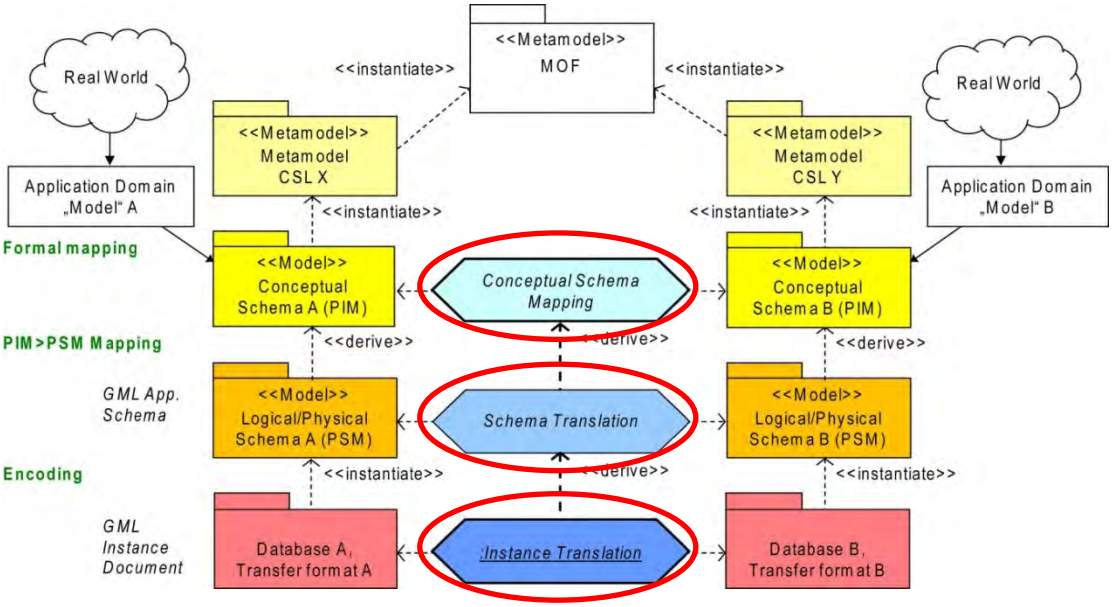


Fig. 8: HUMBOLDT view on schema translation (Staub, 2007)

Conceptual mapping, from the source data to the target data, occurs at the conceptual schema level. Below that there are logical/physical schema translations (e.g. XML schema translations), and finally at the lowest level there is the actual encoding of the data (e.g. into GML files), and instance translations are performed. All these transformations at the different levels can be aided by different software tools. For instance, translating a GML file to another one at the bottom level (i.e. instance translation) is done via XSLT (Extensible Stylesheet Language Transformation) translations, a declarative XML-based language used for the transformation of XML/GML documents. Ideally is to work on the conceptual level, because this will assure a high level of automation to the lower levels. Although there have been several studies and initiatives in this sense, there is currently no commonly agreed standard for a mapping language in the geospatial community (Fichtinger *et al.*, 2011).

As a closing point for this section, it must be mentioned that an integral part of data harmonisation and especially for the development of data specifications, there must be an agreement on technical standards that form the basis for interoperability. A review of existing standards in the geo-information field will be given in the next section.

2.2 Standards in geo-information

Standards are a fundamental part of modern society and an organised way for ensuring best practice, common design, safety and many other benefits across every field of industry and science. There are a series of bodies that coordinate and promote the generation of standards, the most important being at international level. These bodies now play a crucial role in the geo-information field as well, with the most relevant organisations being the International Organisation for Standardisation (ISO), the European Committee for Standardisation (CEN), and Open Geospatial Consortium (OGC). From the integrated interoperability model point of view, standards are being used to ensure syntactic, pragmatic and semantic interoperability.

2.2.1 International Organisation for Standardisation

The international Organisation for Standardisation, widely known as ISO, is an accepted and established international standard-setting body composed of representatives from various national organisations, one member per country, summing up to 161 countries at the moment. ISO has been founded in 1947 and has its headquarters in Geneva, Switzerland. ISO is a non-governmental organisation that forms a bridge between the public and private sectors, having members which are part of the governmental structure of their countries but also members active in the private sector, having been set up by national partnerships of industry associations. Although adhering to ISO standards is optional, many of these standards become law, being implemented in the form of national standards or international treaties. The reason behind this is that international standardisation is market-driven and therefore based on voluntary involvement of all interests in the market-place, with the ultimate goal of providing global solutions to satisfy industries and customers worldwide.

The teams that create the actual ISO standards are called Technical Committees in the ISO structure. These committees deal with standards in specific sectors and domains. The geo-information domain is included in the 'Information processing, graphics, photography and services' sector, and the committee responsible for it is ISO/TC 211 Geographic Information/Geomatics. The scope of ISO/TC 211 *"aims to establish a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the Earth. These standards may specify, for geographic information, methods, tools and services for data management, acquiring, processing, analysing, accessing, presenting and transferring such data in digital/electronic form between users, systems and locations"*. To date, 56 standards were published by this working group, standards that are also known as the ISO 19100 series. Important standards have been published for various aspects of geo-information, some example being like the Reference Model (ISO 19101), Metadata (ISO 19115), Conformance Testing (ISO 19105), Data Product Specifications (19131), and many others.

The ISO/TC 211 is composed of 33 participant members, as well as 30 observing members, represented by national standardisation bodies. There are also several internal liaisons with other ISO Technical Committees, as well as external liaisons with several international professional organisations. One organisation heavily involved in the work of ISO/TC 211 is the Open Geospatial Consortium (OGC), the two organisations having a working arrangement that often results in identical or nearly identical standards often being adopted by both of them. Among the standards developed by OGC that became ISO standards are the Geography Markup Language (GML) (ISO 19136), Web Map Service (WMS) (ISO 19128), and Web Feature Service (WFS) (ISO 19142).

2.2.2 European Committee for Standardisation

The European Committee for Standardisation (CEN) is an international non-profit organisation set up under Belgian law, which provides a platform for the development of European Standards (ENs) and other consensus documents. CEN's national members are the national standard bodies of the 27 EU countries, Croatia, Turkey, plus three countries of the European Free Trade Association (Iceland, Norway and Switzerland), summing up to 32 national members. These members work together to develop standards in a large numbers of sectors to help build the European internal market in goods and services, removing barriers to trade and strengthening Europe's position in the global economy.

An important aspect to note is that compared to ISO standards, CEN standards are not voluntarily adopted, but are an obligation for the national standard bodies that are members.

Just like ISO standards, CEN standards are developed by various Technical Committees with expertise in different domains. The CEN Technical Committee responsible for geo-information standards is CEN/TC 287 Geographic information. This was actually the first initiative of an international standardisation organisation in the geo-information field being set up in 1990, while ISO/TC 211 was formed in 1994. The work of CEN/TC 287 is carried out in close co-operation with ISO/TC 211, but also OGC, as currently CEN/TC287 has stopped developing its own standards and started adopting ISO and OGC standards. Therefore, one of the objectives of CEN/TC 287 is to ensure interoperability by adopting the ISO 19100 series as European standards where appropriate. Other objectives also relate to cooperation for the development of new standards, as well as adopting INSPIRE's requirements as European standards as they become available.

2.2.3 Open Geospatial Consortium

The Open Geospatial Consortium is a non-profit international voluntary consensus standards organisation that is developing standards for the geospatial and location based markets. Having been founded in 1994, the OGC has member organisations all across the world, summing up to 444 organisations belonging to sectors ranging from commercial and governmental to non-profit and research. The declared mission of OGC is to serve as a global forum for the collaboration of developers and users of spatial data products and services, and to advance the development of international standards for geospatial interoperability.

Most of the OGC standards are based on a generalised architecture presented in a set of documents collectively called the 'Abstract Specifications', which describes a basic data model for representing geographic features. Beside the Abstract Specifications, the OGC have developed a growing number of specifications documents, or implementing standards, to serve specific needs for interoperable location and geo-information technology. The OGC have developed to date over 50 standards, many of them becoming ISO standards due to the close cooperation with ISO/TC 211 Geographic Information/Geomatics (and thus CEN/TC 287 Geographic Information), as already mentioned. Like ISO standards, OGC standards are voluntarily, however they are widely adopted because they are built on a consensus of numerous members of the GI community.

2.3 Spatial data infrastructures

In the current age of information, geographic information is one of the most critical elements that aid decision making throughout many sectors. In this sense, many goals that different organisation want to achieve, often by collaborating with each other, can only be achieved if good, consistent spatial data is available and readily accessible. The process of researching interoperability in the GI domain and developing methodologies for spatial data harmonisation according to global industry, or other local standards, will ultimately facilitate the development of spatial data infrastructures (SDI) at larger or smaller scales. This section will review the main aspects of SDI, its components and the implications of the INSPIRE Directive in this context.

Spatial data infrastructures (SDIs) are referred in literature also as geographic information infrastructures (GIIs), geospatial data infrastructure, geographic data infrastructure, or spatial information infrastructure. Although the various terminologies have been used to identify the same phenomenon, Van Loenen (2006) suggests that semantically, a spatial data infrastructure has a different scope than a geographic information infrastructure. This is due to the fact that 'spatial' can include any space, not only the space on the Earth's surface. It can include 3D images of the human body for medical purposes or information on the design of a car, whereas 'geographic' refers specifically to spatial information for the Earth's surface and near surface. Moreover, 'information' represents more than 'data'. Information refers to data with a value added due to processing and human interpretation (Sudra, 2010). However, spatial data infrastructure still proved to be the most popular term, being used as well in the context of the INSPIRE Directive, as it will be used in this thesis.

A Spatial Data Infrastructures (SDI), as a term, is used to denote the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data. It provides a foundation for spatial data discovery, evaluation, and application for users and data providers within all level of government, the commercial sector, the non-profit activities, academia and by citizens in general (Nebert, 2004). This would automatically avoid duplicate expenses that are associated with generation and maintenance of data and its integration with data originating from different sources. Therefore, SDIs have become very important in determining the way in which spatial data is used throughout an organisation, a country, different regions and the world. According to Rajabifard & Williamson (2001), SDI can be defined in many ways by stakeholders from different disciplines or by different nations. For instance:

- The Australian and New Zealand Land Information Council (ANZLIC) defines a National SDI as *"comprising of four components an institutional framework, technical standards, fundamental datasets, and clearinghouse networks. The institutional framework defines the policy and administrative arrangements for building, maintaining, accessing and applying the standards and datasets. The technical standards define the technical characteristics of the fundamental datasets. The fundamental datasets are produced within the institutional framework and fully comply with the technical standards. The clearinghouse network is the means by which the fundamental datasets are made accessible to the community, in accordance with policy determined within the institutional framework, and to agreed technical standards"*.
- The Federal Geographic Data Committee (FGDC) defines the U.S. National SDI as *"an umbrella of policies, standards and procedures under which organisations and technologies interact to generate more efficient use, management and production of geospatial data"*.
- According to Bernard *et al.* (2005), in Europe, where the INSPIRE Directive is setting up a general framework for SDI, the term SDI follows the definition of the SDI Cookbook, meaning conceptually that it *"encompasses the policies, organisational remits, data, technologies, standard delivery mechanisms and financial and human resources necessary to ensure that those working with spatial data, whether at the global or the local scale, are not impeded in meeting their objectives"*.

On the other hand, Van Loenen (2006) identified four different perspectives over the various SDI definitions. These are the identificational, technological, organisational, and productional. The

identificational perspective is focusing more on justifying the investment that is needed to set up a SDI, explaining the major benefits. The technological perspective emphasizes on the structure and functions of SDIs and how there would greatly benefit the users. The organisational perspective widens up the SDI concept by taking into account the organisational context (i.e. policies, financial and human resources, etc.). Finally, the productional perspective refers to SDI as a dynamic concept, emphasizing on the interaction between suppliers and users of geographic information.

Based on the various SDI definitions and perspectives, Rajabifard & Williamson (2001) identify the core components of SDI to be people, data, access network, policy, and standards. Figure 9 below clearly shows the interactions within the SDI framework.

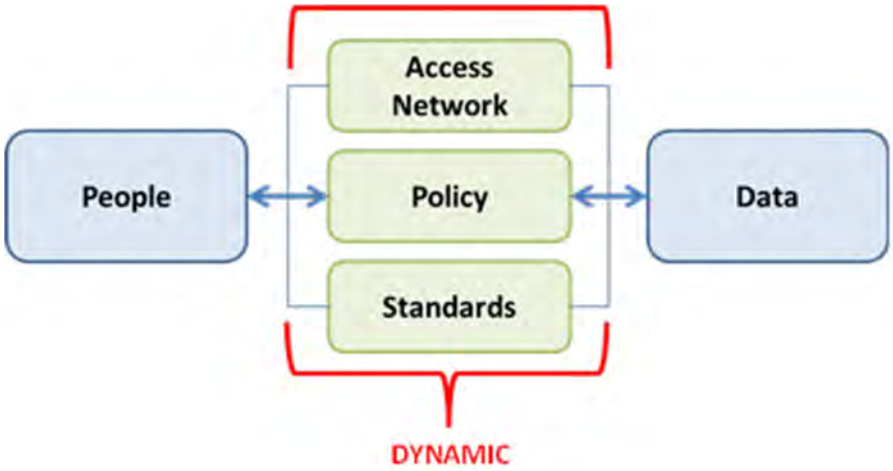


Fig. 9: Nature and relations of SDI components (adapted after Rajabifard & Williamson, 2001)

First of all there is the important and fundamental interaction between people and data as one category, which is based on the productional perspective of SDI described earlier, and second category is consisted of the main technological components: the access network, policy and standards. A very important property of the technological components is dynamism, due to the fast-paced technological developments, and the role for mediation of rights, restrictions and responsibilities between changes in people and data.

SDI initiatives may be developed at different political-administrative levels, ranging from corporate to global level. As a result, Rajabifard *et al.*, 2000, developed a model of SDI hierarchy that includes all SDI levels.

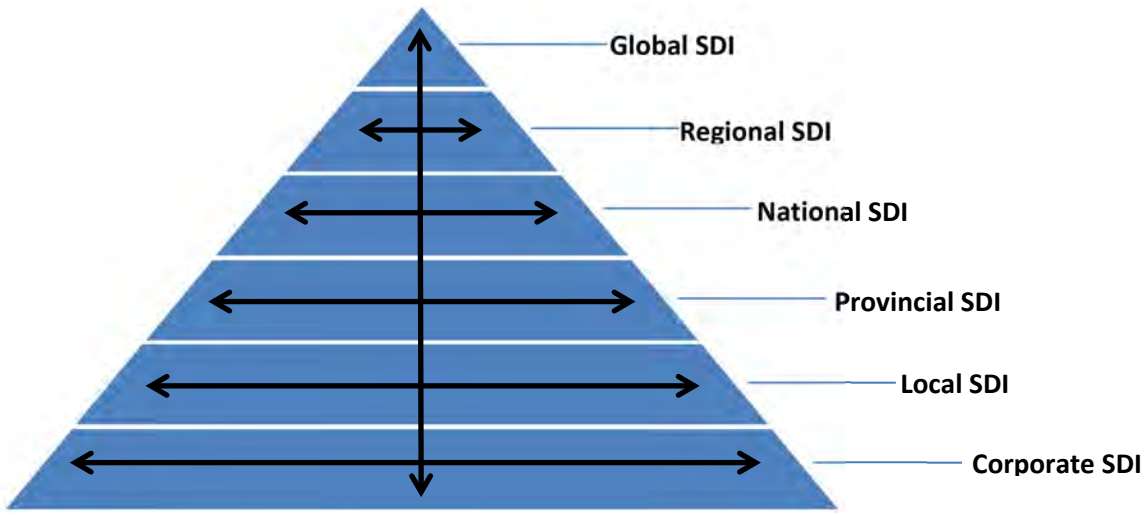


Fig. 10: SDI hierarchy model (adapted after Rajabifard et al., 2000)

Figure 10 illustrates the SDI hierarchy which is made up of inter-connected SDIs at different levels. Each SDI from the local level and above should be formed by the integration of SDIs developed at the lower levels. In addition to that, there can be two views regarding the nature of a SDI hierarchy:

- Umbrella view – the SDI at the higher level encloses all SDI components of those levels below.
- Building block view – any SDI level serves as the building blocks for higher levels in hierarchy.

While corporate, local, and to some extent provincial SDIs, have been already widely implemented across the world for the benefits and efficiency they bring in workflows and decision making at different political/administrative levels, national SDIs are also starting to emerge at a very rapid pace. But these countries also found the need to cooperate with other countries in order to develop a regional SDI, or even a global SDI, to assist in decision-making and collaboration across national borders. Maybe the most representative initiative worldwide at the moment is the emergence of the European Spatial Data Infrastructure (ESDI) through the INSPIRE Directive. In that respect, the ESDI adopted the building block view on SDI hierarchy, since it will be built on top of existing and future provincial and national SDIs. INSPIRE will basically define the requirements for the ESDI, and national SDIs will have to comply. A thorough description of the INSPIRE Directive will be given in the next section of this thesis.

2.4 INSPIRE Directive

INSPIRE stands for Infrastructure for SPatial InfoRmation in Europe, and its primary role is to set down a general framework for a European Spatial Data Infrastructure. This will enable the sharing of environmental spatial information among public sector organisations and better facilitate public access to spatial information across Europe. As mentioned in the problem statement section of this thesis, INSPIRE was initiated because of the increasing need for such a policy at European level, especially between the member states of the European Union, that would help to make geographical information more accessible and interoperable for a wide range of purposes supporting

sustainable development. This need emerged while identifying several issues that were becoming more and more evident at European level, in regard to interoperability of geographic information. Below are some of the most important points:

- Lack of reference and authentic data.
- Inaccessible spatial data due to difficulty to find, sometime cumbersome process to obtain, and very often expensive to purchase.
- Gaps in the availability of spatial data and a lot of duplication at the same time, but also sometimes of a general doubtful quality.
- Very rare spatial data would be harmonised across borders due to total different policies that would be separated by national borders, but also inside the same country between different provinces, counties or even municipalities.
- Very often spatial datasets would be difficult to interpret because of lack of documentation (i.e. missing or incomplete metadata).

An objective response to these issues came in September 2001, when the first INSPIRE, or at that time the E-ESDI expert group (Environmental European Spatial Data Infrastructure) meeting was held in Brussels. Nevertheless, a lot of research and activity was going on already before that moment, through various working groups of the European Commission and EU funded projects. Also, later that year, and ESDI Organisation and E-ESDI Action Plan was published by the Commission, which served as reference document for preparing the ground for the proposal of the INSPIRE legislative framework. Many preparation work followed, and in 2004, the INSPIRE proposal for a Directive was adopted by the Commission, which was a major milestone for the use of geo-information in Europe. In 2005, another important step was made, when it was decided that the definition and preparation of the Implementing Rules cannot be developed in isolation, but need to take into account the various stakeholders of an initiative like INSPIRE. Therefore, an open call was launched for the registration of interest divided in two groups: Spatial Data Interest Communities (SDIC) and Legally Mandated Organisations (LMO). SDICs were identified as self-organised communities bringing together experts in the field, financial resources and policies, producers and end-users of spatial information, organised by country/region, sector or thematic issue. On the other hand, LMOs represent the organisations at local, regional, national, or international level that have a formal legal mandate, giving them the responsibility for specific thematic spatial data resources. These organisations (i.e. SDIC and LMO) were asked to put forward experts and reference material to support the preparation of the Implementing Rules. As a statistic, at present there are 471 SDICs and 269 LMOs involved in INSPIRE. Finally, INSPIRE was officially adopted through the Directive 2007/2/EC of the European Parliament and of the Council, and came into force on 15th May, 2007.

INSPIRE is based on some common declared principles that help to understand the vision of the initiative (http://inspire.jrc.ec.europa.eu/):

- Data should be collected only once and kept where it can be maintained most effectively.
- It should be possible to combine seamless spatial information from different sources across Europe and share it with many users and applications.
- It should be possible for information collected at one level/scale to be shared with all levels/scales; detailed for thorough investigations, general for strategic purposes.

- Geographic information needed for good governance at all levels should be readily and transparently available.
- Easy to find what geographic information is available, how it can be used to meet a particular need, and under which conditions it can be acquired and used.

It also must be noted that INSPIRE applies to geographic information that is under the jurisdiction of public authorities, and that is used by public authorities in the performance of their public tasks. It does not set any obligations on the private sector. Also, member states are not obliged to collect new data, if they do, the data should be INSPIRE compliant. Neither are the member states obliged to give up on the data models set up by the national standards, as long there is access to an INSPIRE compliant version. Finally, the Directive does not affect ownership rights and it does not claim that spatial data should be available for free. However, view services and metadata should be made available by discovery services through the INSPIRE geo-portal. For data download, different conditions may apply that are subject to the data provider policy.

The spatial information considered under INSPIRE is extensive and includes a variety of topical themes, 34 in total, divided into three Annexes as follows:

Annex I

1. Coordinate reference systems
2. Geographical grid systems
3. Geographical names
4. Administrative units
5. Addresses
6. Cadastral parcels
7. Transport networks
8. Hydrography
9. Protected sites

Annex II

1. Elevation
2. Land cover
3. Ortho-imagery
4. Geology

Annex III

1. Statistical units
2. Buildings
3. Soil
4. Land use
5. Human health and safety
6. Utility and government services
7. Environmental monitoring facilities
8. Production and industrial facilities
9. Agricultural and aquaculture facilities
10. Population distribution – demography
11. Area management / restriction / regulation zones & reporting units
12. Natural risk zones
13. Atmospheric conditions
14. Meteorological geographical features
15. Oceanographic geographical features
16. Sea regions
17. Bio-geographical regions
18. Habitats and biotopes
19. Species distribution
20. Energy resources
21. Mineral resources

The classification of the data themes in the three Annexes has several consequences. First of all Annex I themes are considered to be reference data, the type of data that is needed in any GIS application to start with. Annex II and Annex III contain more specific themes. Another consequence derives from here, and it is related to the implementation calendar, as requirements for Annex I have to be fulfilled first, then requirements for Annex II and Annex III. In the same manner, the requirements themselves vary, as there are more requirements for Annexes I and II than for Annex III (Drafting Team 'Data Specifications', 2008a). Figure 11 below provides a road map of the main INSPIRE milestones (actual implementation, not just adoption):

	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Monitoring & Reporting (implementation)	15 May										
Metadata (made available)	3 Dec			3 Dec							
Data & Service Sharing arrangements		19 Oct*		19 Apr**							
Discovery Services (operational)		9 Nov									
View Services (operational)		9 Nov									
Download Services (operational)			28 Dec								
Transformation Services (operational)			28 Dec								
Data & Service Harmonisation (available)				4 Feb*		15 Oct**			4 Feb**		15 Oct**

* - New/Restructured datasets/arrangements
** - Existing datasets/arrangements

Annex I | Annex I & II | Annex II & III | Annex III | All Annexes

Fig. 11: INSPIRE milestones

To ensure that the SDIs of the member states are compatible and usable at the Community level and in a trans-boundary context, the INSPIRE Directive requires that common Implementing Rules are adopted in a number of specific areas:

- Metadata – in order for a user to be able to find spatial datasets and services, and to establish whether they may be used and for what purpose, Member States should provide description in the form of metadata for the these spatial datasets and services. Metadata regulations have been adopted by the Commission in December, 2008, while availability of metadata for spatial datasets and services corresponding to Annex I and II have been made available in December, 2010, and for Annex III has to be made available by December 2013.
- Data Specifications – making it possible to set the framework for interoperability and actual data harmonisation to common standards, as this represents the core component of INSPIRE. With Data Specifications already adopted for Annex I data themes, work is still in progress adopting Data Specifications for Annex II and III. There will be an extended discussion on Data Specifications in the next section of the thesis, as this represents the central INSPIRE aspect for the research.
- Network Services – Member States have to establish and operate several of these services as follows:
    - Discovery Services – making it possible to search for spatial datasets and services on the basis of the content of the corresponding metadata and to display the content of the metadata. Regulations adopted in December, 2009, and have been operational since November 2011.
    - View Services – making it possible, as a minimum, to display, navigate, zoom in/out, pan, or overlay viewable spatial datasets and to display legend information and any relevant content of metadata. Regulations adopted in December, 2009, and have been operational since November 2011.
    - Download Services – enables copies of spatial datasets, or parts of them, to be downloaded and, where practicable, accessed directly. Regulations adopted in November, 2010, and should be operational by December, 2012.
    - Transformation Services – making it possible for spatial datasets to be transformed with a view to achieving interoperability. Regulations adopted in November, 2010, and should be operational by December, 2012.

    - Invoke Spatial Data Services – allow defining both the data inputs and data outputs expected by the spatial service and define a workflow or service chain combining multiple services. It also allows the definition of a web service interface managing and accessing (executing) workflows or service chains. Little work has been done in this aspect, being expected that preliminary regulations will be submitted for discussion by June, 2013.
- Data and Service Sharing – access to spatial data and services represents a central aspect of INSPIRE. Since, the Community institutions and bodies have to integrate and assess spatial information from all the Member States, INSPIRE recognizes the need to tackle the challenge of gaining access to and use spatial data and spatial data services in accordance with an agreed set of harmonised conditions. Adoption of the regulations of this component commenced in March, 2010.
- Monitoring and Reporting – in order to have a solid basis for decision making related to the implementation of INSPIRE Directive and to the future evolution of INSPIRE, continuous monitoring of the implementation of the Directive and regular reporting are necessary. Monitoring and reporting have to cover the 4 main fields of INSPIRE Directive: metadata, spatial data sets and services, network services, data sharing. Monitoring follows a quantitative approach and takes place every year, while reporting covers more qualitative aspects and takes place every 3 years (starting 2010). Monitoring and Reporting regulations have been adopted by the Commission in June, 2009.

The Implementing Rules that apply to the INSPIRE components are adopted as Commission Decisions or Regulations, and are binding in their entirety. The Commission is assisted in the process of adopting such rules by a regulatory committee composed by representatives of the Member States and chaired by a representative of the Commission. Furthermore, the Implementing Rules are supported by additional documentation which serves as implementation guidelines and are comprised of various theme-specific legislative acts, conceptual models, technical guidance reports, dictionaries, application schemas, etc. The creation of the Implementing Rules and of the documentation that accompanies it in each of the five INSPIRE components, is the responsibility of the Drafting Teams (DT). Their work is based on existing standards (i.e. OGC and ISO/TC 211), that have been presented in a previous section. These teams are groups of experts proposed by the SDICs and LMOs and selected by the Commission. The DTs were involved early in the process, during the INSPIRE one call for SDICs and LMOs, back in 2005. Their composition has since then evolved based on the availability of experts and changes in the requirements of the DTs.

Figure 12 below provides a simplified overview of the key elements in the technical architecture of INSPIRE. The core component of the diagram is the actual data which falls under the jurisdiction of the 'Drafting Team Data Specifications' (DT DS), highlighted in the red rectangle. All other resources shown in the diagram (e.g. metadata – MD, network services – NS, etc.) are only needed to find, access, interpret or use the spatial objects in the spatial data sets that form the part of the infrastructure (Drafting Teams 'Data Specifications', 'Metadata', 'Network Services', 2007).

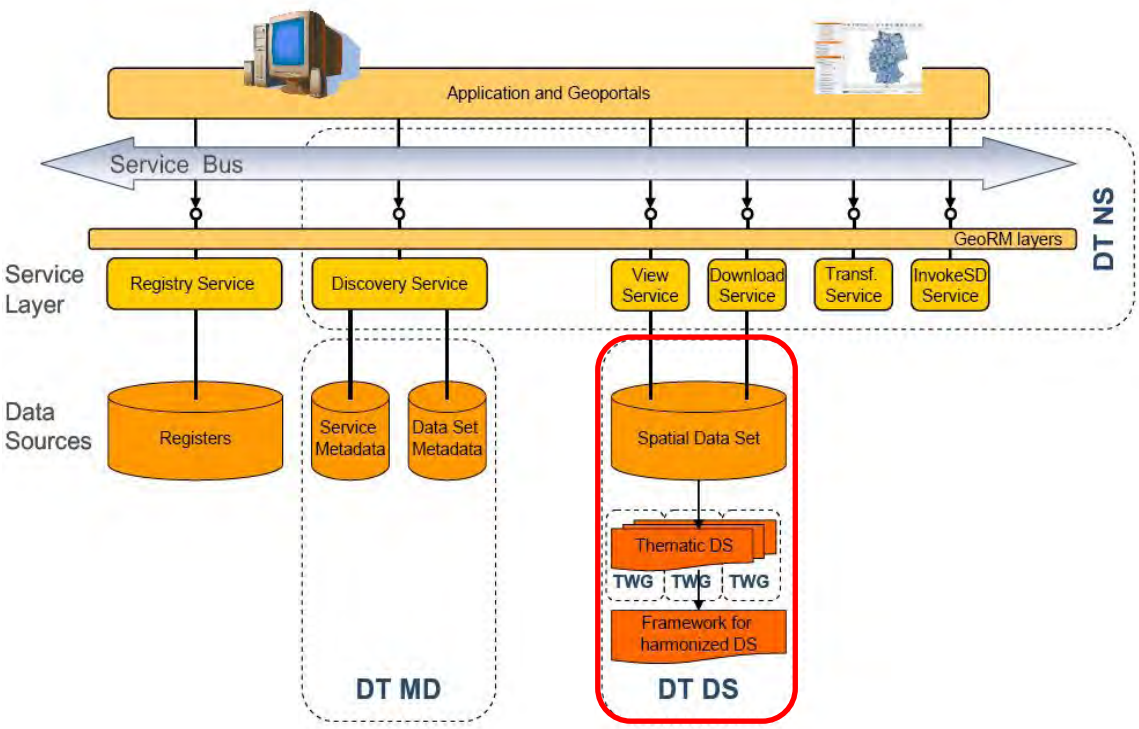


Fig. 12: INSPIRE technical architecture overview (Drafting Teams DS, MD, NS, 2007)

2.4.1 Data specifications

Data is the central component in the purpose of the INSPIRE Directive. The Data Specifications Drafting Team (DS DT) is responsible for setting up a framework that aims at keeping the data specifications of the different data themes coherent. This framework should summarise the methodology to be used for the data specifications development and provide a coherent set of requirements and recommendations to achieve interoperability. The framework is then used by the various Thematic Working Groups (TWGs) to develop the actual data specifications document for each INSPIRE data theme. There is one TWG assigned to each theme, composed of various experts in the field that the specific theme represents. As it can be seen in Figure 13 below, the Data Specifications development framework has as foundation four technical documents (i.e. deliverables D2.3, D2.5, D2.6, D2.7). The figure also illustrates the relationships from the point of view of the data specifications. The square boxes represent INSPIRE requirements documents, the cylinders represent registries, the arrows denote dependencies and the areas with dashed boundaries denote areas of responsibility (Drafting Team Data Specifications, 2008b). To date, data specifications have been published in a final version for Annex I themes (except ATS specifications), while the specifications for Annex II and III are still under development, reaching version 3.0 now, which is open for public consultation.

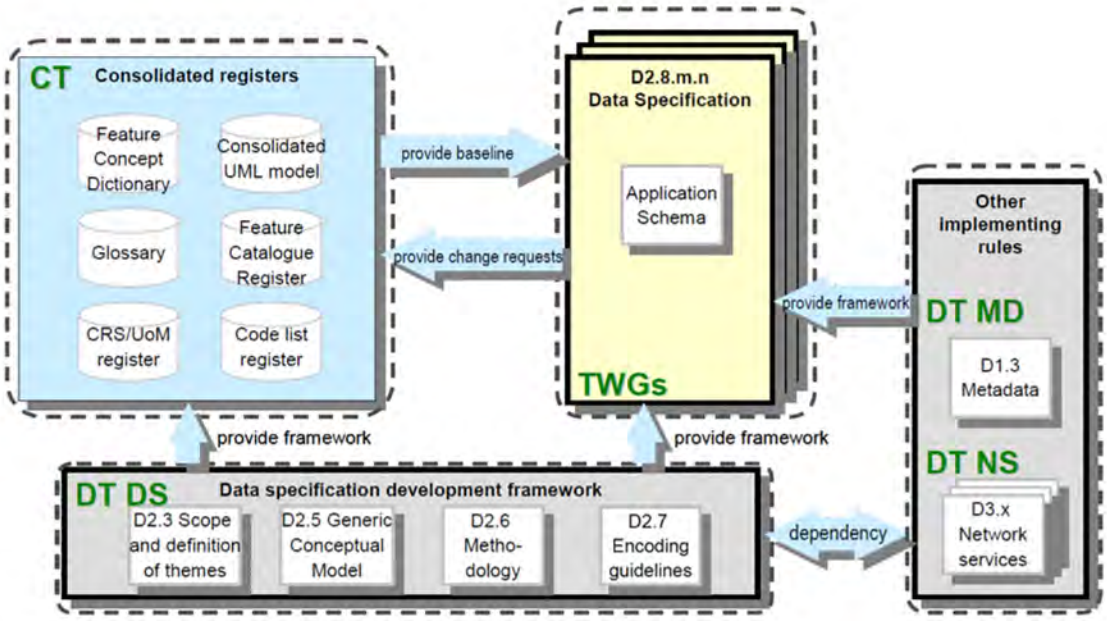


Fig. 13: Data Specifications development framework (Drafting Team 'Data Specifications', 2008)

- Definition of Annex Themes and Scope (D2.3) – describes in detail the spatial data themes defined in the Directive.
- Generic Conceptual Model (D2.5) – represents the central document of the framework and provides general guidelines for developing the data specifications, taking in consideration all kind of data harmonisation aspects. It is based on the ISO 19100 series of geographic information standards and is structured according to the identified components of data harmonisation.
- Methodology for the Development of Data Specifications (D2.6) – defines an iterative methodology enabling the transposition from user requirements to a data specification through a number of steps.
- Guidelines for the Encoding of Spatial Data – defines how spatial datasets can be encoded to enable transfer processes between the systems of the data providers in the Member States. Although it does not set mandatory encoding rule, it recommends GML (ISO 19136) as the default encoding method.

The INSPIRE Generic Conceptual Model identifies 20 different aspects relevant for data harmonisation or so called data interoperability components (see Figure 7), which have to be covered by the provisions in the INSPIRE Implementing Rules and technical guidelines (Drafting Team 'Data Specifications', 2010b). The INSPIRE interoperability components cover a wide range of different aspects of data harmonisation and interoperability in a spatial data infrastructure. There are data model related issues like rules for application schemas or spatial and temporal aspects, and issues related to data instances themselves like spatial reference systems, data quality and consistency. Also, aspects related to data capturing and maintenance as well as visualisation are covered (De Vries *et al.*, 2010).

The methodology of the development of data specification for the INSPIRE themes is aimed at a predictable and repeatable development process model. The approach is directed to identify issues

relevant for data interoperability, as early as possible in the process. It is important to identify as many issues as possible before the implementation takes place. Obviously, with data interoperability being such a complex aspect, it is not feasible to expect that every problem will be caught early in the process. Therefore, performing several iterations of implementing and testing is a normal and recommended practice (DT DS, 2008b). Figure 14 below clearly describes the process of developing INSPIRE data specifications.

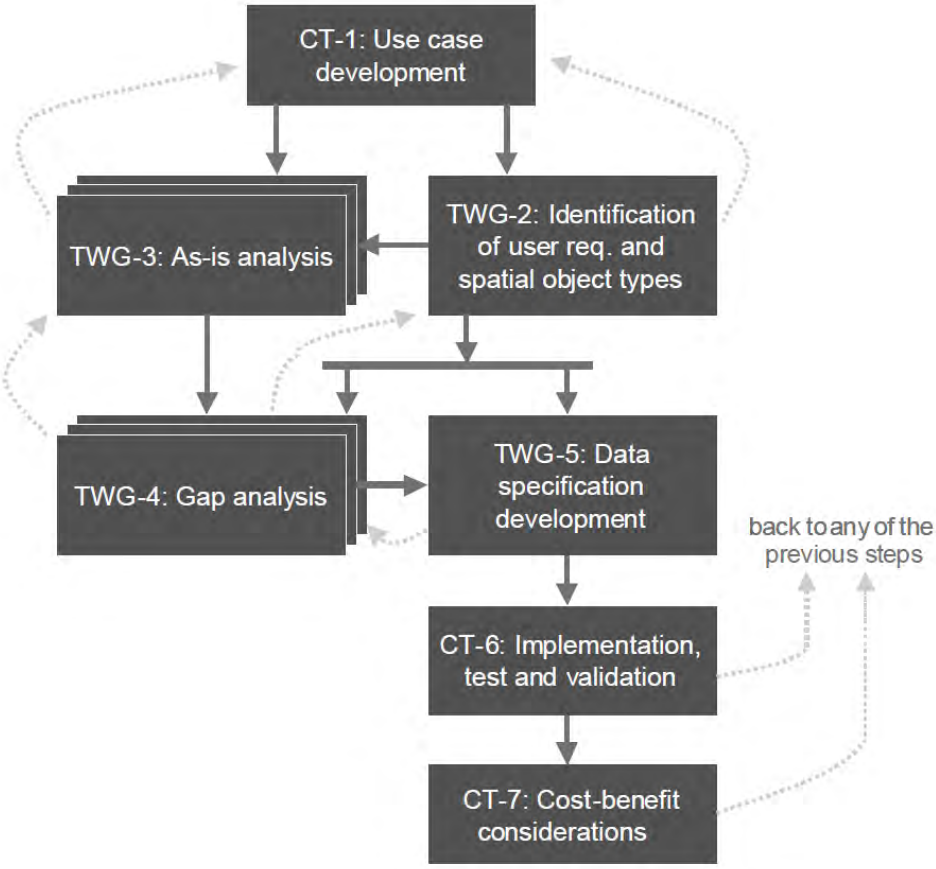


Fig.14: The process of developing INSPIRE data specifications (DT DS, 2008b)

In general, these steps are not carried out sequentially, but with a considerable overlap to allow for rapid feedback, which is intended to be given back to previous steps in every stage, where appropriate. The feedback process involves the drafting teams as well as the involved SDICs and LMOs, by giving recommendations to the Consolidation Team (CT), which is responsible for revision and testing of the proposed specifications. As can be observed the steps starting with CT are the responsibility of the Consolidation Team, and steps starting with TWG are the responsibility of the Thematic Working Group associated with the theme. In general, the Consolidation Team role is to coordinate and support DTs activities and keep a close link to the SDICs. The end result is a technical document (i.e. data specification), that provides the harmonisation requirements for each data theme. In practice, the approach that was followed so far was a '3 beta versions' approach, with the first version being internal to each TWG, the second being shared for comments with the Data Specifications Drafting Team plus all the other TWGs, and finally, the third version being released for public consultation.

Ultimately, according to the different data specifications, for every individual theme a conceptual schema is designed, that is capable of representing data from the various sources and providers that need to publish their data according to INSPIRE requirements. These schemas are developed using the Unified Modelling Language (UML), the chosen conceptual schema language in INSPIRE, and maintained in the Consolidated INSPIRE UML model. Other outputs of the DT DS is the INSPIRE Feature Concept Dictionary, which is used to manage the names, definitions, and descriptions of all spatial object types used in INSPIRE, but also registries, like the Feature Catalogue Register and the Code List Register, and a Glossary (see Figure 13).

2.4.2 Conformance testing

One of the very last interoperability component specified in the INSPIRE Generic Conceptual Model, and that has to obviously be implemented across the data specifications of all themes, refers to conformance, an aspect that will receive an increased attention in this research. The requirement states that *"every INSPIRE data specification shall specify a single conformance class per specification scope. Each conformance class shall reference an abstract test suite that tests all requirements specified in the data specification that are applicable to the specification scope of the conformance class"* (DT DS 2010b). This ensures that the harmonised dataset satisfies all the requirements of the data specifications, being a formal method to certify conformance.

Therefore, conformance testing is the type of testing to determine whether a product or a system meets some specified standard that has been developed for efficiency or interoperability. On the same principle, the objective of standardization in the field of geographic information cannot be completely achieved unless data and systems can be tested to determine whether they conform to the relevant standards. In this case, the relevant standard is ISO 19105 (Conformance and testing), part of the ISO/TC 211 standards for geographical information, which INSPIRE is as well based on.

The international standard ISO 19105 (Conformance and testing) defines two classes of conformance: class A and class B. Class A deals with conformance of the actual data specifications against which the testing refers, while class B concerns conformance of conformance clauses specified in ISO 19105. Conformance requirements can be mandatory, conditional, or optional. In terms of methodology used for conformance testing, it must be first mentioned that there are types of tests: basic tests and capability tests. Basic tests provide preliminary evidence that an implementation conforms, and they are usually used at the start of the conformance assessment process to establish whether or not it is appropriate to perform more thorough testing. Capability tests on the other hand, adopt a more thorough approach of testing, by checking the full range of conformance requirements specified in a standard. Furthermore, although conformance testing should be done automatically by executing software that implements the test, manual testing may be required when automated testing is too complex and/or human judgement is required. Manual testing should be used only for those circumstances where automatic testing is not viable (ISO, 2000).

The procedures to be followed during conformance testing are specified in abstract test suites (ATS), which represent the fundamental notion of conformance testing. An ATS defines a collection of test scenarios and/or test cases that are related or that cooperate with each other (http://www.thefreedictionary.com/). Moreover, it is a formal basis for deriving executable test suites (ETS). An ATS is independent of both the implementation and the values. Therefore, an ETS

results from instantiation of specific values for parameters in the ATS. Also, the ATS has a hierarchical structure, consisting of abstract test modules and abstract test cases. Abstract test cases represent the lowest level in the hierarchy, while abstract test modules are used to classify abstract test cases and other abstract test modules. An overview of such a hierarchical structure is given in Figure 15 below:

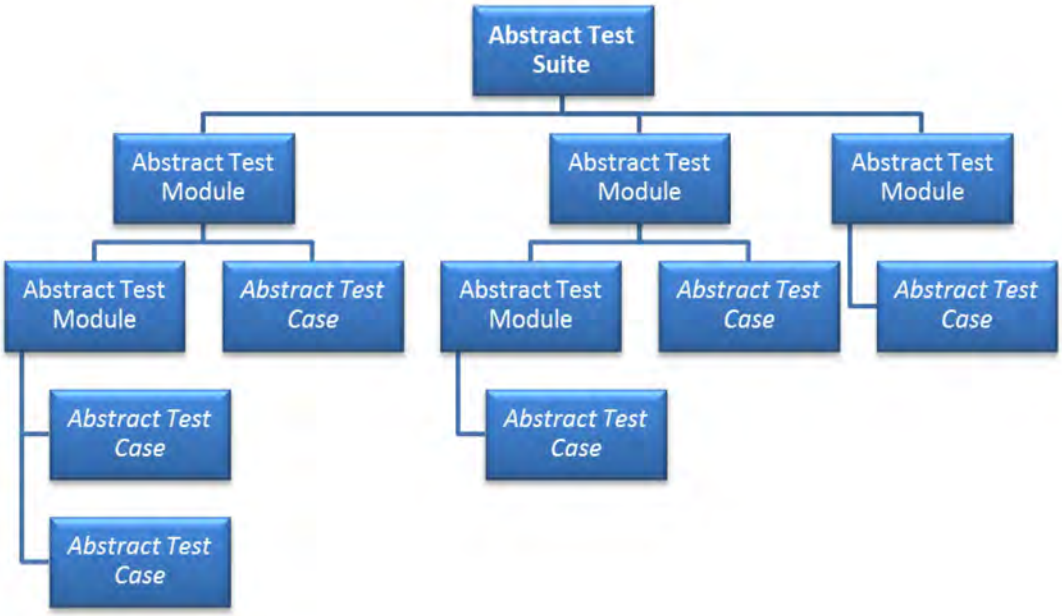
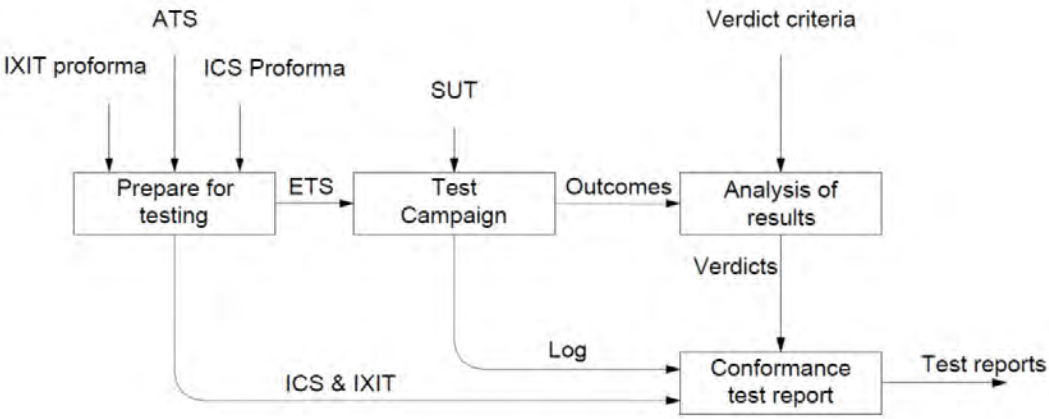


Fig. 15: The hierarchical structure of an abstract test suite (ATS) (adapted after ISO, 2000)

Finally, Figure 16 below presents the conformance assessment process that encompasses all conformance testing activities necessary to determine the conformance of an implementation to the relevant geographic information standards. The conformance assessment process involves four phases: preparation for testing, test campaign, analysis of results, and conformance test report.



IXIT – Implementation eXtra Information for Testing; ICS – Implementation Conformance Statement; SUT – System Under Test; ATS – Abstract Test Suite; ETS – Executable Test Suite

Fig. 16: Conformance assessment process overview (ISO, 2000)

According to the ISO geographic information standard there are two approaches to conformance testing:

- Verification testing – this uses methods that involve rigorous proofs of correctness in which the conformance can be rigorously demonstrated. This is the recommended approach whenever possible. However, the complexity of this approach is most of the times impractical for both technical and economic reasons.
- Falsification testing – this uses a method for detecting inconsistencies in conformance testing by developing a suite of specific tests that would test an implementation against a standard by focusing on critical areas of the standard. However, even if this suite of tests provides a positive verdict, there is no absolute assurance that the implementation conforms to the standard, because unlike the verification testing approach, falsification testing does not guarantee complete coverage of the content of a standard. As being a more practical approach, falsification testing has become an accepted way of doing conformance testing.

In INSPIRE, conformance testing against data specifications is still a blurred aspect that will have to be clarified as soon as possible by the people working in Data Specifications component. In addition to that, the INSPIRE Directive formulates a law to which all Member States have to comply according to the specified deadlines. This aspect complicates conformance to INSPIRE, and consequently the testing process that would validate that conformance, even more because not everything that is written in the INSPIRE documents is legally binding. There is a very fuzzy boundary between legal obligations and other technical guidelines that will inevitably affect the way conformance to INSPIRE have to be formulated. This aspect will be thoroughly tackled in Chapter 6 of this thesis.

2.5 Conclusions

As a chapter that has the scope of putting in context the theme of this research, but also provide an answer to the first research sub-question, this chapter has approached several aspects related to geographic information interoperability, which is the main driving factor behind the INSPIRE initiative. The definition of interoperability in INSPIRE, also considering what INSPIRE proposes to achieve, tends to shift the focus from the pure interaction of different systems to how the users of these systems can benefit by removing the obstacles usually encountered when trying to combine data from various sources. In this context, data harmonisation is necessary when other technical arrangements are not capable to cover the various gaps in interoperability of geographic information, and, therefore, changes in the underlying data structure are required. Giving this circumstance, it can be stated that data harmonisation is capable of removing many inconsistencies that stand in the way of a system interoperability.

Obviously, data harmonisation has to be based on certain agreements, a set of specifications, standards or legal acts. In geographic information there are several, internationally recognised, such agreements, like OGC and ISO/TC 211. Fortunately, the set of specifications that INSPIRE proposes is heavily based on these international standards, which actually protects any prior investment by

organisations in the OGC/ISO standards. Moreover, there are many voices in the industry that claim that INPSIRE could have never happened without the OGC standards.

Zooming out of all these details and trying to visualise the bigger picture of Spatial Data Infrastructures (SDI), it is clear that interoperability arrangements and data harmonisation go hand in hand in a SDI initiative. And this is what INSPIRE is after all, a European Spatial Data Infrastructure (ESDI). And, as any SDI, INSPIRE has its own specific arrangements and specifications for its various components, starting from the common data models to a standardised monitoring and reporting activity. As the scope of this research is centred on the data component of the INSPIRE ESDI, the next chapters of the thesis will give an in-depth and concrete analysis of what adhering to this ESDI means and what it involves, from the data component point of view.

3. Case study: INSPIRE Administrative Units – gap analysis

After clarifying the concepts of interoperability and data harmonisation, and how these are relevant in the context of the INSPIRE initiative, specifically to the Data Specifications component, the next step in the research is to prepare the implementation of INSPIRE requirements on a set of sample datasets. This chapter will describe the case study of the research, specifically, Section 3.1 will analyse the source data and its structure, while Section 3.2 will analyse the main parts of the INSPIRE data specifications and application schema for Administrative Units. The chapter will end with a concluding section which will compare the two data models and try to identify preliminary potential bottlenecks that may be encountered during the transformation.

3.1 Source data

With such an extensive variety of topical themes in INSPIRE, it is very difficult to test the implementation of several data specifications within the scope of this research. Specifications for one theme alone, are very extensive and well elaborated because it needs to take into account all interoperability components presented in the previous chapter. Therefore, the most feasible approach is to pick one theme and focus on an in-depth study on its data specifications and apply the transformation on the sample data. To ensure the most relevant results will be obtained, the preference is to select a theme from Annex I. This is due to the fact that at the present date (July, 2012), data specifications for Annex I have been published in a final version (except ATS specifications), while data specifications for Annex II and III are under public consultation at the moment.

The selection of the source data that will be used to derive INSPIRE conformant datasets, highly depends on availability. In this case, the preference was to go for British data to avoid the extra work load of translating, in this instance, Dutch datasets, although the availability of these would've been higher for obvious reasons. Therefore, the attention was aimed towards UK's Ordnance Survey data products. Ordnance Survey offers several samples of their commercial data products that cover a limited geographic extent, but it also offers free datasets through an open data scheme (i.e. OS OpenData). Based on these resources the choice was to go for the Administrative Units theme.

Fig. 17: Ordnance Survey and OS OpenData logos (http://www.ordnancesurvey.co.uk/)

The Administrative Units will be derived from the OS OpenData BoundaryLine™ product. This contains all levels of electoral and administrative boundaries, from districts, wards and civil parishes (or communities) up to parliamentary, assembly and European constituencies. The BoundaryLine™ is ideal for statistical analysis, or targeting various actions within a specific area. Although this dataset is available for the whole of the UK, the implementation will be based on England's extent only as

there are strong differences in the administrative geography between England, Scotland, Wales, and Northern Ireland. The data comes in ESRI Shapefile format and, as an update frequency, it is revised twice a year in May and October. All boundaries are captured and maintained against a generalised 1:10000 Scale Raster product. This dataset has been specifically designed to show the area of each administrative or electoral boundary.

The Boundary-Line dataset is available in a simplified, layered file structure, supplying the boundaries as individual files. For example, the county file contains only counties, and the unitary authorities file contains only unitary authorities, as is depicted in Figure 18 below.

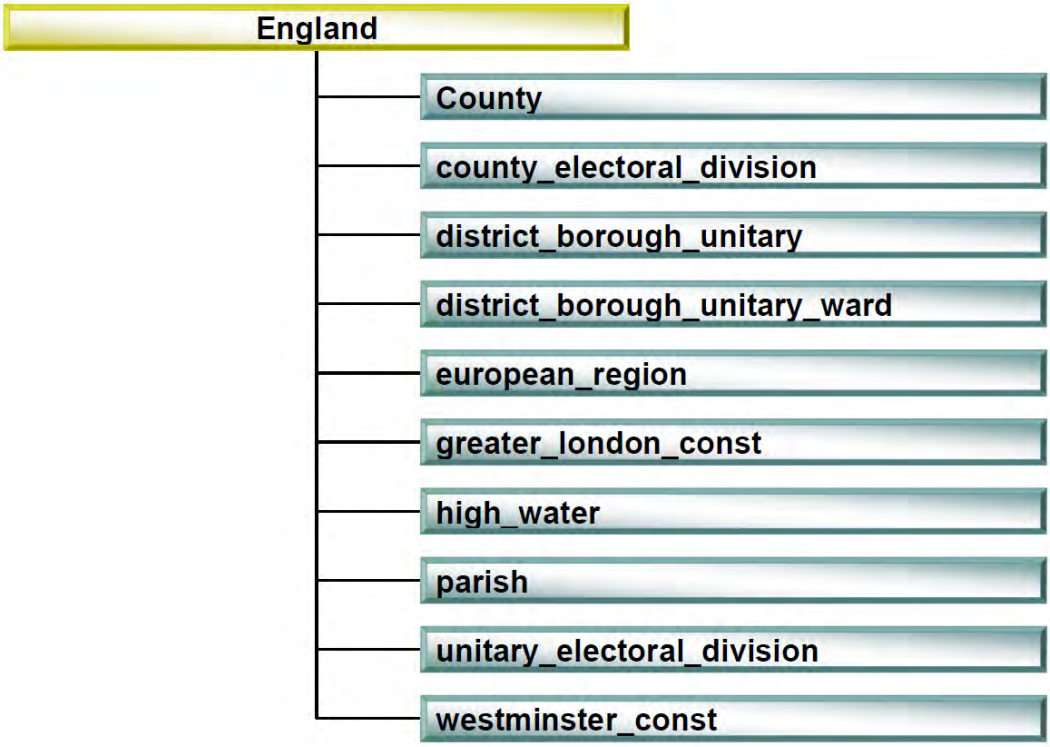


Fig. 18: Layered files of the Boundary-Line dataset (Ordnance Survey, 2012)

After a thorough inspection of the Boundary-Line dataset it became obvious that the administrative geography of the United Kingdom, which concerns itself with the hierarchy of areas relating to national and local government, is very complex, multi-layered and non-uniform. The hierarchy is not only complicated because of the different layers but the structure is also different in each constituent country of the UK. In addition, the boundaries of many of the layers in the hierarchy are subject to periodic or occasional change. For the scope of this project, the extent covered by the datasets that will be used as source data in the transformation, will only be limited to England. One reason for that is to avoid the total distinction between the administrative geography among the UK countries, but also because it will be much easier to track the processing of the transformation. Before making a decision about which layers of the Boundary-Line dataset will be used and if they will require any separate processing beforehand, there is a need to briefly clarify the structure of the administrative geography in England.

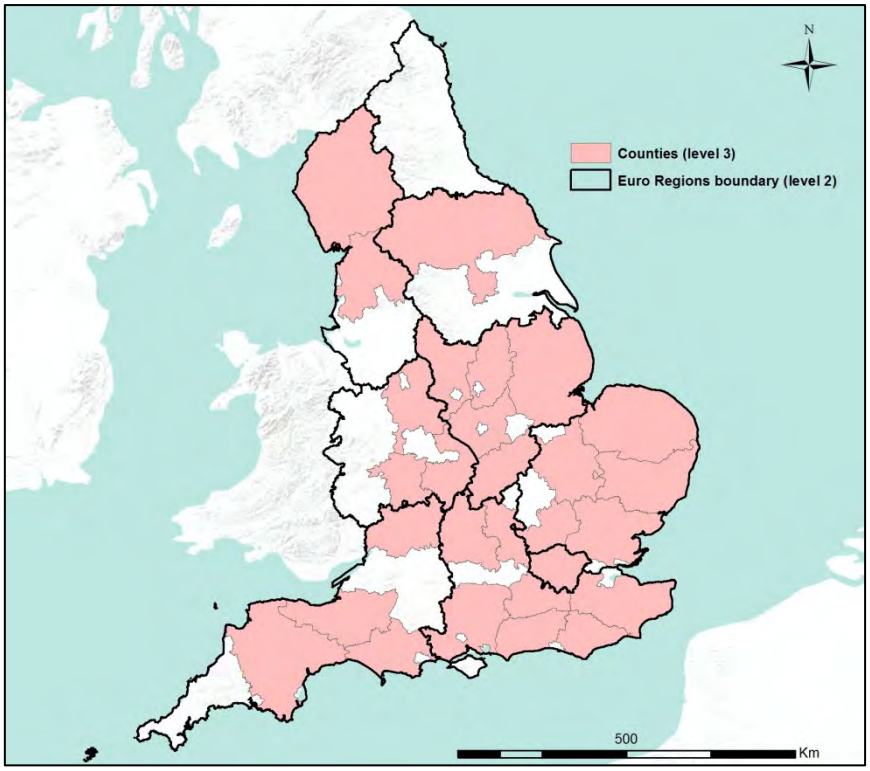


Fig. 19: Counties and Euro Regions of England

Figure 19 above illustrates the counties of England recognised as official administrative units, also being considered the 3rd level in the national administrative hierarchy. It can be clearly seen that they do not cover the entire extent of England, and the relationship with the upper level in the administrative hierarchy (i.e. Euro Regions), is very irregular.

Figure 20 below illustrates to a certain extent the different types of administrative units in England and the hierarchy between them. The colour of the boxes also depict a certain relation because, although theoretically on the same hierarchical level, Greater London, the (non-metropolitan) counties, and the metropolitan counties, were established at different periods in time and for different reasons, therefore they do not necessarily have a comparable administrative structure or administration body. On the other hand, the electoral wards/divisions are the key building block of England's administrative geography, being the units used to elect local government councillors in all the levels directly above them (http://www.ons.gov.uk). These lower level units are the most uniform administrative units in England, no matter what type of unit sits above them, thus they are represented with the same colour.

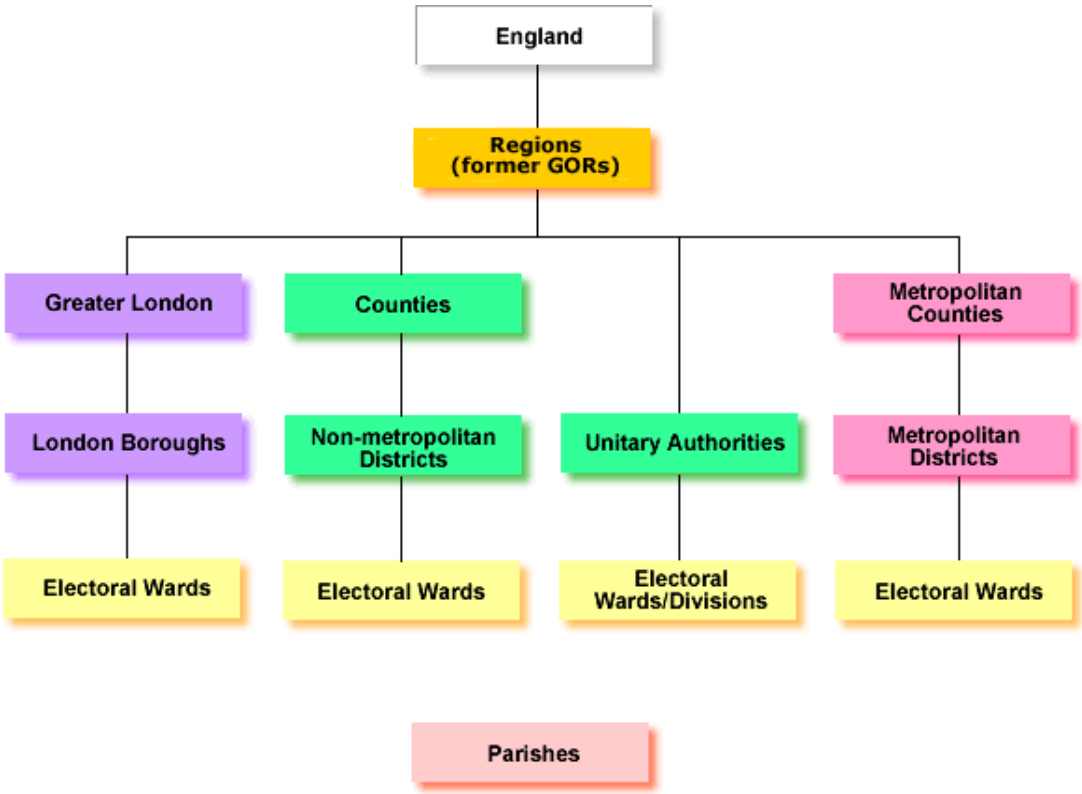


Fig. 20: Administrative geography of England (http://www.ons.gov.uk/ons/resources/england_tcm77-188215.gif)

Having that clarified, the highest level subdivisions of England are the Regions. The London Region (i.e. one of the Regions of England), known as Greater London, is further divided into the 33 London Boroughs, each being further divided into Electoral Wards. The other England Regions are made up of Metropolitan Counties, (non-metropolitan) Counties, and Unitary Authorities. The counties, both metropolitan and non-metropolitan are further divided into metropolitan and non-metropolitan districts, while the Unitary Authorities effectively combine the functions of counties and districts. All of these, are also further divided into Electoral Wards as it can be seen in Figure 20. Finally, Parishes are the smallest type of administrative areas in England, but again, not uniformly. Parishes, or town councils, exist for villages and small towns; they only rarely exist for communities within urban areas. They are even prevented from existing within Greater London. Another important aspect to Parishes is that although they are affected by the frequent boundary changes of the county Districts or Unitary Authorities in which they fall, they are not contiguous with Electoral Wards (http://www.ons.gov.uk). Many Parishes are a similar size to Wards, but some can contain several Wards, and Ward boundaries are not necessarily followed. Therefore, Parishes are not directly connected to the administrative geography hierarchy of England in the figure.

Giving that the layer files in the Boundary-Line dataset contain both administrative and electoral boundaries, not all of them are needed to reproduce the administrative hierarchy presented in Figure 20. Only the following layers are required:

- european_region – England Regions;
- county – (non-metropolitan) Counties;

- district_borough_unitary – Unitary Authorities, London Boroughs, Metropolitan (Metropolitan Counties can be derived) and Non-metropolitan Districts;
- district_borough_unitary_ward – Electoral Wards;
- unitary_electoral_division – Electoral Divisions (Wards equivalent in some Unitary Authorities);
- parish_region - Parishes;

The remaining layers are exclusively electoral boundaries and do not have any local, regional, or national administration significance. The 'high_water_polyline' layer represents the mean high water (springs) mark and enables to build and explicitly identify dry land areas within coastal polygons. All the mentioned layers are of polygon type with the exception of 'high_water_polyline' which is of line type.

The six layers that represent the administrative geography of the UK and that are going to be used in the transformation as source data, all have the same table structure as the one depicted in Table 1 below.

Table 1: Shapefile table structure of the Boundary-Line dataset

Title	Type	Description
FID*	Object ID	Shapefile unique identifier
SHAPE*	Geometry	Shapefile field depicting geometry type
NAME	Text	Name of the administrative unit
AREA_CODE	Text	Code allocated by OS to depict area type
DESCRIPTO	Text	Description of AREA_CODE
FILE_NAME	Text	Name of a file linking to the administrative unit
NUMBER	Double	Unknown
NUMBER0	Double	Unknown
POLYGON_ID	Double	Unique OS identifier across the whole BL dataset
UNIT_ID	Double	Unknown
CODE	Text	Unique national identifier by Office for National Statistics (ONS)
HECTARES	Double	Area of polygon to a 0.001 hectares precision
AREA	Double	Non-inland area, including foreshore and tidal water
TYPE_CODE	Text	OS code to depict unit type
DESCRIPT0	Text	Description of TYPE_CODE
TYPE_COD0	Text	Unknown (empty in all layers)
DESCRIPT1	Text	Unknown (empty in all layers)

* - Shapefile automatically created fields

To have a better understanding on how these fields are populated, Table 2 below depicts an example field record for the 'county_region' layer.

Table 2: Example field record for the 'county_region' layer

FID*	11
SHAPE*	Polygon
NAME	Lancashire County
AREA_CODE	CTY
DESCRIPTO	County

FILE_NAME	LANCASHIRE_COUNTY
NUMBER	31
NUMBER0	1145
POLYGON_ID	121912
UNIT_ID	5156
CODE	E10000017
HECTARES	308299.429
AREA	17846.579
TYPE_CODE	AA
DESCRIPT0	CIVIL ADMINISTRATION AREA
TYPE_COD0	
DESCRIPT1	

* - Shapefile automatically created fields

Most attributes are easily understandable and do not require further clarifications. It may be only relevant to mention that three of the attributes, described as 'Unknown' in Table 1, didn't have any description even in the technical document of the dataset. The last two columns labelled as 'Unknown' were empty across all layers being possible that they represent extra attributes created for further administrative unit coding types in the future.

3.2 Administrative Units data specifications

This section will dive a bit deeper in the INSPIRE data specifications for the Administrative Units theme in order to understand the target data model and how the source data will be able to relate to that model.

INSPIRE defines Administrative Units as *"units of administration, dividing area where Member States have and/or exercise jurisdictional rights, local, regional and national governance, separated by administrative boundaries"* (European Commission, 2007). The definition has been interpreted not to include administrative units such as census districts, post office regions and other sector-specific regions (Drafting Team 'Data Specifications', 2008a). However, the Administrative Units theme contains reference to the Nomenclature of Territorial Units for Statistics (NUTS). The NUTS is an EU geocode standard for referencing the subdivisions of countries for statistical purposes at European level, and at national level there are Local Administrative Units (LAU). The NUTS and LAU are defined for all Member States, so the relation to the administrative levels in the context of INSPIRE might prove very relevant. There are three NUTS levels (1, 2, and 3) and two LAU levels (1 and 2). LAUs are basic components of the NUTS regions, and it is important to mention that NUTS regions do not necessarily match with the national administrative units.

The European wide harmonised data of all the administrative units and their boundaries is extremely relevant for any kind of cross-boundary spatial analysis, important in operations and management, and also in geo-referencing of thematic/statistical information, based on linkage to NUTS/LAU units. It is also relevant to mention that the data specifications for the Administrative Units theme have been derived from the specifications of EuroGeographics EuroBoundaryMap (EBM™) product. EuroGeographics represents 56 National Mapping and Cadastral Agencies from 44 countries across

Europe. According to the Administrative Units Thematic Working Group (2010), this theme will mainly support the following high level cases:

- Filtering data – users are able to select other datasets based on the spatial relation to an administrative unit.
- Linking thematic information – easy and rapid access to comparable thematic information across the entire European Union.
- Disaster management – administrative units that are affected by an environmental phenomenon can be selected.
- Boundary-based analysis – verification of data of thematic features located at the boundaries of administrative units. This also covers the aspect of edge-matching.
- Discovery of unit related data – search catalogues to discover available data sets with respect to administrative unit geometry, name, or code.

Below is an overview of the Administrative Units package and referenced packages. The 'AdministrativeUnit' spatial object type uses the 'GeographicalNames' type from the Geographical Names package, and it also refers to the Base Types package. The Base Types package contains elements not defined in the foundation schemas, but which are required in INSPIRE (e.g. Identifier). This package is referenced by all other INSPIRE application schemas. Finally, as already explained, there is a link to 'NUTSRegion' of Annex III Statistical Units theme. This is referenced as a preliminary specification, since the Annex themes are still under development.

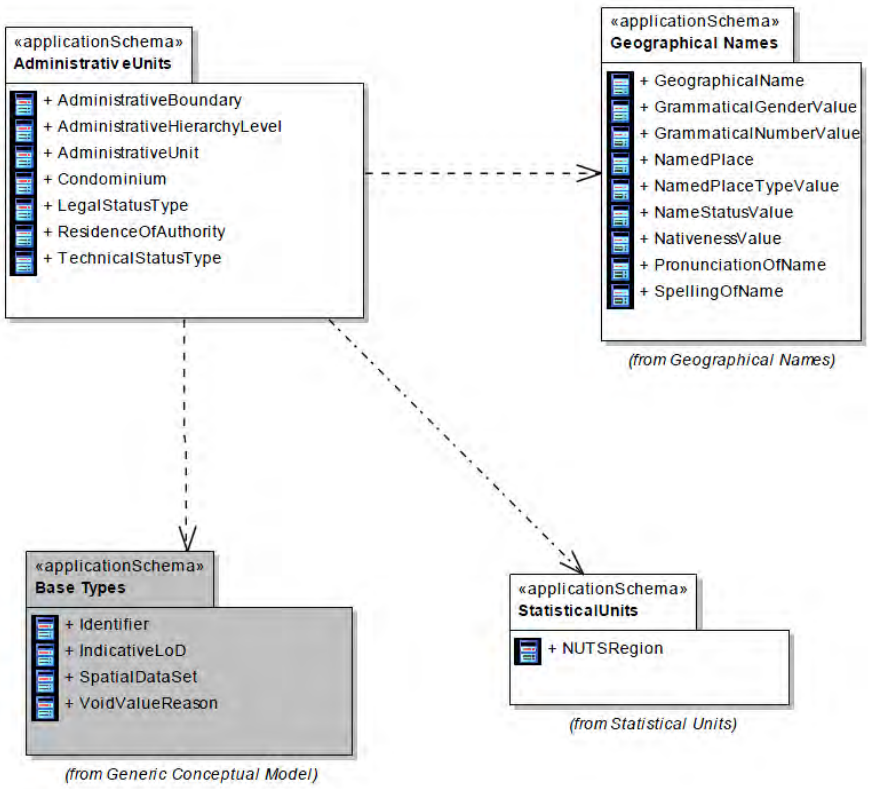


Fig. 21: Administrative Units package overview (INSPIRE TWG AU, 2010)

The next page illustrates the entire application schema in UML of the Administrative Units theme.

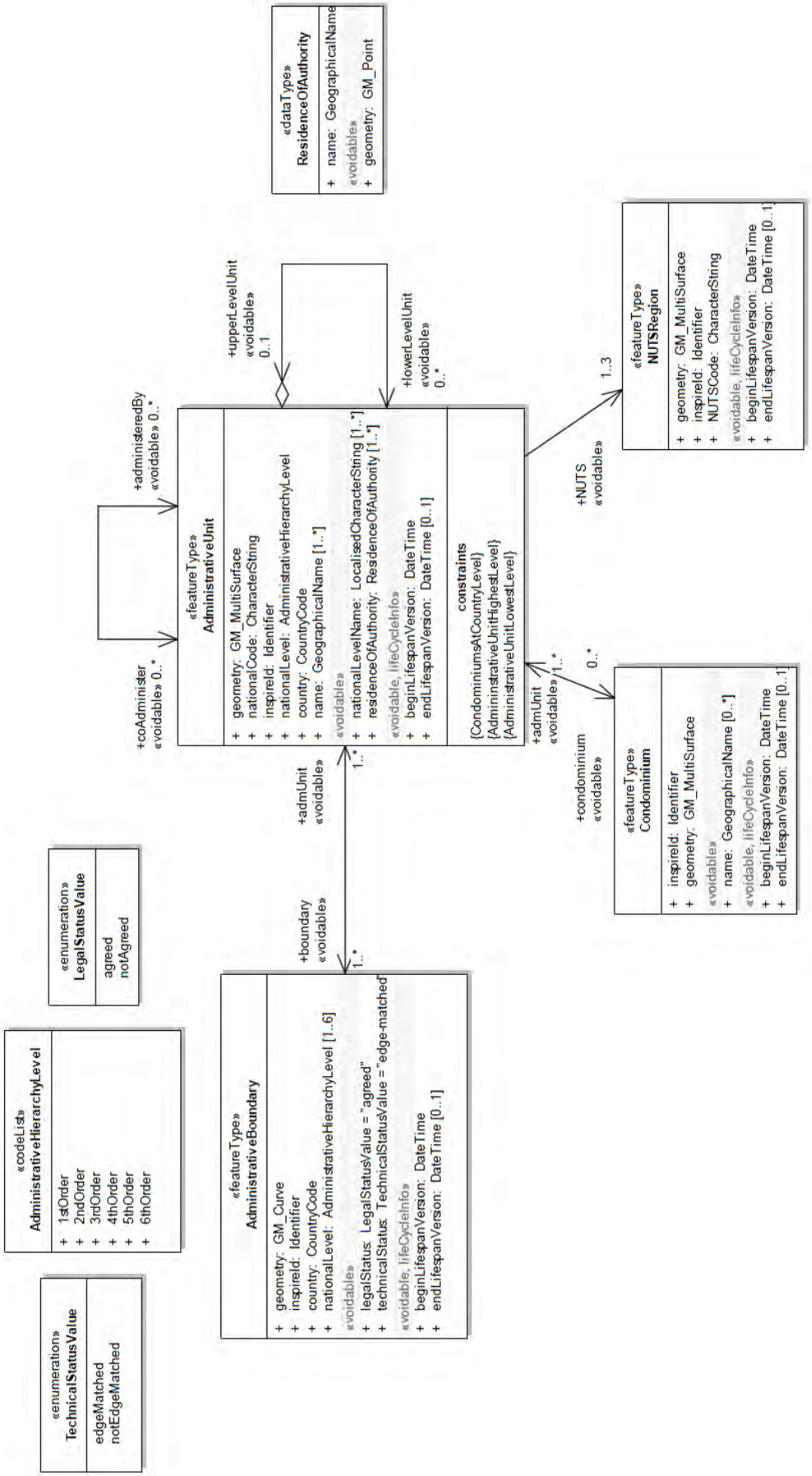


Fig. 22: Administrative Units UML application schema (INSPIRE TWG AU, 2010)

The UML diagrams offer a very efficient way to visualise the main elements of the specifications and their relationships. To fully understand the content of the data model, especially for non-UML specialists, INSPIRE has in place a Feature Catalogue which includes the definition of the spatial object types, attributes, and relationships. This will be very useful when data providers across Europe will have to use people with strong thematic expertise, to aid in the mapping process, but who are not familiar with UML. The catalogue covers of course all INSPIRE themes. Those parts of the catalogue that cover the Administrative Units theme are attached in Annex A of this document.

The main spatial object of the application schema is the 'AdministrativeUnit' (see Figure 23). It represents administrative units at all level of administrative hierarchy, and each single unit belongs to exactly one level of the respective national hierarchy. Because administrative units from a higher level do not necessarily always aggregate the units at lower levels (as it was already explained in the case of England's administrative units), semantic relations between the units of subsequent upper and lower levels were introduced and documented in the mandatory 'nationalLevel' attribute. A relevant detail to keep in mind is that because administrative units can consist of inlands and exclaves, or other non-inland territories (i.e. islands), their geometric representation is of multi-polygon type (i.e. GM_MultiSurface).

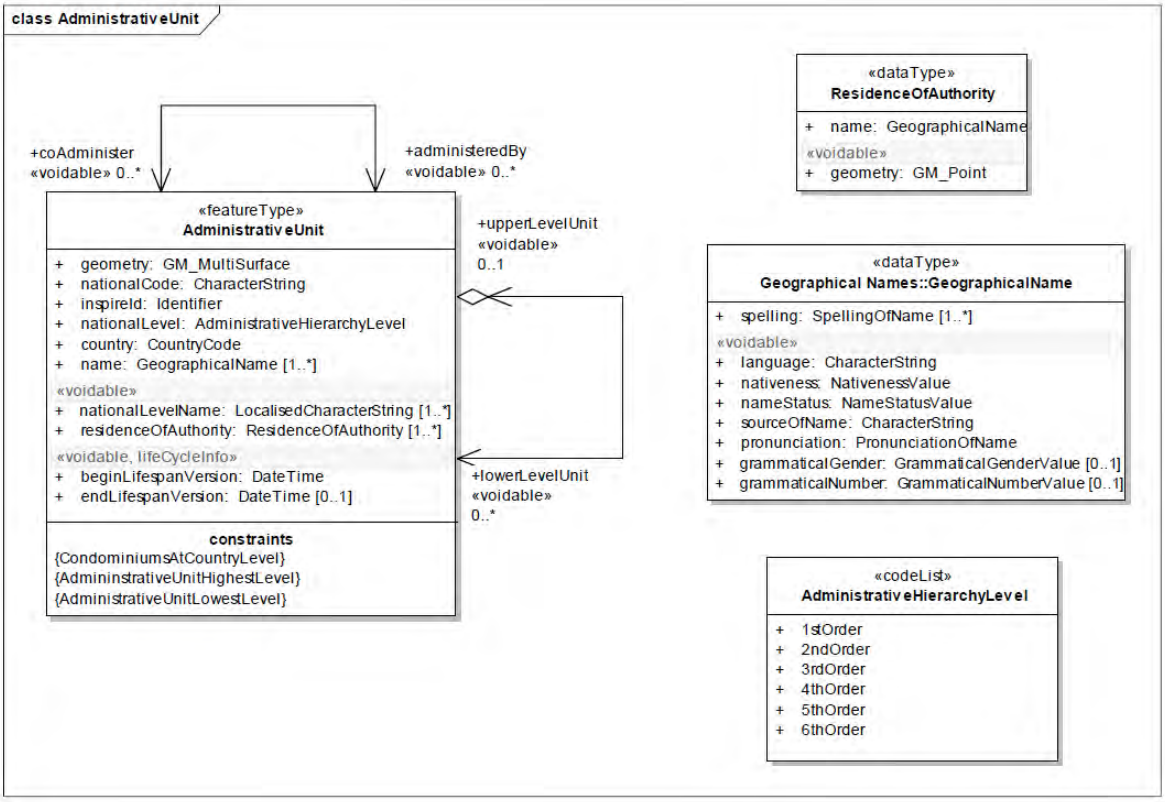


Fig. 23: The 'AdministrativeUnit' spatial object and its references (INSPIRE TWG AU, 2010)

For each administrative unit it is possible to specify the location of the authority residence, along with its name which is mandatory. To clarify, all attributes, or spatial objects under the 'voidable' tag are optional but there is a requirement to specify why these are not specified. There are two reasons that are grouped in a code list (i.e. VoidValueReason) that can be used:

- Unpopulated – the characteristic is not part of the source dataset maintained by the data provider, although it might exist in the real world. The characteristic receives this value for all objects in the dataset.
- Unknown – the value is not known and not computable by the data provider, however a correct value may exist. The value is applied on an object-by-object basis.

The name of the administrative unit is of 'GeographicalName' type, and, as already mentioned, it links to the 'GeographicalName' object of the Geographical Names application schema, belonging to the Geographical Names theme.

The second spatial object type in the application schema is the 'AdministrativeBoundary'. This object represents the boundaries between neighbouring administrative units (see Figure 24).

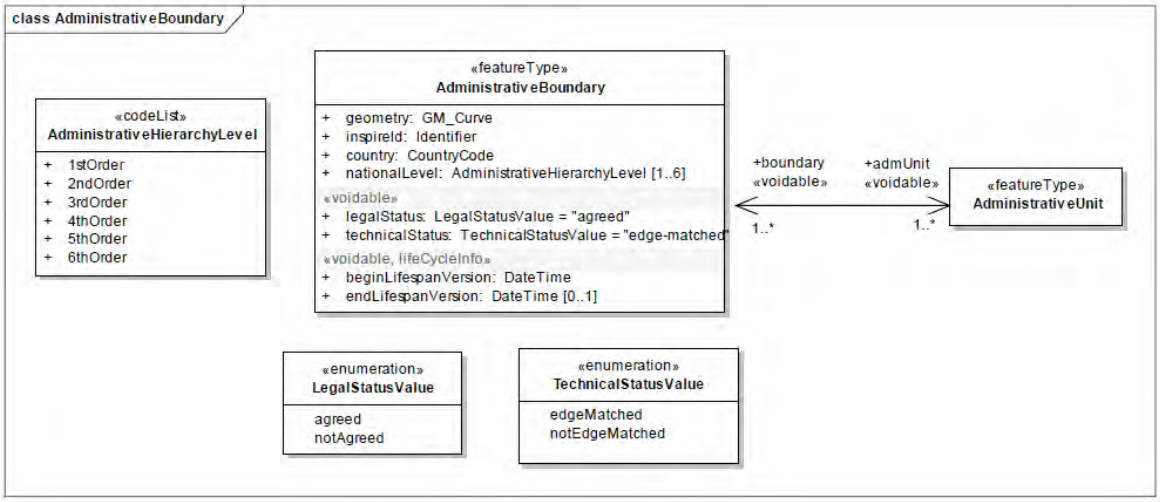


Fig. 24: The 'AdministrativeBoundary' spatial object and its references (INSPIRE TWG AU, 2010)

Besides providing information about administrative division, the boundaries also have a legal and a technical property. The legal status refers to the political consensus between the administrative units (including country boundaries), while the technical status refers to edge-matching issues that may arise after the harmonisation to the INSPIRE standard. The boundaries also have a 'nationalLevel' attribute that plays the same role as in the 'AdministrativeUnit' spatial object. As it would be expected there is an association, voidable though, with the 'AdministrativeUnit' spatial object type to support topological and semantic relationships.

Another important aspect relates to the specifications related to the geometric structure of the boundary features. There are two alternative geometric structures for boundary features:

- Flat model – aims to support download of mass information collected in the form of data files.
    - Each boundary feature corresponds to the curve established between the two significant nodes of topological graph established with respect to the lowest level of national administrative hierarchy.
    - Each single boundary feature might refer one or more hierarchical levels (e.g. part of 2nd level boundary and part of 3rd level boundary).

    - Each administrative unit feature will associate to all touching boundary features.
    - Each boundary feature will associate to all touching administrative units from all the levels of administrative hierarchy.
- Multi-layer model – better suited for View Services, as well as selective download of small portion of feature instances using direct access download services.
    - Each single boundary feature represents exactly one administrative boundary established at certain level of national administrative hierarchy.
    - The boundary features between administrative units are identified at each hierarchical level.
    - The geometry of boundary features corresponds to the entire line of demarcation determined for the administrative unit at the same level as the boundary level.
    - Each administrative unit feature will associate only to boundaries established at the level corresponding to the level of the respective administrative unit.
    - Each administrative boundary feature will associate only to administrative units that are separated by the respective boundary and have the same administrative level as the level of the boundary.

Another spatial object of the Administrative Units application schema is the 'Condominium'. In very rare cases, some administrative units would fall under this category if the area is administered by two or more countries. The relation between the condominium and the governing countries is represented by a link between the 'AdministrativeUnit' and 'Condominium' spatial object types as depicted in Figure 25.

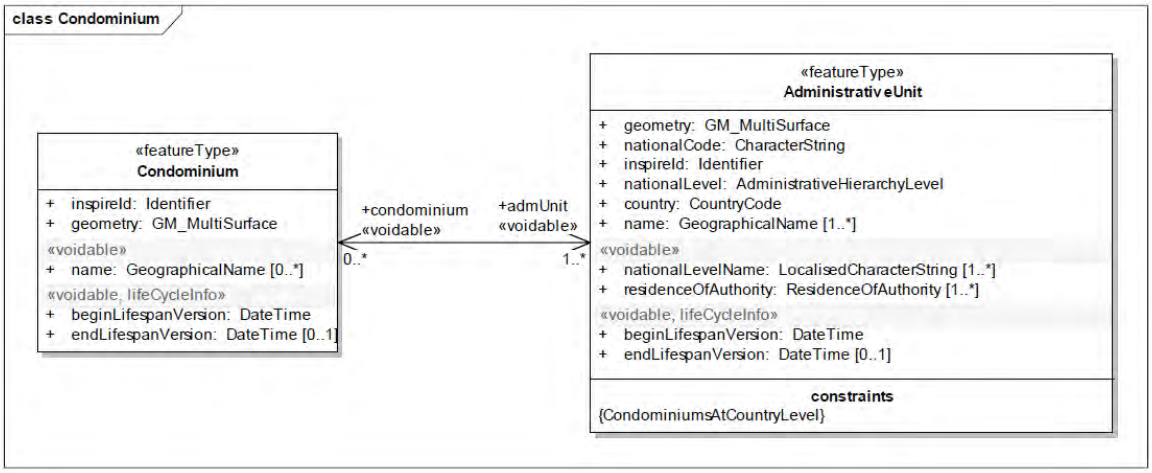


Fig. 25: The 'Condominium' spatial object type (INSPIRE TWG AU, 2010)

Giving the study case of this project and the source data used for the transformation, there will be no administrative units that might fall under the condominium type. It might be worthy to mention for the record though that in 2001, the British Government held discussions with Spain with a view for a proposal for joint sovereignty to the people of Gibraltar. This proposal was rejected by a high majority of Gibraltarians during a referendum in 2002. Gibraltar is a British overseas territory, but it governs its own affairs, though some powers, such as defence and foreign relations, remain the responsibility of the UK Government.

Finally, the last spatial object type that appears in the Administrative Units application schema is the 'NUTSRegion' of the Statistical Units application schema. As already explained, this is an association of each administrative unit to a NUTS unit for statistical purposes at European level. One administrative unit can be associated to up to 3 NUTS regions as these are classified on three hierarchical levels (i.e. NUTS1, NUTS2, NUTS3), and, for instance, an administrative unit on the lowest hierarchical level can be part of all NUTS levels. This association is also a voidable because Statistical Units is part of Annex III themes, which do not have a stable version of the data specifications.

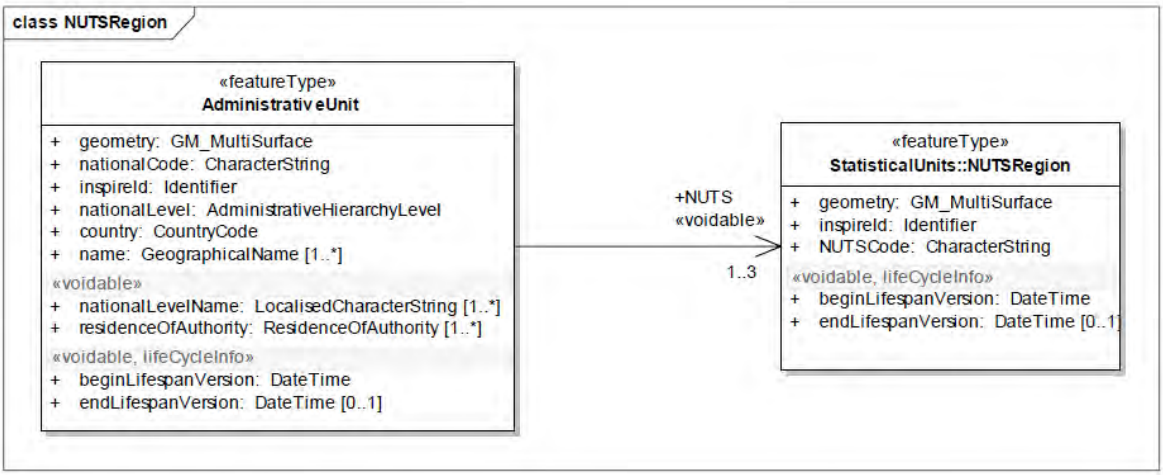


Fig. 26: The 'NUTSRegion' spatial object type (INSPIRE TWG AU, 2010)

The illustrated specifications for 'NUTSRegion' in Figure 26 are preliminary and they contain a geometry type, the INSPIRE identifier of course and the NUTS code for the statistical unit.

3.3 Conclusions

When comparing the source and target data model in this pre-transformation stage, there are already few inconsistencies that can be highlighted which raise many question marks, which might or might not be solved during the actual transformation.

First of all, regarding the actual spatial objects, the target data model contains four spatial objects: 'AdministrativeUnit', 'AdministrativeBoundary', 'Condominium', and 'NUTSRegion'. In that respect, the source data correlates with only one spatial object in the target data model, the 'AdministrativeUnit'. It is very true that boundaries can be derived, but giving the requirements for the topology structure of the boundaries, there might be some issues that will be encountered. Also, establishing the association between administrative units and their respective boundaries might be problematic for the same reason. In addition, although voidable, the 'NUTSRegion' spatial object doesn't have any correspondent in the source data, so it is clear that it will not be possible to implement it if no additional dataset, containing the NUTS regions for England, will be brought in the transformation. It also must be noted though, that 'NUTSRegion' is actually an association to a spatial object in another INSPIRE theme (i.e. Statistical Units – Annex III). Finally, 'Condominium' is

applicable only in extraordinary circumstances, and these do not happen to exist in England's administrative affairs.

Going further and analysing the preliminary gaps between the attributes of the two data models, there are several observations that need to be made. First of all, the identifiers required by the INSPIRE data model do not exist in the source data, so these will have to be intuitively created and making sure they would be unique at European level. Besides that, there are many other attributes that will have to be created during the transformation that do not exist in the source data. Some of these are constant (being values maintained in code lists or enumerations) across an entire target data model spatial object, like 'country' (i.e. UK), or constant across an entire source shapefile (i.e. within one administrative level), like 'nationalLevel' (i.e. each national level being comprised of at least one shapefile of the source data); therefore, it should be relatively easy to implement them. In other cases though, these might be very hard or even impossible to derive, like 'residenceOfAuthority', which in fact appears to be an entire separate dataset containing point locations of the responsible authority for an administrative unit. Lifecycle information is also not available in the source data, as well as administrative units nesting information, which is needed to derive the 'upperLevelUnit' and 'lowerLevelUnit' associations.

Although many of the enumerated issues can be skipped because the specific attributes or associations are voidable, therefore not compulsory, this is not the scope in this research. This also raises another aspect that refers to all the voidable elements. Some countries might be very tempted to take the 'shortcut' and just provide what is compulsory, but then it is a question of what is the real value of the data if all these elements are missing. It is also true that INSPIRE is aimed specifically at data providers, which will probably have access to more complex data models of the source data that will provide much more information towards the transformation than a dataset which is part of an Open Data scheme, like the case is with the source data used in this case study.

All the points made in this section will be referred back to in the next chapters, but especially in the concluding section of the data transformation chapter where a post-transformation review of all these issues will have to be made.

4. Data transformation

This chapter will review several tools in Section 4.1, that provide data harmonisation functionality and that claim to offer solutions to transform datasets according to the INSPIRE requirements. Ultimately one of those tools will be used for the entire Extract-Transform-Load (ETL) chain using the British administrative units as source data and aiming to output INSPIRE conformant Administrative Units data. Section 4.2 will briefly present the approach to be used to produce INSPIRE compliant data, and therefore highlighting key requirements needed from the software that will be used. These requirements are mainly based on the findings from the literature review chapter. The choice of the tool will be primarily reflected upon in Section 4.3, taking into account the pros and cons of the functionality, with a more in-depth and concrete analysis to follow in the next chapter after the actual implementation.

4.1 Tools overview

Spatial ETL tools provide the data processing functionality of traditional ETL software, but with a primary focus on the ability to manage spatial data. Spatial data commonly consists of a geographic element that physically places the different features on the earth surface, and the related attribute data. Therefore, spatial ETL transformations are often described as being either geometric transformations (transformations of the geographic element), or attribute transformations (transformations of the related attribute data).

Common geometric transformations that should be present in any spatial ETL relate to:

- Re-projection – the ability to convert spatial data between one coordinate system to another.
- Spatial transformation – the ability to model spatial interactions.
- Topological transformations – the ability to create topological relationships between different datasets.
- Data clean-up – the removal of errors from a dataset.
- Data merging – the process of bringing together multiple datasets into a common framework.
- Quality assessment – comparison of multiple datasets for verification and quality assurance purposes.
- Data translation – conversion between different data formats and data models.

In the following sections several software solutions that are capable to provide the above functionality are presented.

4.1.1 Safe Software FME

Safe Software Inc. is a Canadian based software development and consulting services company that focuses on managing the exchange of both spatial and non-spatial data between GIS applications and/or relational databases with differing file formats and structures. The company's core data translation and transformation product is FME (i.e. Feature Manipulation Engine), an integrated collection of spatial ETL tools that provides functionality similar to a traditional ETL tool, but also has

additional capability to manage spatial datasets and the complexity that comes with such data, like feature geometries, attribute tables and coordinate systems. FME is able to convert data between over 200 spatial data formats, including ESRI, MicroStation, AutoCAD, raster and database formats. There is also a server edition of FME, which further expands the functionality, enabling the distribution and delivery of data in a variety of formats and services such as WFS, WMS, SOAP, KML, as well as basic streaming.

Fig. 27: Safe Software and FME logos (http://www.safe.com/)

As it would be expected, FME is capable to perform schema mapping and schema transformation processes that are required to transform datasets according to the INSPIRE data specifications. In addition, with the release of the latest version of the software (i.e. FME 2012), the support for INSPIRE transformations and other related processes has been increased significantly, making FME a really strong solution for data providers that have to publish their data to INSPIRE standards in the coming years. There also has been a noticeable increase in promoting the FME solution for INSPIRE among several consultancies across Europe.

4.1.2 Snowflake Software GO Publisher / GO Loader

Snowflake Software is a UK-based company with a strong background in 'Commercial off the Shelf' (COTS) software to facilitate data exchange. Snowflake's solutions are based on open standards and are mainly focused on loading, translating and publishing XML and GML data. They do not claim to offer an 'out of the box' spatial ETL tool, but the combination of their two main products, GO Publisher and GO Loader, perform like one. Besides offering data exchange solutions for the data providers, Snowflake Software also specialises in the aviation and defence domains, specifically in the information systems of the air traffic management infrastructure, and support for the system integrators in defence and intelligence. They also have a close collaboration with Ordnance Survey, not only through the functionality of their main tools which are widely used and recommended by Ordnance Survey, but also by developing a specialised tool specifically tailored for one of the OS data products, the OS MasterMap (i.e. highly detailed topographic data), called OS MasterMAp Viewer. Being so much reliant on XML data formats, Snowflake solutions undoubtedly are of a very relevant fit for INSPIRE requirements. The company is also registered as a SDIC organisation, has participated in previous INSPIRE pilot projects, and has a member of their staff working with one of the TWGs.

Fig. 28: Snowflake Software GO Loader / GO Publisher logos (http://www.snowflakesoftware.com/)

Regarding the software tools, GO Loader is a complete solution for loading and integrating XML data formats into a wide variety of spatial databases like Oracle, PostGIS or SQL Server. The main advantage of GO loader is that it is 'schema aware', by adapting itself to support the dataset based on its XML schema, so there are no concerns regarding any specific new data models not being supported. Also, it helps the user model and store the data in a manner tailored to the individual or business needs, making it more valuable than just a simple loading tool. On the other hand, GO Publisher ensures the easy translation and publication of information from a variety of data models or database models, being as well 'schema aware', allowing the user to map the data from the source database model to a target XML model. Therefore, the combination of GO Loader and GO Publisher perform as a regular spatial ETL tool, but mainly based on XML data formats and database models.

4.1.3 Esri ArcGIS for INSPIRE

ESRI is a company that doesn't require a very thorough presentation. Anyone that had any tangency with the geographic information domain at all would be familiar with Esri, the incontestable world leader in GIS software solutions and services. Their main product, ArcGIS, which comes in various flavours for desktop, server, mobile, and more recently online, have had an extension developed in the last years, called ArcGIS for INSPIRE. It is claiming to be advancing and enhancing Esri's proven ArcGIS technology base for use by EU Member States and other EU constituent organisations to meet the INSPIRE requirements.

In contrast to the other solutions presents so far, ArcGIS for INSPIRE aims to bring ArcGIS into full compliance with INSPIRE rules and requirements, not only relating to data transformation but also in terms of creation and maintenance of metadata and network services. In short, the ArcGIS for INSPIRE product includes:

- An ArcGIS for Desktop extension to create and maintain INSPIRE-compliant spatial data and metadata.
- An ArcGIS for Server extension to serve INSPIRE-compliant view and download services.
- INSPIRE-compliant geo-database templates for extracting, transforming, and loading geospatial information from existing databases into INSPIRE-compliant geo-databases.
- The open source Esri Geoportal Server including enhancements for cataloguing and publishing INSPIRE-compliant metadata and providing INSPIRE-compliant discovery services.

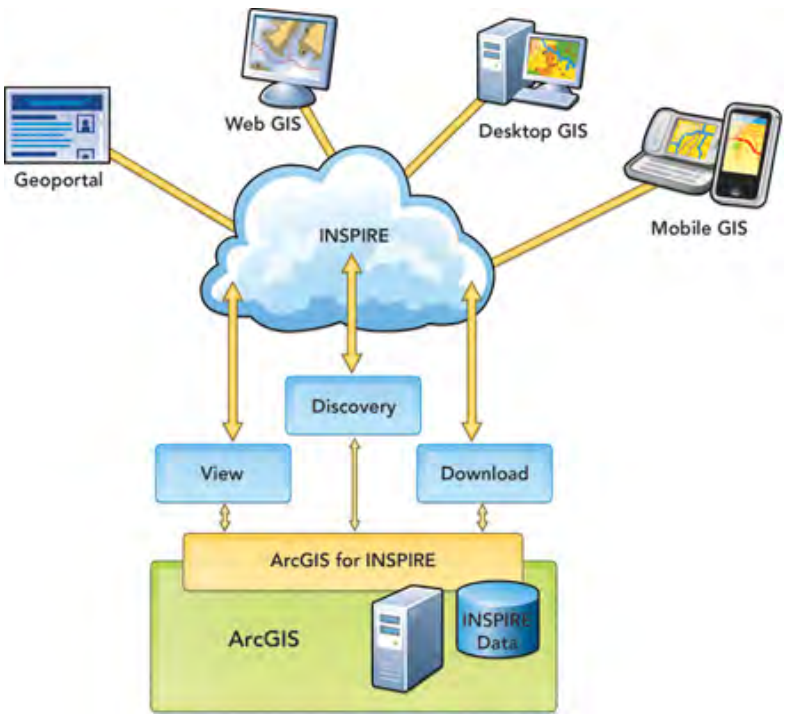


Fig. 29: Relationship of ArcGIS for INSPIRE to ArcGIS platform (http://resources.arcgis.com/sites/default/files/images/ArcGIS%20for%20INSPIRE/INSPIRE_network_dgm(4).jpg)

Although it seems like an all-in-one solution, the current state and availability of the product tends to be rather an unknown variable. It wasn't possible to obtain a trial version for the scope of this research in order to perform some tests, and even if it would've been, ArcGIS for INSPIRE seems to require a complete, and rather complicated for this level, ArcGIS platform deployment, including Desktop and Server extensions.

4.1.4 The HUMBOLDT Framework

HUMBOLDT is a European funded research project coming out of the GMES (Global Monitoring and Security) program. The HUMBOLDT project consortium consists of a heterogeneous partner structure, covering universities and research institutes, national mapping agencies, GI users and other GI industry stakeholders. The scope of HUMBOLDT was to contribute to the implementation of a European Spatial Data Infrastructure (ESDI) that integrates all the diversity of spatial data available from the multitude of European organisations. The most important part of the project is to enable these organisations to document and harmonise their spatial information. The HUMBOLDT framework has three main components:

- HUMBOLDT tools – contain applications that can be used by spatial data experts to create conceptual schemas and to define transformation between them.
- HUMBOLDT services – define both cross-cutting service components that allow the collaborative usage across organisations, countries and domains.
- Individual transformation services – each of those are able to tackle a specific harmonisation problem on specific given scenarios like border security, urban planning, forestry, protected areas, and many others.

The HUMBOLDT software is open source, thus available for free. Some examples are the Model Editor (UML editor for the creation of UML application schemas), Alignment Editor also known as HALE (helps define conceptual schema transformations), Workflow Editor, Mediator Service (a proxy service that executes transformation chains to provide harmonised spatial data), and others.

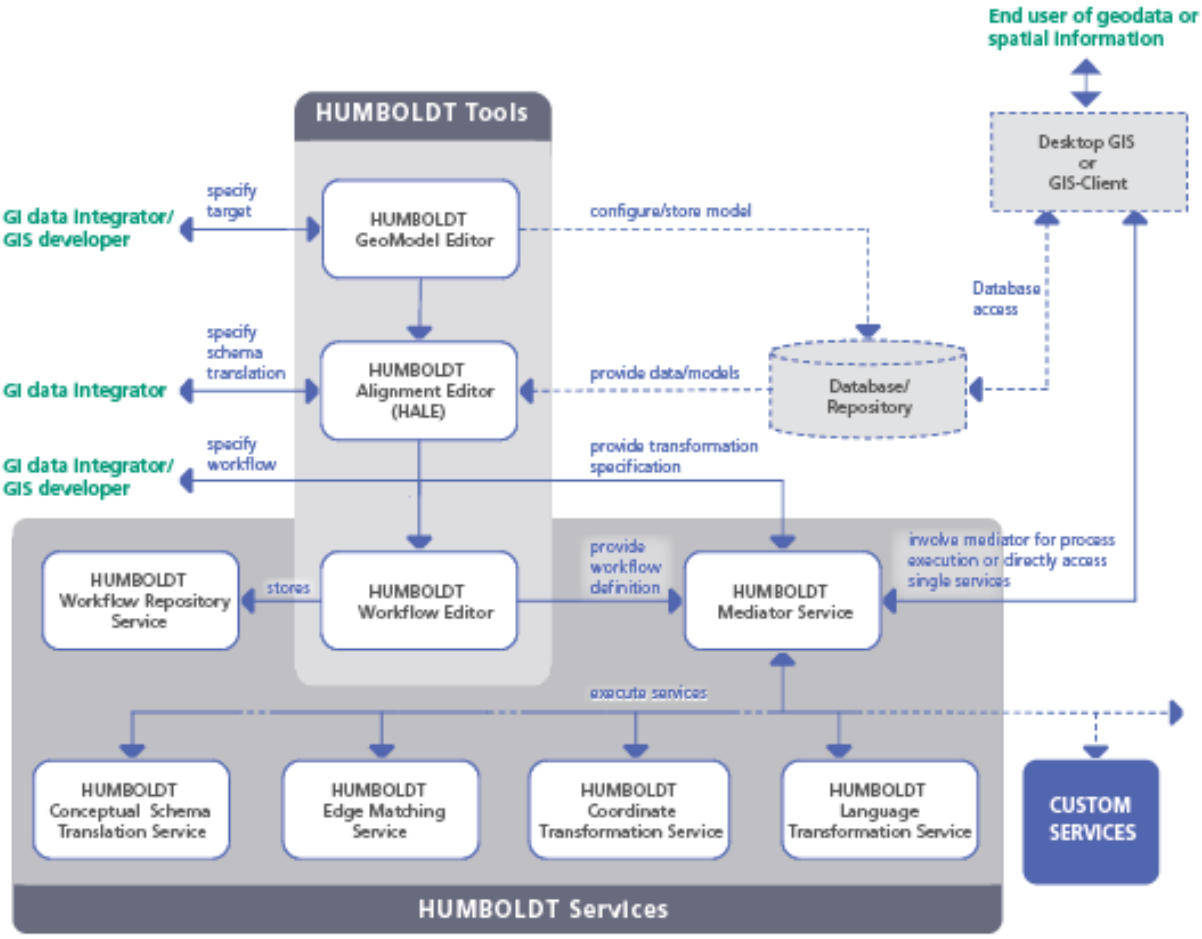


Fig. 30: HUMBOLDT Framework for data harmonisation (Fichtinger et al., 2011)

The HUMBOLDT Framework can be considered a response at European level to the methodology, rules, and guidelines developed by INSPIRE, to facilitate the creation of a ESDI, and thus test these guidelines in the project. However, although the current functionality of the HUMBOLDT framework provides a good start-up, further development is needed in order to be able to deal successfully with INSPIRE requirements and to be able to compete as a solution with the other commercial products. Being more of a research project, funded by the EU and with limited budget of course, it is hard to estimate whether HUMBOLDT is going to take the step further, given that the project was ended in 2010 according to the initial time planning.

4.1.5 Other FOSS solutions

When it comes to currently available open source spatial ETL, worthy to mention are Spatial Data Integrator and GeoKettle. Both are based on 'regular' ETL suites, Talend, respectively Pentaho. These tools are considered to be rather entry-level spatial ETLs and their functionality is far from what FME is offering for example. In addition, the lack of support for GML 3.2.1, which is the required version for INSPIRE, among other things, makes it hard to believe that these solutions are ready to support

INSPIRE requirements just yet. Nonetheless, there are attempts within the OSGeo (Open Source Geospatial Foundation) community to investigate and put together a fully FOSS stack for INSPIRE implementation (http://code.google.com/p/inspire-foss/).

4.2 Transformation approach

Based on the literature study in Chapter 2, specifically Section 2.1.2, where the concept of data harmonisation was tackled, this section will briefly present the implementation of that concept in the case study of this research.

To adhere to INSPIRE data requirements there is clearly the need to change the structure of the source data to the structure specified in the target (i.e. INSPIRE) data model. This process involves all the identified steps during Chapter 2, namely, schema matching, schema mapping, and ultimately schema transformation, but also CRS transformation, assuring geometry and topology consistency, and encoding the data to a specified format. Although it would be desirable to have the entire process automated, this is hardly possible, especially in INSPIRE where every source data (i.e. from the different Member States) is different. In this sense, schema mapping is probably the most work-intensive part of the process and the hardest to automate.

With schema mapping it is usually best to have domain experts involved that have a thorough knowledge of the source data. Depending on the complexity of the source or target data structure, the expertise of such people can be crucial and can save a lot of time. As source and target data models for the case study of this research were presented in the previous chapter, it became clear that there is certain information in the target model that is missing from the source data. It is a matter of understanding the descriptive information of the source data, but in some cases this may also be heavily dependent on the capabilities of the software that will be used for the transformation.

In addition to that, the source data will most definitely need pre-processing. As the Administrative Units application schema includes a separate spatial object for administrative boundaries (i.e. polyline geometry), these will have to be derived separately as the source data only contains the administrative units in polygon geometry. Not to mention the topology model that INSPIRE requires for some of the spatial objects. To perform these pre-processing operations it may be necessary to use additional tools, like traditional desktop GIS packages that offer a wide variety of tools, although some operations might also be possible with the capabilities of a spatial database. In this case, source data will be initially stored in a spatial database from where it will be fetched by the transformation tool. Therefore, a spatial database can also be added to the list of required software.

Ultimately, the format the transformed data will be encoded in must be considered in the transformation software capabilities. INSPIRE data specifications recommends using a specific GML version standard, which means this has to be supported and easy to publish into in the transformation tool that will be used.

Taking into consideration all the aspects presented above, the next section will provide a brief comparison between the software tool presented, and which one of them will be eventually used to implement the INSPIRE requirements, as well as other software that will be needed.

4.3 Discussion

Based on the individual software tools review that was given in Section 4.1 and on the approach to transformation described in Section 4.2, the table below summarises a comparison of some of the core functionality needed to transform the source data according to INSPIRE requirements, but also availability and dependency on other extensions.

Table 3: Software tools comparison

	Availability for this research	Schema mapping	Coordinate transformation	Spatial database support	Encoding to GML 3.2.1	Status
FME	Yes	Yes	Yes	Yes	Yes	Commercial
GO Publisher / GO Loader	Yes	Yes	Yes	Yes, but limited	Yes	Commercial
ArcGIS for INSPIRE	No	Yes, but requires FME extension	Yes	Yes, but requires ArcSDE	Yes, but requires FME extension	Commercial and also dependent on other extensions
HALE	Yes	Yes	Yes	No	Yes	Open source, and still under development

Judging solely on these points it would seem that the most appropriate decision will be to use either FME or GO Publisher / GO Loader. After a more detailed analysis of the two products it has been decided to use primarily use GO Publisher as the transformation tool. At this pre-transformation stage this was a relatively personal choice as both products seemed capable to deliver the required results, but also because GO Publisher come from a vendor relatively newer on the market that attracted a higher interest. Finally, the fact that the entire functionality of GO Publisher is based on the 'data exchange' idea by only supporting publishing to XML/GML formats from an existing database, and is not so 'format conversion' centric as FME is, made it, at this stage, to seem much more suitable in the INSPIRE context. Of course, the performance of GO Publisher and whether is really a stable solution to publish data according to INSPIRE requirements, will be thoroughly discussed and analysed after the implementation in the next chapter.

On the other hand, GO Loader will not be necessarily required as it is primarily used to load GML data into a spatial database, which in this case means loading the produced INSPIRE-compliant GML data back in the database. For the scope of this research this is not required, therefore, once the GML data is produced, the transformation will be considered as finalised.

Besides GO Publisher, other software that will be used comprises of a spatial database and a desktop GIS package that will be used for pre-processing. For the spatial database solution, Oracle Express Edition is used, which is an entry-level, small-footprint database. The current version (i.e. 11g) is

based on the Oracle Database 11g Release 2 code base. It is free to develop, deploy, and distribute, and it is also worth to mention that its spatial capabilities are a cut-down version of the commercial Oracle Spatial extension (i.e. Oracle Locator) (http://www.oracle.com). In addition, Oracle SQL Developer is used as well (also downloadable for free from Oracle's website). This is an integrated development environment (IDE) for working with SQL in Oracle databases. It is a database administration and query tool that provides a single consistent interface for various databases. It uses the Java Development Kit (JDK) to run on a machine, and maybe very important to note is that GO Publisher is also relying on the same development kit. One of the reasons to use Oracle therefore, was, and this came as a recommendation from Snowflake Software as well, that GO Publisher works best with Oracle spatial databases, although SQL Server is also supported. PostgreSQL is supported as well, but not its spatial extension, PostGIS.

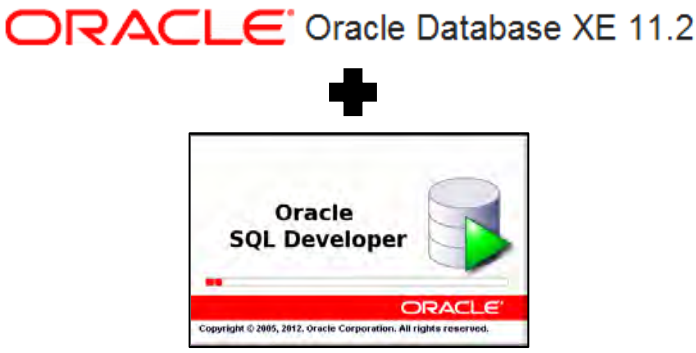


Fig. 31: Oracle Database Express Edition 11g & Oracle SQL Developer logos (http://www.oracle.com)

Finally, a third-party extension for Oracle SQL Developer is used, GeoRaptor, which adds functionality for viewing and managing spatial data, including a shapefile loader plugin. In regard to the desktop GIS package, ArcGIS 10 for Desktop is used.

With this chapter preliminary clarifying the choice behind the software stack, especially in regard to GO Publisher, which will stand as the basis for the data transformation, it provides the first part of the answer to the third research sub-question. The next chapter will be describing the actual implementation and the process of producing INSPIRE compliant datasets, and linked with this chapter, will fully answer the third research sub-question.

5. Producing INSPIRE compliant data

In this chapter the process of achieving INSPIRE-compliant datasets will be presented by using one of the software tools reviewed. The choice was to use Snowflake Software's solutions consisting of GO GO Publisher, mainly because this tools have a much more narrowed scope and focus solely on data exchange using open standards, in this case these being XML and GML. The second choice would have been Safe Software FME (for a parallel transformation comparison) but due to the time limit allocated for this research it will not be possible go through the data transformation with both solutions.

The software versions used in the implementation are 2.1 for GO Publisher, which is the current version at the moment of writing (July 2012). It's important to mention that GO Publisher comes in three flavours:

- GO Publisher Desktop – configuration of the transformations required for desktop and enterprise data publishing.
- GO Publisher Agent – enterprise bulk data publishing service via an ordering system.
- GO Publisher WFS – direct access web service for data exchange over the internet using WFS and RESTful services.

Although a more complete solution for the scope of INSPIRE would be GO Publisher WFS, the version used for the implementation in this research is the Desktop one, as the focus here is solely on data transformation and doesn't include INSPIRE network services.

During the course of the chapter, Section 5.1 will deal with necessary pre-processing steps, while Section 5.2 will summarise the actual transformation procedure in GO Publisher for the spatial objects of the Administrative Units application schema. Finally, Section 5.3 will review the transformation process and will analyse the results and the encountered problems, as well as limitations of software or data specifications.

5.1 Data preparation

The very first step in the data preparation is to clip the Boundary-Line dataset to England's extent only, because as already discussed there is a significant difference between the administrative geographies of England, Scotland, Northern Ireland and Wales. This is achieved via several 'select and export' operations in ArcMap. Although the Oracle database supports several spatial functions that might be capable to perform the same operations, it has been decided to do all the pre-processing externally, and load in the database the ready-to-transform data. Below are the resulted layers that will be used in the transformation:

- Entire country – 1st level (this dataset is derived from the Regions layer by dissolving the boundaries)

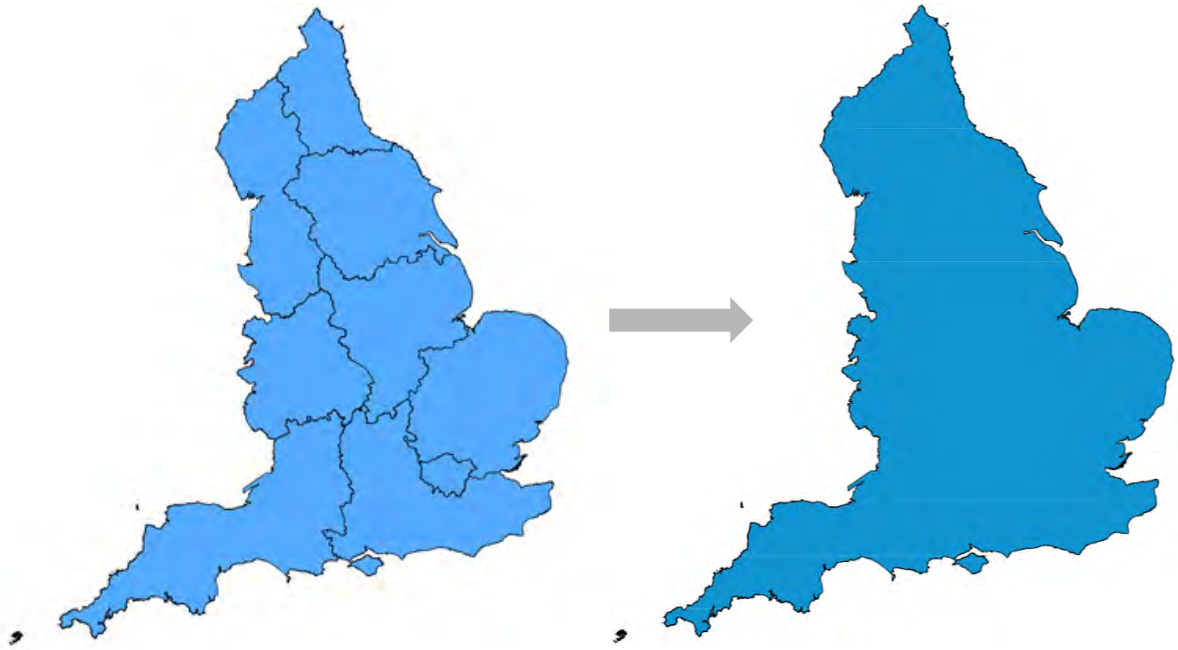


Fig. 32: Dissolve applied on Regions to obtain the 1st level of administrative level (country)

- Regions – 2nd level
- Counties – 3rd level
- Districts, London Boroughs, Unitary Authorities – 4th level
- Wards – 5th level (formed by joining 'district_borough_unitary_ward' and 'unitary_electoral_division', as these together cover the entire extent of England and are considered to be part of the same administrative hierarchy level)

Fig. 33: Joining of the two layers with administrative units the 5th level in the administrative hierarchy

All the layers above will be used for a direct transformation to derive the 'AdministrativeUnit' spatial object from the INSPIRE application schema. As it has been already identified in the gap analysis

chapter, the issue encountered here is that the INSPIRE application schema also requires the boundaries of the administrative units as a separate spatial object, of a polyline type geometry.

To derive the boundaries from the administrative units' layers, there are two approaches depending on which geometric structure for the boundary features will be used, as described in Section 3.2 of Chapter 3. For this case study, the 'flat model' will be used, which means only boundaries of the administrative units at the lowest level in the administrative hierarchy are needed. To achieve the required structure, the administrative units at level 5 (i.e. wards) are first transformed to line features in ArcMap with the 'Feature to Line' tool, and then the output is processed with the 'Unsplit Line' tool. The reason the 'Unsplit Line' tool is used is that although the first operation should split the boundary features only at intersections (i.e. between two nodes in the topological graph), it occurs that, randomly, some boundary features (i.e. edges) are split along without an occurring intersection. Figure 34 below demonstrates the difference. When taking the 'multi-layer model' approach for the geometry structure of the boundary features, the same procedure is applied on all administrative units layers, resulting in a topological boundary graph for each administrative hierarchy level.

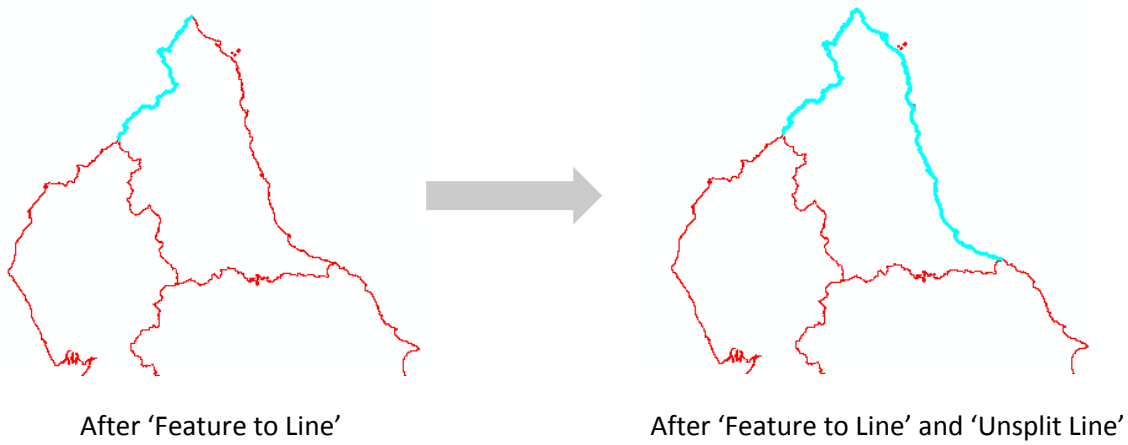


Fig. 34: Processing the boundary topology graph

Once all datasets are processed and boundaries available, there is one more aspect that needs to be taken care of before the transformation. GO Publisher needs to fetch the data from a database in the form of tables with a geometry column. It doesn't read a variety of file formats, like FME does for example. Although this might seem as limited functionality, in real-life situations any organisation would be storing and managing their data inside a database. GO Publisher can connect to various versions of Oracle, SQL Server, PostgreSQL, MS Access, MS Access, JNDI (Java Naming and Directory Interface), JDBC (Java Database Connectivity). A very important observation is that PostgreSQL is supported but its spatial extension, PostGIS, is not. Support will be added in future versions, so the only reliable databases left that have a decent support for spatial data are Oracle and SQL Server. This is something to keep in mind, especially for those organisations or individual users that might not be storing their data in one of these databases.

As already mentioned in the previous chapter a third party shapefile loader plugin (i.e. GeoRaptor) will be used to upload the datasets in the Oracle database. Based on the original table structure of the layers, presented previously in Table 1, the following alterations have been applied on all tables:

Table 4: Attributes alterations during import to database

Original dataset table	Oracle XE table upon import
NAME	NAME
AREA_CODE	AREA_CODE
DESCRIPTO	renamed to DESCRIPTION
FILE_NAME	FILE_NAME
NUMBER	filtered out
NUMBER0	filtered out
POLYGON_ID	POLYGON_ID
UNIT_ID	renamed to ADMIN_UNIT_ID
CODE	renamed to UNIT_CODE
HECTARES	renamed to AREA
AREA	renamed to NONINLAND_AREA
TYPE_CODE	renamed to AREA_TYPE_CODE
DESCRIPT0	renamed to DESCRIPTION1
TYPE_COD0	filtered out
DESCRIPT1	filtered out

It goes without saying that a geometry column will be added as well to the table that will store the geometry of each individual feature (i.e. SDO.GEOMETRY), multi-polygon for administrative units and line for boundaries.

5.2 Data transformation

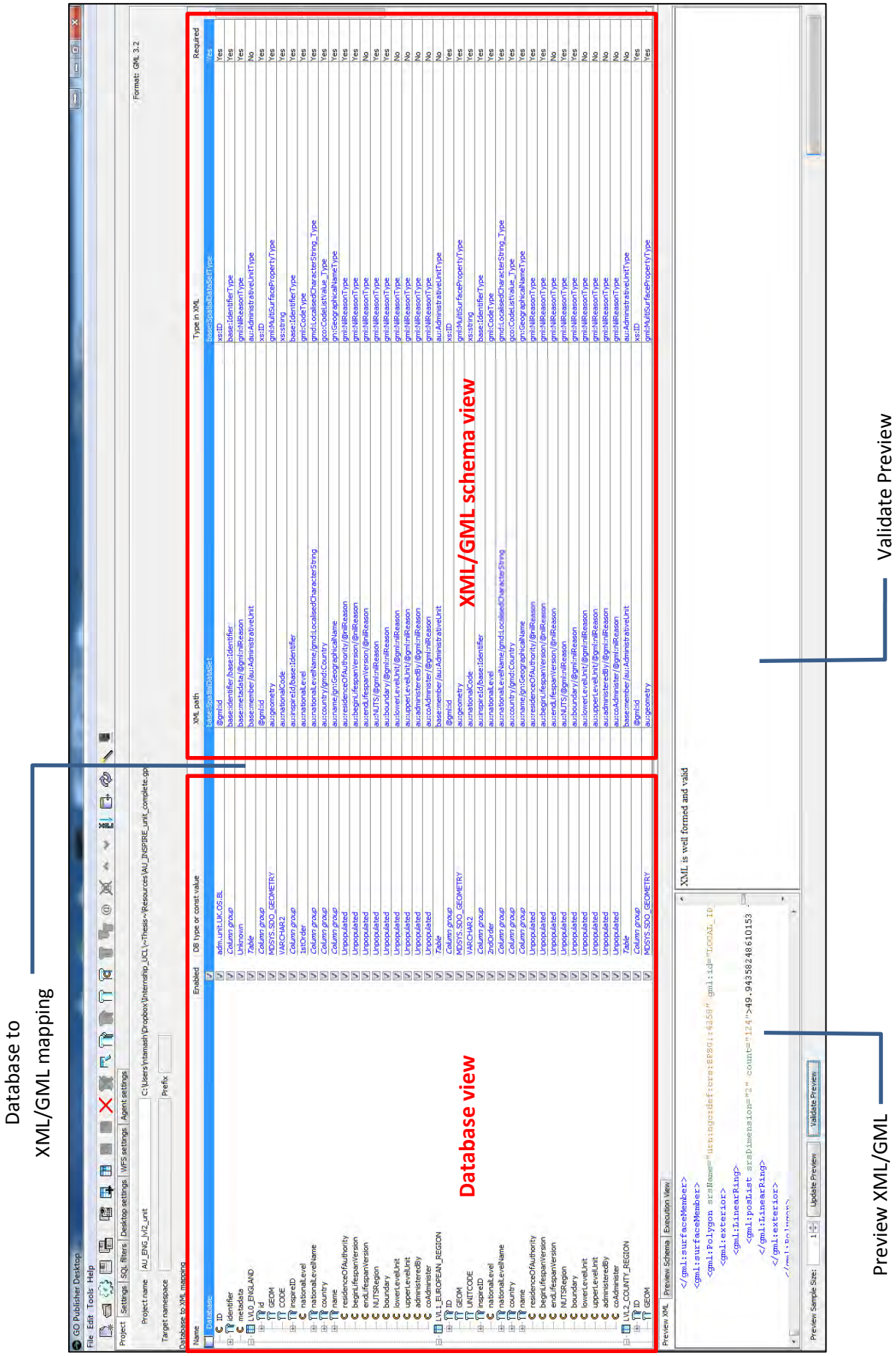
Once the data is loaded in the Oracle database, the actual transformation can be processed in GO Publisher. GO Publisher is quite user-friendly in the process of setting up a project by providing a 'Project Wizard', which allows defining the target application schema, setting the GML root element, connecting to the database, and selecting the required tables from the database.

The target application schema (i.e. XSD for Administrative Units) is downloadable from the INSPIRE (http://inspire.jrc.ec.europa.eu), as well as all the other schemas for the rest of the INSPIRE themes. It is also possible to connect to web schemas by providing the URL. The root element in an INSPIRE GML dataset (also called GML instance document) (i.e. 'base:SpatialDataSet') is defined in the Base Types application schema, which is an additional schema that is needed besides the Administrative Units schema. Base Types contains types not defined in the foundation schemas, but which are required in INSPIRE. After setting the root element, connection to the database is established and tables to be used are selected.

The main parts of the interface are the 'Database to XML mapping' panel, which consists in the left-hand side of the database view (i.e. the source data) and on the right hand side the XML/GML

elements that source data is going to be mapped into (i.e. XML/GML schema view). Right below there is a 'Preview XML' tab which shows a sample of the output data based on the mappings created in the 'Database to XML mapping' panel, on the fly. Next to the 'Preview XML' tab there is also a validator that checks if the preview generated validates against the INSPIRE schema. Errors will be displayed and will point to the invalid XML line in the 'Preview XML' panel. The 'Validate Preview' in combination with the 'Preview XML' is extremely useful because it allows the user to see in real-time the results of the created mappings, and track the possible errors, in a structure that will eventually represent the actual GML output. It also helps understanding the structure of the INSPIRE schema, especially if the user is not very familiar with the GML data format. Figure 35 on the next page illustrates the main interface of GO Publisher as described. The mappings are built from the 'XML Path' column in the 'Database to XML mapping' panel using a context driven drop-down list to select the elements that are being mapped to. All changes made to the 'XML Path' are reflected in the 'Preview XML' tab. The 'XML path' values can be edited manually as well by changing the name in the XML path, or by just simply selecting a different mapping from the available drop-down list. Columns from the database that are not required can be left out from the mapping by unchecking the boxes in the 'Enabled' column. Any additional processing required is mainly performed in the database view, where various SQL queries can be applied on the tables, new columns can be created based on constant values or other SQL queries, columns grouping and concatenations, etc.

Fig. 35: Go Publisher interface



5.2.1 Administrative Units

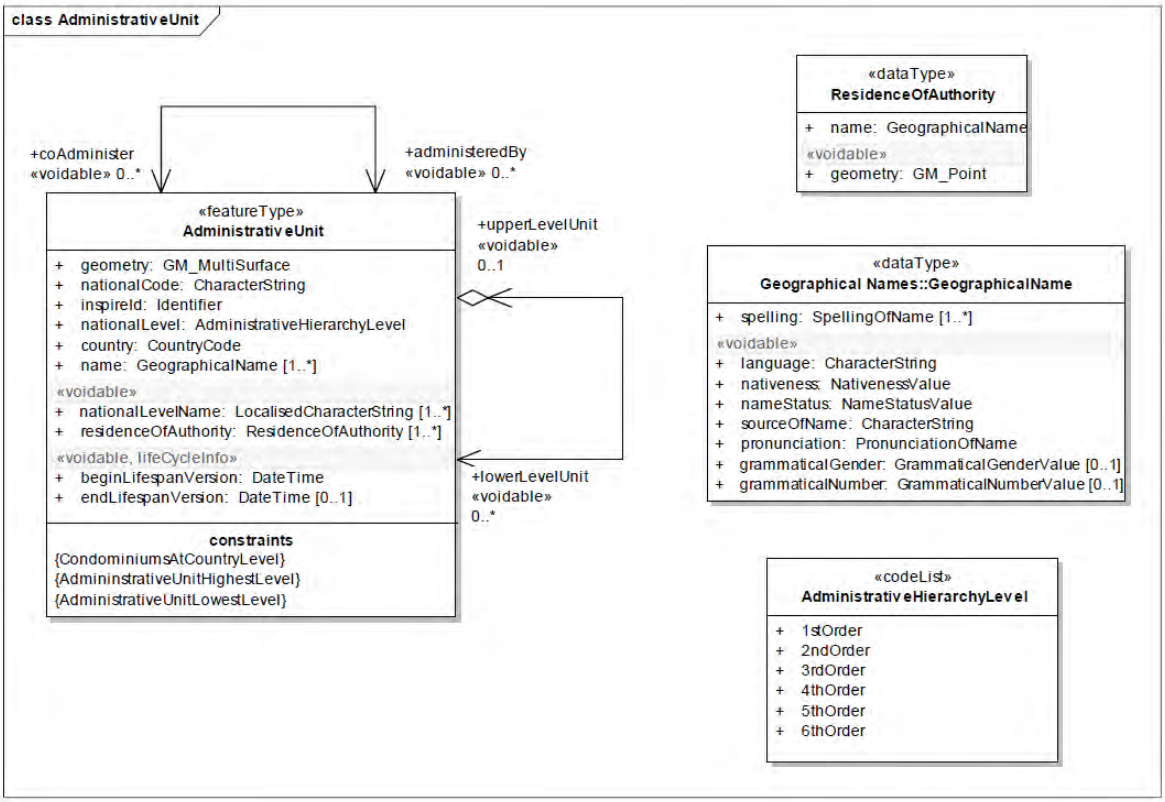


Fig. 36: The 'AdministrativeUnit' spatial object and its references (INSPIRE TWG AU, 2010)

The GML root element has already been defined during the wizard, so the first thing would be to define the spatial object from the INSPIRE application schema that the source table will be mapped to. In this case, an 'AdministrativeUnit' spatial object is bound to be created.

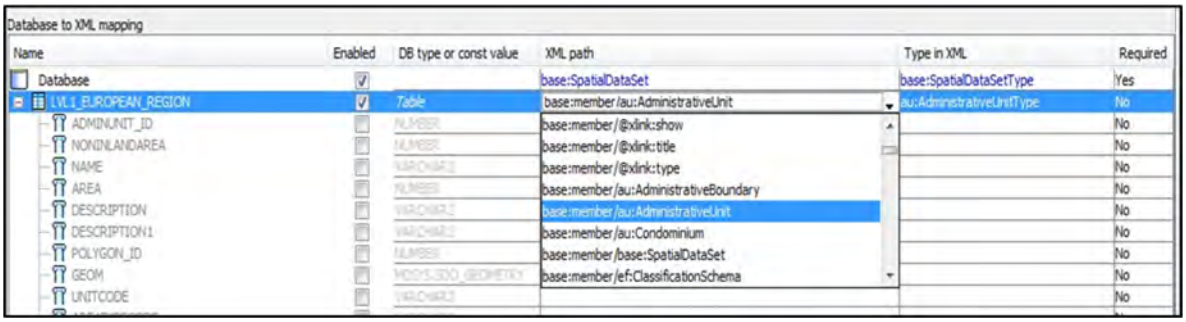


Fig. 37: Defining the INSPIRE spatial object

As it has been initially identified in the gap analysis in Chapter 3, mapping to the target schema do indeed turn out not to be that straight-forward by just directly mapping all the table columns to specific XML paths, but will require creating column groups, table joins, constant values, and other bespoke mappings. All these mapping techniques will be described on an element by element basis, with a more specific explanation of the implementation in GO Publisher, using screenshots, in Annex C of the thesis. In general, the process of schema matching is highly recommended to be performed by, or with the aid of thematic experts, people that have a very good understanding of the source data. Although the Administrative Units theme is considered to have one of the less complex data

specifications in INSPIRE, the fact that the source data, and especially the phenomenon it represents, the administrative geography of the UK, was not a domain the researcher was very familiar with, had serious consequences on the amount of time actually required to define the transformation. To start with, Figure 38 below depicts a preliminary schema matching for the 'AdministrativeUnit' spatial object.

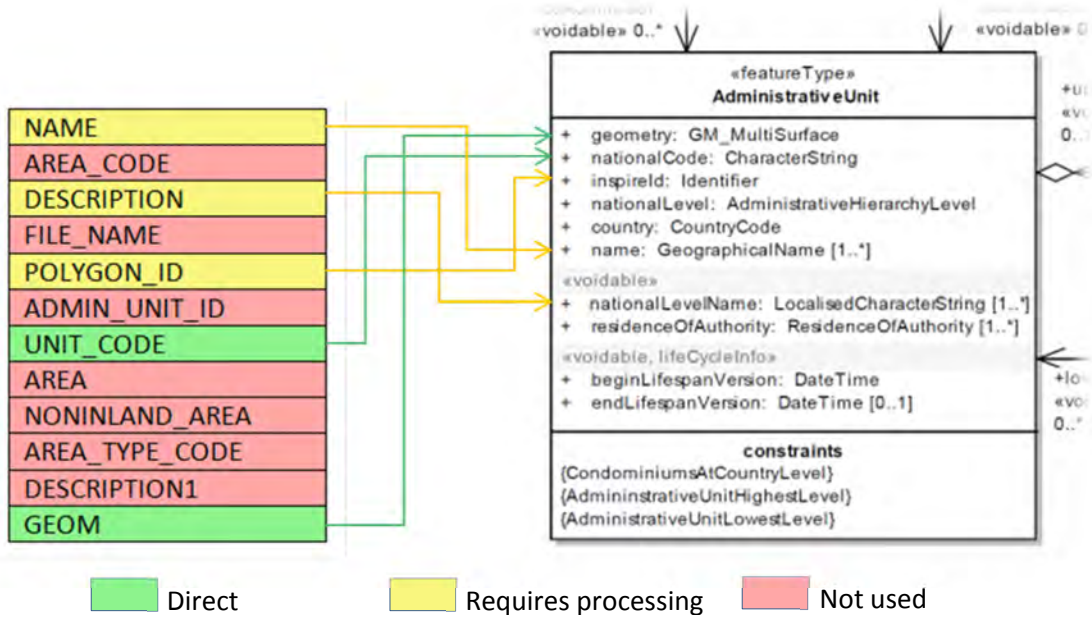


Fig. 38: Schema mapping

As it can be seen, out of all the source data attributes, only two have direct correspondents in the target schema, while another three will be used in bespoke mappings. All the remaining attributes in the source schema will not be used, with the rest of the target attributes to be created during the transformation where possible, while for others the 'voidable' property will be used.

Geometry and National code

Table attributes marked in green can be directly mapped to the target model. In GO Publisher this means as much as selecting the correct XML path from the dropdown list for each element. The attributes that can be mapped directly are the geometry attribute (i.e. GEOM), that was created automatically when the data was loaded in the Oracle database, and the UNIT_CODE attribute, which matches perfectly the description of the 'nationalLevel' attribute from the INSPIRE Feature Catalogue. The geometry attribute will also hold the coordinate reference system information, which in the source data is British National Grid. The transformation to ETRS89 will be covered later in the process, as GO Publisher is capable to perform coordinate transformation as well.

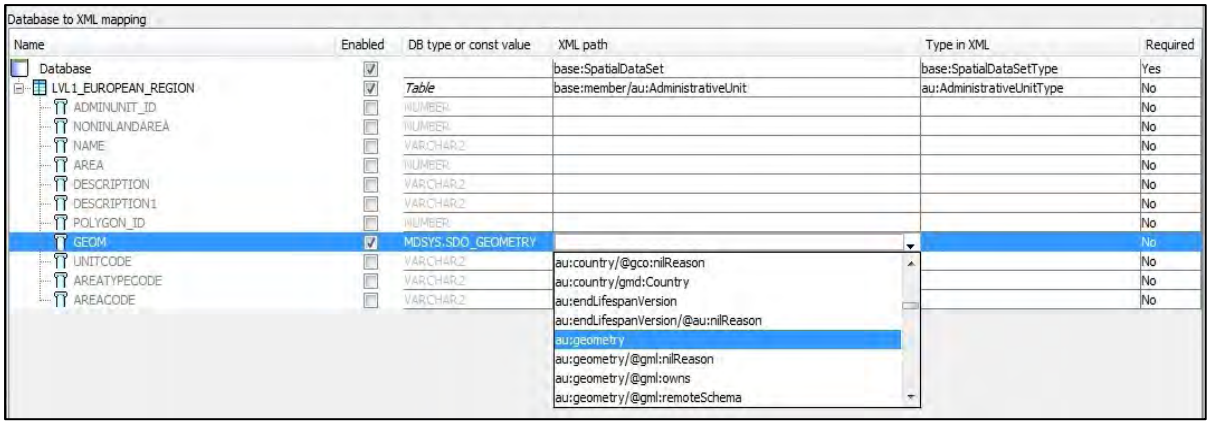


Fig.39: Direct mapping of the geometry attribute

Inspire ID

First attribute in the target schema that requires processing in order to be mapped properly is the 'inspireId' attribute. This is an 'identifier' type attribute, defined by INSPIRE, and has the following description in the Feature Catalogue: *"An external object identifier is a unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object. The identifier is an identifier of the spatial object, not an identifier of the real-world phenomenon"*. This definition is highly ambiguous for an element that is supposed to be internationally (or at least European) unique for every single spatial feature of any given dataset that would fall under the INSPIRE scope. In addition, it all falls down to the (many) data providers to come up with a unique combination, having in the same time a certain consistency at European level of how this combination is constructed. The data specification document slightly clears this issue by stating that *"the identifier shall consist of two parts: the namespace and a local id. The namespace is to uniquely identify a national registry wherein the identifier is registered, whereas id is to uniquely identify an object within this namespace. The pragmatic approach to making it internationally unique is to add a prefix of the Member State identifier along with a theme specific identifier for the namespace"*. Although using the recommended pragmatic approach will be the way to go for the scope of this project, it would still be interesting to see how the data providers across Europe will interpret the specification and to what extent there will be a consistence among the INSPIRE identifier type attribute. Therefore, the 'namespace' here will be 'UK.INSPIRE.AU.adm.unit' and the 'localId' will be given the value of the 'POLYGON_ID' source data attribute, which as according to the source data technical document (Ordnance Survey, 2012), 'POLYGON_ID' is a unique identifier across the entire Boundary-Line dataset.

The technique here is to create a so called column group that will hold the two elements. The 'namespace' gets the constant value mentioned, while the 'localId' will be unique across all features based on the 'POLYGON_ID' column. The entire group is then mapped to the 'inspireId' attribute in the target schema. The grouping of columns allow the creation of nested XML/GML elements, but this feature has to be used with caution as it is very easy to nest elements in a way that is not conformant to the target application schema. It is a matter of selecting the correct XML path from the drop-down list in GO Publisher for both the column group first, and then the individual columns that form the group, as the XML paths selection for the individual columns adapt based on the path selected for the group. Figure 40 below illustrate the mapping in GO Publisher and the resulting XML/GML for two 'inspireId' alternatives.

inspireID	☑	*Column group*	au:inspireId	base:IdentifierPropertyType	Yes
POLYGON_ID	☑	NUMBER	base:Identifier/base:localId	xs:string	No
namespace	☑	UK.INSPIRE.AU.adm.unit	base:Identifier/base:namespace	xs:string	No

a) – not valid

```
<au:inspireId>
  <base:Identifier>
    <base:localId>124433</base:localId>
  </base:Identifier>
  <base:Identifier>
    <base:namespace>UK.INSPIRE.AU.adm.unit</base:namespace>
  </base:Identifier>
</au:inspireId>
```

inspireID	☑	*Column group*	au:inspireId/base:Identifier	base:IdentifierType	Yes
POLYGON_ID	☑	NUMBER	base:localId	xs:string	Yes
namespace	☑	UK.INSPIRE.AU.adm.unit	base:namespace	xs:string	Yes

b) – valid

```
<au:inspireId>
  <base:Identifier>
    <base:localId>124433</base:localId>
    <base:namespace>UK.INSPIRE.AU.adm.unit</base:namespace>
  </base:Identifier>
</au:inspireId>
```

Fig. 40: Nesting XML/GML elements (a) – not valid against schema; (b) – valid against schema

This issue can appear upon mapping any nested elements so it is a matter of having a very good knowledge of the target GML application schema, or, in the case of using GO Publisher, make heavy use of the 'Preview XML' and 'Validate Preview' functionality to identify errors and to rectify the XML paths in order to create a XML/GML structure that validates against the target schema.

National level

Going further, to fill in the required attributes of the target schema, there is the 'nationalLevel' attribute which specifies the national administrative hierarchy at which the administrative units are established. This attribute doesn't exist in the source data, and even if it did, the data specifications contain a code list to populate it. For the mapping in GO Publisher a constant column is created, that holds the value in regard to the level in the national administrative hierarchy. In the case of Regions for example, that value is '2ndOrder'. There are 6 administrative levels defined in the data specifications with the '1stOrder' belonging to the country level.

National level name

Another target attribute that can be derived from the source data attributes is 'nationalLevelName' which refers to the name of the level in the national administrative hierarchy. The source attribute that holds that information is 'DESCRIPTION' and has a constant value across the entire dataset (i.e. European Region), hence it would've been easy to define it the source data wouldn't have held it.

The target attribute is of 'LocalisedCharacterString' type which means it is a character string with a locale. A locale is a combination of language, potentially a country, and a character encoding in which localised character strings are expressed. In this case, the character encoding is the content of the DESCRIPTION attribute, while the locale will just be set to 'ENG', depicting the English language.

Database to XML mapping

Name	Enabled	DB type or const value	XML path	Type in XML	Required
Database	☑		base:SpatialDataSet	base:SpatialDataSetType	Yes
LVL1_EUROPEAN_REGION	☑	Table	base:member/au:AdministrativeUnit	au:AdministrativeUnitType	No
GEOM	☑	MDSYS.SDO_GEOMETRY	au:geometry	gml:MultiSurfacePropertyType	Yes
UNITCODE	☑	VARCHAR2	au:nationalCode	xs:string	Yes
inspireID	☑	Column group	au:inspireId/base:Identifier	base:IdentifierType	Yes
nationalLevel	☑	2ndOrder	au:nationalLevel	gml:CodeType	Yes
nationalLevelName	☑	Column group	au:nationalLevelName/gmd:LocalisedCharacterString	gmd:LocalisedCharacterString_Type	Yes
DESCRIPTION	☑	VARCHAR2	.	gmd:LocalisedCharacterString_Type	No
locale	☑	ENG	@locale	xs:anyURI	No
country	☑	Column group	au:country/gmd:Country	gco:CodeListValue_Type	Yes
name	☑	Column group	au:name/gn:GeographicalName	gn:GeographicalNameType	Yes
ADMINUNIT_ID	☐	NUMBER			No
NONINLANDAREA	☐	NUMBER			No

```
<au:nationalLevelName>
  <gmd:LocalisedCharacterString locale="ENG">European Region</gmd:LocalisedCharacterString>
</au:nationalLevelName>
```

Fig. 41: Defining the 'nationalLevelName' attribute and resulting XML/GML

Country

For the 'country' attribute, again, a code list value will populate. This is a two character country code, in this case being 'UK'. There have been some issues with the definition of this specific mapping in GO Publisher because it was expected that adding a constant column with value 'UK' would validate against the schema. It occurred that a nested element was needed that contained two values: the name of the code list and the code list value itself. Therefore, another constant column was added containing the code list name (i.e. 'countryCode'), and grouped with the 'codeListValue' column.

Database to XML mapping

Name	Enabled	DB type or const value	XML path	Type in XML	Required
Database	☑		base:SpatialDataSet	base:SpatialDataSetType	Yes
LVL1_EUROPEAN_REGION	☑	Table	base:member/au:AdministrativeUnit	au:AdministrativeUnitType	No
GEOM	☑	MDSYS.SDO_GEOMETRY	au:geometry	gml:MultiSurfacePropertyType	Yes
UNITCODE	☑	VARCHAR2	au:nationalCode	xs:string	Yes
inspireID	☑	Column group	au:inspireId/base:Identifier	base:IdentifierType	Yes
nationalLevel	☑	2ndOrder	au:nationalLevel	gml:CodeType	Yes
country	☑	Column group	au:country/gmd:Country	gco:CodeListValue_Type	Yes
codeList	☑	countryCode	@gco:codeList	xs:anyURI	Yes
country	☑	UK	@gco:codeListValue	xs:anyURI	Yes
ADMINUNIT_ID	☐	NUMBER			No

```
<au:country>
  <gmd:Country codeList="countryCode" codeListValue="UK"/>
</au:country>
```

Fig. 42: Mapping the 'country' attribute and the resulting XML/GML

An observation that has to be made here, and this is probably a unique situation across the Member States, is that the United Kingdom, which is a country in its own right, actually consists of four countries: England, Wales, Scotland, and Northern Ireland. As the case study is solely based on England, due to the differences in the administrative geography between the four countries, it is sensible to say that this code list should maybe modified, and contain a separate value for each of the four countries.

Name

The last compulsory attribute is the name of the administrative unit. This attribute is of 'GeographicalName' type, a data type belonging to the Geographical Names application schema, which is another data theme in INSPIRE. In other words, the name attribute to nest all attributes contained in the 'GeographicalName' data type. These are depicted in the UML diagram in Figure 36 (page 58). The only required attribute in 'GeographicalName' is 'spelling', which itself contains nested elements, one of those (i.e. 'text') eventually holding the administrative units name. Everything else in the 'GeographicalName' data type is voidable, therefore this option will be extensively used in this case as most of these attributes are not known. An appropriate 'voidValueReason' will be specified, which in this case is 'Unpopulated'. Therefore, constant columns will be created for all these attributes, and then mapping them to the correct XML path.

name	☑	Column group	au:name/gn:GeographicalName	gn:GeographicalNameType	Yes
language	☑	ENG	gn:language	gn:(anonymous)	Yes
nativeness	☑	Unpopulated	gn:nativeness/@nilReason	gml:NilReasonType	Yes
nameStatus	☑	Unpopulated	gn:nameStatus/@nilReason	gml:NilReasonType	Yes
sourceOfName	☑	Unpopulated	gn:sourceOfName/@nilReason	gml:NilReasonType	Yes
pronunciation	☑	Column group	gn:pronunciation/gn:PronunciationOfName	gn:PronunciationOfNameType	Yes
pronunciationSoundLink	☑	Unpopulated	gn:pronunciationSoundLink/@nilReason	gml:NilReasonType	No
pronunciationIPA	☑	Unpopulated	gn:pronunciationIPA/@nilReason	gml:NilReasonType	No
spelling	☑	Column group	gn:spelling/gn:SpellingOfName	gn:SpellingOfNameType	Yes
NAME	☑	VARCHAR2	gn:text	xs:string	Yes
script	☑	Unpopulated	gn:script/@nilReason	gml:NilReasonType	Yes
transliterationScheme	☑	Unpopulated	gn:transliterationScheme/@nilReason	gml:NilReasonType	No
grammaticalGender	☑	Unpopulated	gn:grammaticalGender/@nilReason	gml:NilReasonType	No
grammaticalNumber	☑	Unpopulated	gn:grammaticalNumber/@nilReason	gml:NilReasonType	No

```
<au:name>
  <gn:GeographicalName>
    <gn:language>ENG</gn:language>
    <gn:nativeness nilReason="Unpopulated" xsi:nil="true"/>
    <gn:nameStatus nilReason="Unpopulated" xsi:nil="true"/>
    <gn:sourceOfName nilReason="Unpopulated" xsi:nil="true"/>
    <gn:pronunciation>
      <gn:PronunciationOfName>
        <gn:pronunciationSoundLink nilReason="Unpopulated" xsi:nil="true"/>
        <gn:pronunciationIPA nilReason="Unpopulated" xsi:nil="true"/>
      </gn:PronunciationOfName>
    </gn:pronunciation>
    <gn:spelling>
      <gn:SpellingOfName>
        <gn:text>London Euro Region</gn:text>
        <gn:script nilReason="Unpopulated" xsi:nil="true"/>
        <gn:transliterationScheme nilReason="Unpopulated" xsi:nil="true"/>
      </gn:SpellingOfName>
    </gn:spelling>
    <gn:grammaticalGender nilReason="Unpopulated" xsi:nil="true"/>
    <gn:grammaticalNumber nilReason="Unpopulated" xsi:nil="true"/>
  </gn:GeographicalName>
</au:name>
```

Fig. 43: Mapping the 'name' attribute and the resulting XML/GML

The value for the 'language' attribute of the 'GeographicalName' data type is also added, which refers to the language the name is spelled in, in this case being English (i.e. ENG). This is also a code list value.

Residence of authority

The 'residenceOfAuthority' attribute is supposed to refer to the centre for national or local administration for the respective administration unit. The attribute is a 'ResidenceOfAuthority' data type which itself contains two attributes, the location of the 'residence of authority' (point geometry type), and name of the residence of authority ('GeographicalName' type). The 'GeographicalName' data type has already been described and used for the 'Name' attribute.

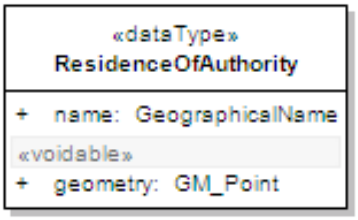


Fig. 44: 'ResidenceOfAuthority' data type

Residence of authority information is obviously missing from the source data, as it would imply having another dataset holding the location (i.e. point geometry) of the centres of administration for each administrative unit. Although this attribute is voidable, a small use case will be created to demonstrate functionality. Virtual centres of administration will be created manually for a limited extent of the dataset. The limited extent will cover the London Boroughs, level 4 in the national administrative hierarchy, on the same level with Districts and Unitary Authorities throughout the rest of England. The point data representing the centres of Administration are generated in ArcMap with the 'Feature To Point' tool applied on the London Boroughs (polygon data). The generated points are technically equivalent to the polygon centroids. This operation would also be possible directly in the Oracle database if the full spatial extension would be available (i.e. the SDO_GEOM.SDO_CENTROID function). Oracle Express edition, used in this case study, only supports a cut-down version of the full spatial extension with only few basic spatial functions available.

Fig. 45 Residence of authority points

A new column is added to the newly generated point geometry shapefile that will contain the name of the centres of administration (for London these being called London Borough Councils). The names are then added manually for all 34 boroughs. Finally, the shapefile is then loaded in the Oracle XE database with GeoRaptor and added to the GO Publisher project. As the point dataset inherited all the attributes of the polygon features, it is possible to join the latter with the new added table in GO Publisher. The entire joined table containing the centres of administration is mapped to the 'residenceOfAuthority' attribute and then the nested elements (i.e. geometry and name) are mapped to their own counterparts in the schema. It must be noted that for 'name' the entire 'GeographicalName' data type structure has to be re-created (see Figure 46).

USE_CASE_RES_AUTH_LVL4	☑	Table	au:residenceOfAuthority/au:ResidenceOfAuthority	au:ResidenceOfAuthorityType	Yes
name	☑	Column group	au:name/gn:GeographicalName	gn:GeographicalNameType	Yes
language	☑	ENG	gn:language	gn:(anonymous)	Yes
nativeness	☑	Unpopulated	gn:nativeness/@nilReason	gml:NilReasonType	Yes
nameStatus	☑	Unpopulated	gn:nameStatus/@nilReason	gml:NilReasonType	Yes
sourceOfName	☑	Unpopulated	gn:sourceOfName/@nilReason	gml:NilReasonType	Yes
pronunciation	☑	Column group	gn:pronunciation/gn:PronunciationOfName	gn:PronunciationOfNameType	Yes
pronunciationSoundLink	☑	Unpopulated	gn:pronunciationSoundLink/@nilReason	gml:NilReasonType	No
pronunciationIPA	☑	Unpopulated	gn:pronunciationIPA/@nilReason	gml:NilReasonType	No
spelling	☑	Column group	gn:spelling/gn:SpellingOfName	gn:SpellingOfNameType	Yes
RES_AUTH	☑	VARCHAR2	gn:text	xs:string	Yes
script	☑	Unpopulated	gn:script/@nilReason	gml:NilReasonType	Yes
transliterationScheme	☑	Unpopulated	gn:transliterationScheme/@nilReason	gml:NilReasonType	No
grammaticalGender	☑	Unpopulated	gn:grammaticalGender/@nilReason	gml:NilReasonType	No
grammaticalNumber	☑	Unpopulated	gn:grammaticalNumber/@nilReason	gml:NilReasonType	No
GEOM	☑	MDSYS.SDO_GEOMETRY	au:geometry	gml:PointPropertyType	Yes
POLYGON_ID	☐	[illegible]			No

```
<au:residenceOfAuthority>
  <au:ResidenceOfAuthority>
    <au:name>
      <gn:GeographicalName>
        <gn:language>ENG</gn:language>
        <gn:nativeness nilReason="Unpopulated" xsi:nil="true"/>
        <gn:nameStatus nilReason="Unpopulated" xsi:nil="true"/>
        <gn:sourceOfName nilReason="Unpopulated" xsi:nil="true"/>
        <gn:pronunciation>
          <gn:PronunciationOfName>
            <gn:pronunciationSoundLink nilReason="Unpopulated" xsi:nil="true"/>
            <gn:pronunciationIPA nilReason="Unpopulated" xsi:nil="true"/>
          </gn:PronunciationOfName>
        </gn:pronunciation>
        <gn:spelling>
          <gn:SpellingOfName>
            <gn:text>Royal Borough of Kingston upon Thames</gn:text>
            <gn:script nilReason="Unpopulated" xsi:nil="true"/>
            <gn:transliterationScheme nilReason="Unpopulated" xsi:nil="true"/>
          </gn:SpellingOfName>
        </gn:spelling>
        <gn:grammaticalGender nilReason="Unpopulated" xsi:nil="true"/>
        <gn:grammaticalNumber nilReason="Unpopulated" xsi:nil="true"/>
      </gn:GeographicalName>
    </au:name>
    <au:geometry>
      <gml:Point srsName="urn:ogc:def:crs:EPSG::4258" gml:id="LOCAL_ID_2">
        <gml:pos>51.387890740283254 -0.2869144497634939</gml:pos>
      </gml:Point>
    </au:geometry>
  </au:ResidenceOfAuthority>
</au:residenceOfAuthority>
```

Fig. 46: Mapping the 'residenceOfAuthority' attribute and the resulting GML/XML

It must be noted that at the end, the 'AdministrativeUnit' spatial object, will now hold two geometry fields of different types (i.e. polygon and point).

Begin lifespan version / End lifespan version

These attributes are used in most of the spatial objects defined in the INSPIRE application schemas and they refer to the lifespan of the spatial object. It is important to note that lifecycle information refers to the lifespan of the version in the spatial dataset itself, which is different from the temporal characteristics of the real-world phenomenon described by the spatial object. Because of that, there is a recommendation in the data specifications saying that "*if life-cycle information is not maintained as part of the spatial data set, all spatial objects belonging to this data set should provide a void value with a reason of 'unknown'*", instead of 'unpopulated' as it would've been expected. And indeed, the source data used in this study case does not hold lifecycle information on the dataset itself, therefore, it will be given the recommended value as void reason.

Upper Level unit / Lower level unit

The 'upperLevelUnit' and 'lowerLevelUnit' attributes refer to the unit established at an upper level or units established at a lower level in the national administrative hierarchy. These associations will basically help to identify the nesting of the administrative units (i.e. what counties can be found in a specific region, or what districts does a county contain). In addition there are two constraints defined for these associations which specify that the administrative unit at the highest level in the national administrative hierarchy (i.e. nationalLevel = 1stOrder) cannot have an 'upperLevelUnit' association, while the administrative units established at the lowest level in the national administrative hierarchy (i.e. nationalLevel = 6thOrder, or whichever is lowest) cannot have a 'lowerLevelUnit' association. These are common-sense constraints that have to be put in place to avoid inconsistencies.

As associations, the two attributes will refer to upper and lower level units by 'xlinks', as defined in the GML standard. Xlink components are used in GML to implement associations between (spatial) objects by reference. The most important 'xlink' component is 'xlink:href', which represents the identifier of the resource which is the target of the association, given as a URI (unique object identifier). Therefore, the appearance of an 'xlink:href' on the, for example, 'upperLevelUnit' property, indicates that the value of the property (i.e. in this case the upper level administrative unit) shall be found by traversing the link, that is the value is pointed to by the value of the 'xlink:href' attribute (Open Geospatial Consortium Inc., 2007). For associations occurring between objects in the same GML document, the 'xlink' can be as simple as the GML identifier of the referenced object, or any other unique identifier element of the referenced object.

Since the source data doesn't hold nesting of administrative units information, a small use case will be created manually for a limited extent (i.e. London area) and also limited to one level in the administrative hierarchy, so that the units established at the specific level will have associations defined for at least one of the 'upperLevelUnit' / 'lowerLevelUnit' associations. The website of the Office for National Statistics offers for download several lookup tables in CSV format that hold nesting information the administrative units. These are limited though to Wards within Districts, and Districts within Counties. The administrative units are identified in the table by their unique national code (i.e. UNIT_CODE in the source data). The tables have the following structure:

Table 5: Lookup table of administrative units nesting

District code	(Old) District code	District name	County code	(Old) County code	County name
E07000004	11UB	Aylesbury Vale	E10000002	11	Buckinghamshire
E07000005	11UC	Chiltern	E10000002	11	Buckinghamshire
…	…	…	…	…	…

As each record in this table corresponds to a district, it is possible to join it in GO Publisher with the table holding the Districts, London Boroughs, and Unitary Authorities (i.e. 4th level in the national administrative hierarchy) based on the district code. The entire table can be mapped to the 'upperLevelUnit' attribute, and the column holding the national code of the county can be mapped to the 'xlink:href' element. This is a valid association giving the GML standard specifications since all administrative units will be sitting in the same GML instance document (i.e. dataset), so it is very easy to reference one unit to another by using any unique identifier field. As an alternative to the retrieved lookup tables it is also possible to use the 'SDO_OVERLAPS' function in the Oracle database and create own lookup tables.

The problem that arises though is with the 'lowerLevelUnit' association. This is a 'one to many' relationship, and even though the other available lookup table provides Wards within Districts nesting, it is not possible to join it in GO Publisher on the same basis, because each record (i.e. district) in the spatial data table would have to accommodate multiple records from the lookup table (i.e. wards). One solution to avoid multi-valued attributes would be to have an additional parent-child class that would model this part of the schema. For the scope of this case study though, the voidable property is used.

Administered by / Co-administer

'administeredBy' and 'coAdminister' are association attributes as well, which will also refer to other administrative units. 'administeredBy' will refer to units established at the same level of national administrative hierarchy which co-administer the respective administrative unit, while 'coAdminister' refers to a unit at the same level of administrative hierarchy that is co-administered by the respective administrative unit. Between the two, there is an inverse role. After a brief research it has been concluded that there is no administrative unit in the study case that might be administered by other units. Therefore, constant values will be created for the two attributes and given the value of 'Unpopulated' as void value reason. Even though the multiplicity of these associations is 0…*, and in this case the two associations simply don't exist, they still have to be defined in the GML instance with a void reason value.

Administrative boundary

The administrative boundary is a voidable association to all the boundary features that form the boundary of the respective administrative unit. This information does not exist in the in the source data. A solution to this issue would to be to create a topology relation table in the database between the administrative units and the boundary features. The encountered issue was that none of the available spatial functions in Oracle Express edition serve the purpose (i.e. select all the boundary

features that are coincident with the boundary of the administrative unit). Therefore, a void value reason is provided.

Condominium

The condominium, as already explained in Chapter 3, is a territory administered by two countries. In the schema there should be an association between an administrative unit and such a territory. As a constraint, it is specified that the administrative unit that receives this association can only be an administrative unit at the highest level in the national administrative hierarchy, which means at country level (i.e. nationalLevel = 1stOrder). Again, such an association won't be necessary as there is no territory administered by England and any other country. A constant is created that receives the value 'Unpopulated' as void reason.

NUTS region

The last element is also an association, this time to a NUTS region spatial object. Each administrative unit or group of administrative units would normally have a NUTS region correspondent which is part of a different application schema (i.e. Annex III theme). As NUTS regions data, this association will be passed to the 'Unpopulated' void value reason.

Standard GML elements

Besides all the attributes of the INSPIRE data model for Administrative Units, there are also some standard XML/GML attributes that have to be defined in order to have a valid GML 3.2.1 instance. These have no connection whatsoever to the Administrative Units application schema, but are elements that are defined by the INSPIRE GML standard and that have to appear on every GML dataset. These attributes are an external object identifier and a container for metadata.

The external unique object identifier is published by the responsible body (i.e. data provider), and it is used by external applications to reference the spatial object. Therefore, this identifier has the same structure as the 'inspireId' attribute already defined during the mapping, the difference being that 'inspireId' is referencing each individual feature of the spatial dataset, while the Base Type identifier references the entire spatial dataset. Again, the identifier is given intuitive values for the 'localId' element (i.e. adm.unit) and the 'namespace' element (i.e. UK.INSPIRE.AU). In the case of metadata, an 'Unknown' value is given as a void value reason because INSPIRE metadata doesn't fall within the scope of this project. Would've there been a valid INSPIRE metadata record for the spatial dataset, that would've have to be reflected within the metadata attribute.

Finally, every GML document must have a GML identifier to it, much like other formats have their own identifier element (e.g. shapefile unique identifier). Once again, there must be a GML identifier for the dataset as a whole, but also for each individual feature of the dataset. To keep things simple and coherent, the features GML IDs will be defined using the same elements as the 'inspireId' attribute; a copy of the 'POLYGON_ID' attribute from the source table, and a constant value (i.e. UK.INSPIRE.AU.adm.unit _). These will be concatenated, so an example of a feature GML identifier will appear under the following form: 'UK.INSPIRE.AU.adm.unit _124430'. The main GML identifier will receive a constant value (i.e. 'UK.INSPIRE.AU.adm.unit'). The structure of the GML identifiers doesn't have any requirement defined, so it will also be on the latitude of the responsible organisations to come up with their own identifiers. The GML identifiers though, do not have the same importance as the INSPIRE defined identifiers, which are required to be unique across the

entire European spatial data infrastructure. They have to be created because they are 'hard-coded' elements of the GML standard.

Coordinate reference system

The last step in the data transformation is to define the spatial reference system. GO Publisher has coordinate reference system transformation capabilities, which is as simple as defining the source spatial reference system (if it hasn't already been extracted from the source data), and defining the coordinate reference system that the data has to be published in. In the INSPIRE data specifications ETRS89 (European Terrestrial Reference System 1989) has been set as the coordinate reference system to be used for making available spatial datasets. As it would be expected ETRS89 is the suitable CRS for use in Europe, onshore and offshore. The generic ETRS89 is used in this study case (SRID = 4258, according to the European Petroleum Survey Group – EPSG database), which is suitable for the any European region. There are also other, more specific, projected ETRS89 coordinate reference systems, which are only suitable for a given region/country.

Exporting / Publishing the GML dataset

At this stage the GML document can be published, consisting of a dataset that at least schema-wise, should be INSPIRE compliant. On the main menu bar of GO Publisher there is a 'Create a XML/GML file' button, which pops up a window where several options are available like output location on the local machine, and the option to only export/publish a limited number of features from the dataset as a sample. At this point the produced INSPIRE Administrative Unit GML dataset is ready to go through conformance testing, which will be tackled in Chapter 6.

5.2.2 Administrative Boundaries

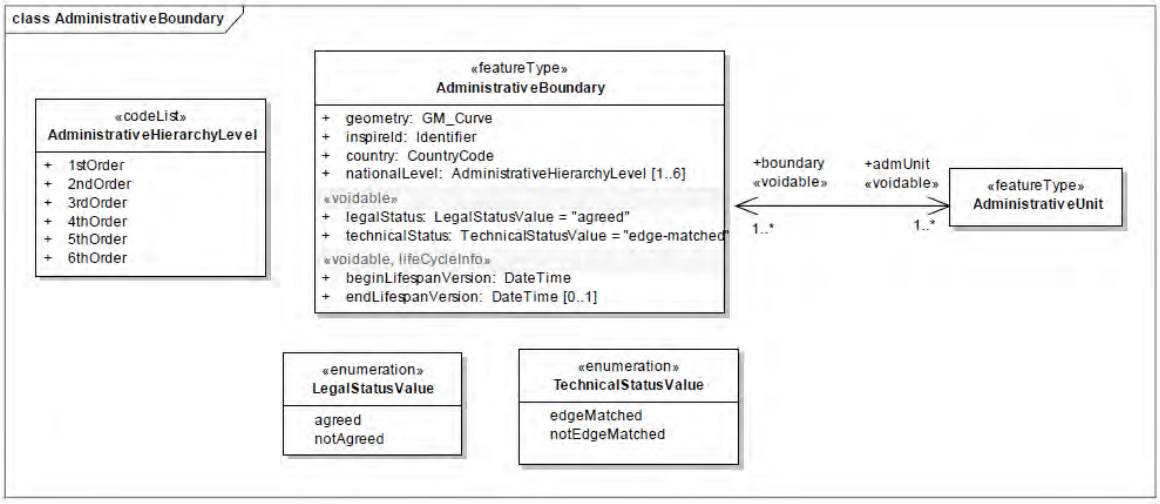


Fig. 47: The 'AdministrativeBoundary' spatial object and its references (INSPIRE TWG AU, 2010)

In regard to the second main spatial object of the Administrative Units schema, the 'AdministrativeBoundary', the mapping process is very similar. The 'AdministrativeBoundary' has less attributes than the 'AdministrativeUnit', and most of them are the exact same attributes, as depicted in the UML diagram in Figure 47.

There are only two new attributes. One of them refers to the legal status of the boundary (i.e. 'legalStatus') which is an of enumeration type attribute, being possible to use only two values:

‘agreed’ and ‘notAgreed’. The legal status is considered in terms of political agreement or disagreement of the administrative unit separated by this boundary. Since the source data is official Ordnance Survey data, during the mapping in GO Publisher, all boundary data will receive the ‘agreed’ value for the ‘legalStatus’ attribute.

The other attribute refers to the technical status of the boundary (i.e. ‘technicalStatus’) and is also an enumeration type attribute, being able to hold the following values: ‘edgeMatched’ and ‘notEdgeMatched’. The technical status of the boundary is considered in terms of its topological matching or not matching with the borders of all separated administrative units. Edge-matched literally means that the same set of coordinates is used. Fortunately, in the study case used in this project, England does not have any neighbouring countries that may pose edge matching issues of administrative units’ borders. It is surrounded by either ocean or sea, while the administrative data for neighbouring regions, Wales and Scotland, is included in the same OS Boundary-Line product. In addition, it must be noted that in the cases where there are neighbouring countries, not all edges in the boundary graph need to be edge matched. In practice, the ‘technicalStatus’ attribute only applies to edges that contain in the ‘nationalLevel’ attribute the value ‘1stOrder’, which makes them part of the national boundary.

One aspect that has to be noted in the ‘AdministrativeBoundary’ spatial object is the different approach to the ‘nationalLevel’ attribute, which has a multiplicity of 1…6. This means that an edge part of the boundary topological graph can be part of up to six levels in the national administrative hierarchy. This is especially true for the ‘flat model’ geometry structure for boundary features that was also used in this case study. To overcome this challenge 5 new columns were added to the boundary dataset, each corresponding to one level in the national administrative hierarchy. Then, using the ‘SDO_EQUAL’ function in the Oracle database and some generic SQL queries, each edge in the boundary dataset was checked against geometric equality with the polygon boundaries, but each level of administrative hierarchy at a time. Each time, the selected edges were given the value ‘1’ in the column that corresponded to the level of administrative hierarchy of the comparison polygon, and the value ‘0’ for the unselected edges. The result was that an edge that is part of, for example, a boundary at all levels in the national administrative hierarchy, would receive value ‘1’ for all the five columns. Based on this ‘boolean’ approach, when performing the mapping in GO Publisher, a custom SQL query is placed on each of the 5 columns that has the following condition: if the value in the column corresponding to the 1st level of national administrative hierarchy is 1, then the value of the mapping is ‘1stOrder’; if it’s 0, the mapping is null.

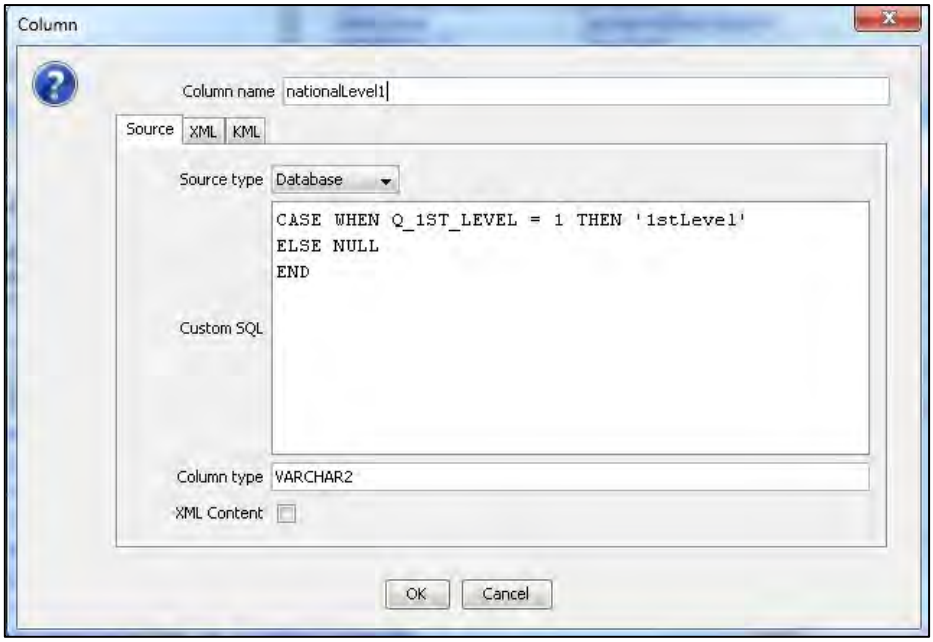


Fig. 48: Conditional mapping of the 'nationalLevel' attribute

Therefore, the XML/GML resulting for a boundary feature that is part of multiple levels in the national administrative hierarchy looks like this:

```
<au:nationalLevel>2ndOrder</au:nationalLevel>
<au:nationalLevel>3rdOrder</au:nationalLevel>
<au:nationalLevel>4thOrder</au:nationalLevel>
<au:nationalLevel>5thOrder</au:nationalLevel>
```

Fig. 49: 'nationalLevel' XML/GML for boundary features

In case the 'multi-layer model' is used, the 'nationalLevel' attribute follows the same approach like in 'AdministrativeUnit' because there will be one GML instance document for each level in the national administrative hierarchy.

Another aspect that slightly differs from the 'AdministrativeUnit' mapping process is the structure of the identifiers. As it has been explained in Section 5.1, during the pre-processing stage of the data, when the boundaries are derived, they lose all attributes that were inherited from the administrative units since the geometry structure was changed. Therefore, there is no unique identifier in the boundary data that is need to be mapped as 'localId'. Luckily, GO Publisher has an automatic id generation function. This is used to generate unique ids on the fly during the mapping. These generated ids have a simple structure: 'local_id_2' (the number always changes with each feature).

Finally, there is the 'admUnit' association, which is the exact inverse of the 'boundary' association in the 'AdministrativeUnit' spatial object. This will receive a void value, for the same reasons explained during the 'AdministrativeUnit' mapping. Same recommendations apply though.

5.2.3 Condominiums

Since it has been established that there are no territories falling under the 'Condominium' spatial object scope in this study case, there is no further action to be performed. A short overview of a possible transformation will be given for cases where condominiums exist.

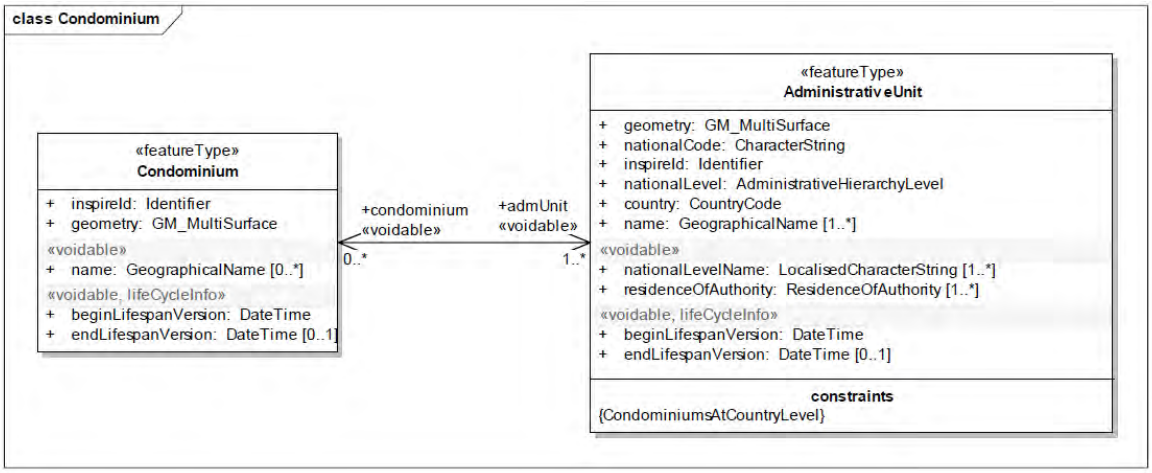


Fig. 50: Condominium UML model (INSPIRE TWG AU, 2010)

As can be observed in Figure 50, the attributes of the 'Condominium' spatial object are a cut-down version of the 'AdministrativeUnit' spatial object, with only the INSPIRE identifier and geometry, which are mandatory, and name and life cycle information which are optional. Spatially, the 'Condominium' is also an administrative unit, so the exact same mapping procedures apply as with 'AdministrativeUnit'. There is also the 'admUnit' association, which is the exact inverse of the 'condominium' association in the 'AdministrativeUnit' spatial object, and only if it applies of course. Again, important to note that 'Condominium' can only associate to administrative units at 1st level in the national administrative hierarchy (i.e. country level), and only units at the 1st level in the national administrative hierarchy can associate to a condominium.

5.2.4 NUTS Regions

NUTS Regions is a spatial object type belonging to the Annex III Statistical Units theme. Since NUTS Regions data was not available for this study case, there is no transformation applying to this spatial object.

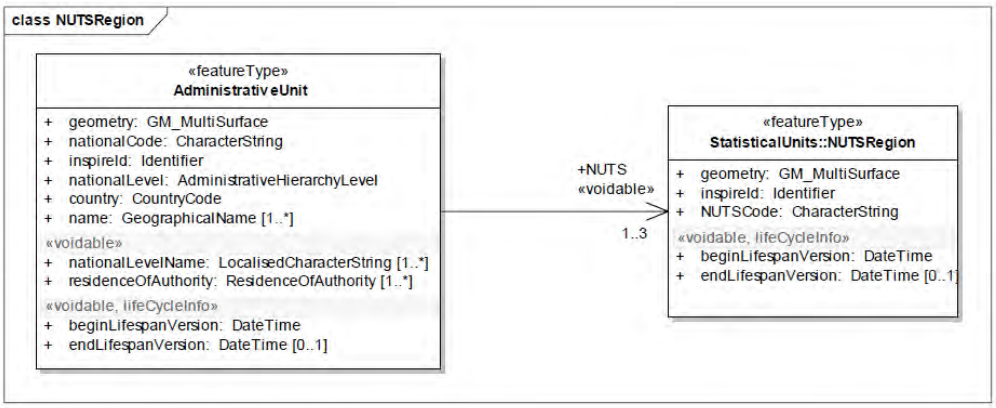


Fig. 51: NUTSRegions UML model (INSPIRE TWG AU, 2010)

Similarly to the 'Condominium' though, a brief overview of the attributes and the transformation approach is given, in cases where NUTS data is available. NUTS Regions contains as well very similar attributes to the 'AdministrativeUnit' spatial object. There is geometry, INSPIRE identifier and life cycle information which would follow the exact same mapping process as with administrative units, and there is a 'NUTSCode' attribute which is the unique European code identifying the statistical units, similarly to the national administrative codes. NUTS datasets should definitely contain this attribute, therefore, mapping is straightforward. The 1..3 multiplicity, as also explained in chapter 3, refers to the 3 levels of NUTS regions (1, 2 and 3). An administrative unit can associate to all three at a time, as there is a spatially nesting relationship between them, just like between the different level of national administrative hierarchy.

Compared to the other associations between different spatial objects where in most cases the association applies in both directions, there is no associations from the 'NUTSRegions' spatial object to the 'AdministrativeUnit' spatial object.

5.3 Evaluation and conclusions

This section will first review GO Publisher as a tool and then reflect on the data transformation process itself and discuss the main bottlenecks encountered, in reference to the issues that were primarily identified in Chapter 3 before the transformation. Overall, this section should also be able to connect to the conclusions from the previous chapter, related to the general context of data transformation and available software solutions, in order to respond to the third research sub-question.

5.3.1 GO Publisher

Getting started with GO Publisher was not as straight-forward as wished due to lack of previous experience with the software but mainly the lack of XML-based grammar knowledge and GML in general. Therefore, a lot of time was spent in understanding the true functionality of GO Publisher and how to create valid mappings. Unfortunately, current documentation of GO Publisher is not as extensive as it would be expected, and for some features of the software is incomplete or missing. This might also be due to the fact that the software developers provide regular training sessions where the functionality is thoroughly explained and demonstrated, this being the main source of 'documentation' for the users. Besides general knowledge about XML/GML, maybe more importantly is to have solid understanding of the target data model and its correspondent XSD schema. The way GO Publisher as a piece of software is constructed also aids in the process, especially through the possibility to constantly view and validate in real-time the GML instance document that is created. In addition, GO Publisher can be considered as an extension to the source database, since it integrates a lot of SQL functionality that allows efficient manipulation of the source data. Many database native spatial functions can be used as well, especially if the source database is an Oracle one. The entire GO Publisher interface is strongly focused on the relation between the source database elements and target XML schema elements, therefore the mapping process becomes intuitive. The drop-down list selection approach of the target XML/GML elements is very efficient as the available options always adapt based on the element mapped at the higher level in the nesting structure of the XML/GML hierarchy. For instance if a group of columns or tables

is mapped to the 'name' attribute (case of the Administrative Units schema), the drop-down list options for the columns/tables in that group is limited only to elements belonging to the 'GeographicalName' data type. One drawback in general might be that although GO Publisher provides several database types connection options, when it comes to spatial functionality, it seems that using Oracle is the only solid solution. For instance, although PostgreSQL is supported, PostGIS is not. This is expected to be improved in the future versions. Moreover, source data can only be fetched from a database; it cannot be imported as a file format (e.g. shapefile).

In what concerns the mapping process, GO Publisher is relatively powerful. It can create constant or custom (based on SQL queries) columns when the data is missing from the source database or needs to be processed; 'xlink:href' attributes which create the associations to other features. It is also able to create XML/GML standard attributes and there are some native functions that will automatically derive some of these attributes (e.g. GML id, bounding box). Furthermore, it is possible to create various levels of nested tables and columns by joining and concatenation, post-process with style-sheets, or translating values with custom Java classes. It is fairly straightforward to use multiple spatial tables and with different geometry types in the same GML dataset. This is very helpful as some attributes have complex data types that have their own geometry, like 'residenceOfAuthority' for instance. Finally, GO Publisher allows publishing the created GML data in any required coordinate reference system, performing an on-the-fly coordinate conversion. However, the quality of that conversion is an open discussion, which will be further discussed in Chapter 6 during conformance testing.

As a general conclusion, GO Publisher is quite a powerful tool for publishing data into an INSPIRE conformant GML dataset as long as the user has a good knowledge of its functionality, but as well of the GML standard and especially the source and target data models and XSD schemas. The most efficient result will probably be obtained when it is used in conjunction with a desktop GIS package or using the full functionality of Oracle Spatial, and not just the Express edition which is much more limited. GO Publisher seems, and probably is at the moment, the most efficient solution exclusively for the mapping process, but when more advanced data processing is needed before or during the mapping, another software solution needs to be brought in the equation.

5.3.2 Transformation issues

As initially identified in Chapter 3, during the source-target data models gap analysis before the transformation, there were some aspects that confirmed to be problematic during the transformation. These aspects are related to both the structure of the source data, but also the requirements in the target data model.

First of all, only two of the four spatial objects in the target data model were produced (i.e. 'AdministrativeUnit' and 'AdministrativeBoundary'). This was because one spatial object (i.e. 'Condominium') was not applicable within the administrative structure of the case study, and to current knowledge it is hard to think of a national administrative structure in the EU that would make use of this type of spatial object. Finally, the last spatial object (i.e. NUTSRegions), requires an external data source that was not available for this case study, and in addition, this is a spatial object type belonging to another theme (i.e. Statistical Units – Annex III) for which data specifications are still in progress.

Regarding the source data model, the main issue was related to the boundaries that didn't exist, so had to be derived and processed to adhere to the required geometric model, and while doing so, all inherited attributes from administrative units were lost. In other words, the only information available was the geometry. Therefore, the attributes for boundaries (to correspond to the target data model) were sometimes derived straight from code lists (e.g. country), but also by automatic and random (and unique) generation (i.e. localId), or by additional pre-processing combined with SQL queries during the mapping (i.e. nationalLevel). Some information was impossible to derive, like the life cycle information, which was also the case for the 'AdministrativeUnit' spatial object type. In addition, nesting information of administrative units was not available in the source data, although with a full Oracle spatial extension a topology model can be created to extract such information. Despite that, a small use case was still implemented by the aid of an external lookup table that contained nesting information, but only for upper level units.

With the target data model the main issue was that it often required information that sits in another data source or INSPIRE theme. The most challenging aspect in this sense was the 'residenceOfAuthority' attribute that required point geometry data representing the location of the residence of authority for each administrative unit. As this information doesn't belong to any other INSPIRE theme, nor is it expected to be maintained by any data provider in the source data model of administrative units, it has to be brought in from external sources. Again, a small use case was implemented here (points derived by polygon centroids and names manually inserted) to have a complete as possible GML dataset for at least a small part of the administrative units. In addition, the 'name' required the study of the data model of another INSPIRE theme, Geographical Names, while the 'NUTSRegions' association the Statistical Units theme, although this data was not available, hence not possible to create the association. It is also important to note that beside the geometry data for 'NUTSRegions', it is very important to have in the administrative units source data, information about the correspondent NUTS regions. A last observation is related to the identifiers, which although have been intuitively implemented and with a good chance of uniqueness across INSPIRE, a more administrated and standardised approach is required in the data specifications concerning the content of the 'namespace' and 'localId' components (and this is a valid observation across all themes), so that just by looking at an identifier useful information can be derived, like belonging theme, country, etc.

6. Conformance testing

As already described in the literature review chapter, conformance testing is a very important requirement in the INSPIRE data specifications because there is no other formal method to declare a dataset to be conformant to the specifications. The very first highlighted requirement in all INSPIRE data specifications states the following: "*Any dataset claiming conformance with this INSPIRE data specification shall pass the requirements described in the abstract test suite presented in Annex A to this specification*". As the Abstract Test Suite (ATS) for none of the INSPIRE themes have been defined yet, this chapter will pursue to develop a testing procedure for the Administrative Units data specification requirements, and assess to what extent the GML datasets produced with GO Publisher can be labelled INSPIRE compliant / conformant (i.e. propose an ATS). It must be noted though that during the course of the research there has been a very generic ATS template published internally by the JRC, which had the scope of proposing standard test cases, which would be afterwards adapted to all the individual themes. The test cases developed in this chapter are partly based on the mentioned template. In addition, 10 days before the submission of this thesis (16th July, 2012) a newer, more complete ATS template was published by the JRC to the different TWGs that would have to apply and adapt it for each individual theme, and afterwards publish a newer version of the data specifications for all themes with the ATS included. This newer version was also consulted as much as the remaining time allowed.

Although compliance and conformance tend to express the same thing, in INSPIRE there is a tendency to draw a semantic difference between the two terms. This was also one of the reasons the two words were in the title of this research project (i.e. INSPIRE Compliant Datasets – Transformation & Conformance Testing). There is a very ambiguous difference in INSPIRE to what is legally binding and what is not legally binding, and this directly affects the testing process and the verdict of the testing. Semantically, compliance tends to be attributed to a system that has the ability to operate in a way defined by a standard, and represents an informal industry term generally accepted to express that the system provides support for some parts of a given standard. On the other hand, conformance is attributed to a system that has the ability to operate in the way defined by a specification, and its recognition occurs upon formal testing that proves that the system provides 100% support for a given standard (Tóth *et al.*, 2012).

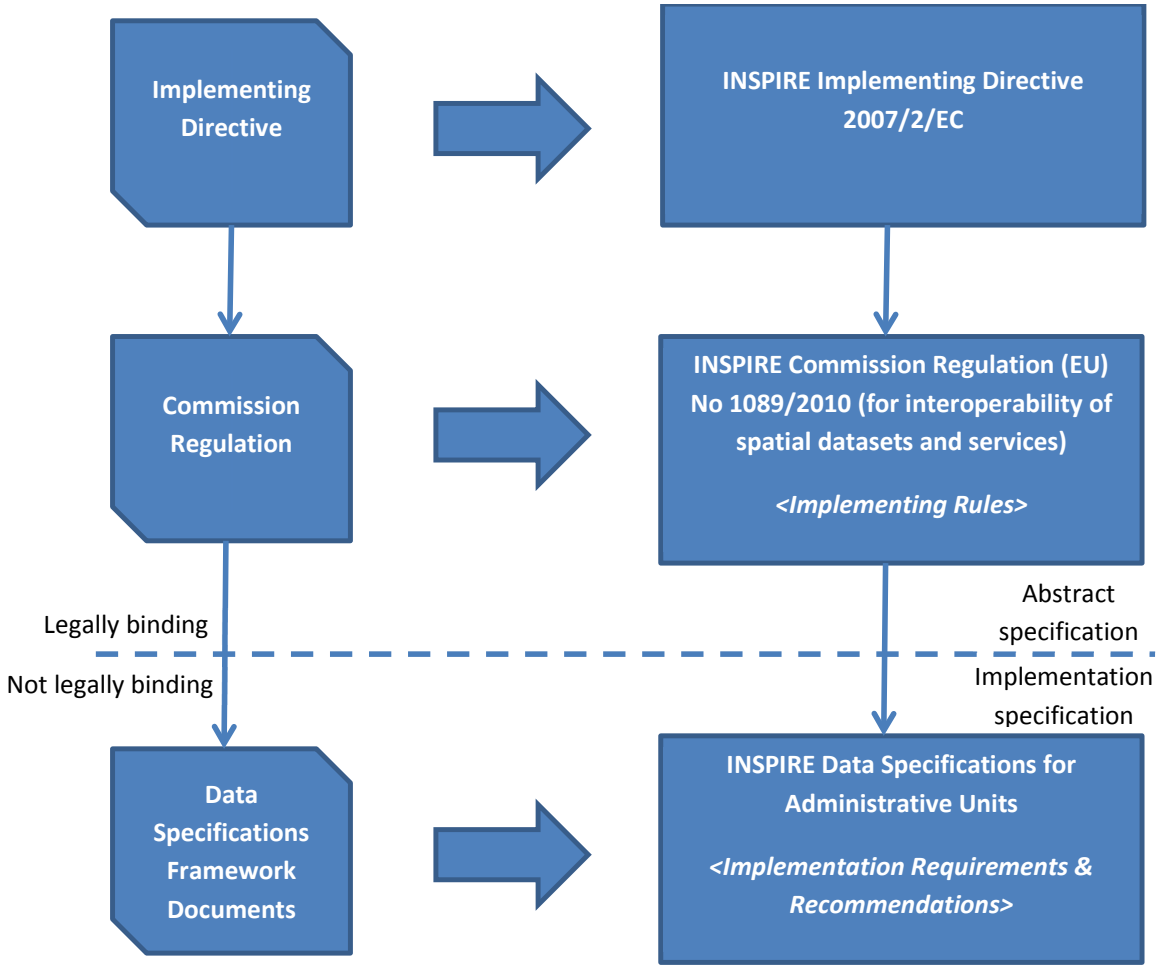


Fig 52: Relationship between INSPIRE Implementing Rules and Technical Guidance (adapted after Vowles, 2012)

Figure 52 above provides a good overview of the fuzzy boundary between what are considered to be legally binding implementing rules that will assure 'legal compliance' with the Regulation (i.e. not very troublesome to implement but not very interoperable either), and not legally binding implementation requirements and recommendations that will assure 'technical conformity' with the data specifications (i.e. more troublesome to implement, not compulsory, but important for true interoperability). It is important therefore to understand the difference between the two terms and also reflect on the consequences of the two approaches.

In the next Section the testing procedure will be presented taking into consideration the existing standards for spatial data compliance testing, but also the ATS template available from the JRC. Section 6.2 and its sub-sections will be presenting the various conformance classes with their afferent test cases applied on the produced GML datasets in the previous chapter. Finally Section 6.3 will provide some conclusions on the developed ATS for the Administrative units theme and its instantiation, which will answer the forth and the last research sub-question.

6.1 Testing procedure

As it has been already explained the Joint Research Centre (JRC) of the European Commission have to prepare a general ATS procedure, which will then be used by the different Thematic Working Groups to develop the ATS for each specific theme. During the course of this research, there has been an intermediary version published internally by the JRC, with a more complete version published internally as well just a few days before the submission of this thesis. Throughout the entire conformance testing implementation on the study case, the guidelines defined in the draft ATS template will be followed where applicable, and customised for the Administrative Units theme. In addition, new elements for conformance testing will be defined and implemented as required by the Administrative Units theme, but also a general review will be given to what is proposed in the current draft ATS version and what is applicable, or missing, in the general context of INSPIRE data specifications, and in particular in the context of the Administrative Units theme.

Testing adherence to the requirements of the Administrative Units data specifications can be grouped within the ATS in several conformance classes, each covering a specific aspect of the specifications (e.g. application schema, reference system, etc.). The same principle can be applied to any other data specifications for the other themes. As it is stated in the INSPIRE Generic Conceptual Model, all test cases that are defined for a specific conformance class, have to be passed in order for the dataset to be conformant to that conformance class. Consequently, for the dataset to be conformant to the entire data specification, it has to pass all conformance classes defined in the Abstract Test Suite. Generally, each conformance class has to tackle a set of requirements from the data specifications. In the Administrative Units data specifications there are 22 requirements and 11 recommendations, with at least the requirements, having to be specifically tackled by the different test cases. The requirements and recommendations defined in the Administrative Units data specifications are attached in Annex B of this document.

After a thorough study on the Administrative Units data specifications and the general draft template on the Abstract Test Suite for INSPIRE data specifications by the JRC, the following conformance classes are proposed:

- Application schema – spatial datasets that fall under the specifications of the Administrative Units theme must be made available according to the spatial object types and data types specified in the Administrative Units application schema (i.e. attributes, associations, constraints, definitions).
- Geometry/Topology – data has to go through essential geometric and topology checks as well as checks required by the data specifications.
- Reference systems – spatial datasets have to be made available using one of the three-dimensional, two-dimensional, or compound coordinate reference systems specified in the data specifications, that make use of the European Terrestrial Reference System 1989 (ETRS1989) datum.
- Data quality – spatial datasets may have to meet target quality results that may be listed in the data specifications.
- Metadata – the metadata describing the spatial datasets must comply with the INSPIRE metadata standard.
- Delivery – conformance of the specified encoding and delivery platform for the spatial datasets.

- Portrayal – although the Directive doesn't formally have any requirements for portrayal, basic rules are necessary. Therefore, for each layer a default style will have to be used.

The division of conformance classes is rather intuitive as they correspond to the different chapters in the data specifications document, and it is worthy to note that structure of the data specification documents has been kept the same for all themes, at least for the finalised ones from Annex I. Not all the conformance classes will be tackled, while some might be tackled partially, in this research project due to the fact that some of the requirements that fall in certain conformance classes are not within the scope of the research.

For instance, metadata conformance refers to the metadata component that should in the end be linked with every INSPIRE dataset. This aspect was not tackled during this research therefore it will not be tested. In INSPIRE, metadata has its own drafting team that developed a separate set of implementing rules for this component on its own. In essence, in a complete conformance testing, the metadata conformance class will have to make sure that the metadata associated with the dataset complies with the INSPIRE metadata standard. The same metadata standard will obviously apply across all INSPIRE themes.

Another conformance class that slightly falls out of the scope of this research is portrayal. Portrayal would apply in-between the INSPIRE Data Specifications component and INSPIRE Network Services component, specifically view services, even though the requirements for portrayal have been included in the data specifications. Furthermore, there is no clearly specified requirement in the Directive about portrayal, but it has been implemented in the data specifications to guarantee that spatial data will be portrayed consistently in the view services, so basic rules are necessary. The view services, as defined in the INSPIRE Directive, are not understood to offer the capability to create high-quality maps, but the basic capability of viewing the data. Although not a priority at this stage, the styles of the layers will have to be implemented in such a way that, combinations of cross-theme layers will provide cartographically good results.

The delivery conformance class is to a certain extent under the same circumstances as portrayal. Most of the requirements in the data specifications refer to network services as the delivery platform of the spatial data. Therefore, a lot of the testing should be performed on network services. The only aspect of the delivery conformance class that could be tested within the scope of this research refers to encoding requirements, which relate to a specific GML version (i.e. 3.2.1) or netCDF.

In the wider context of general conformance to all INSPIRE requirements, an important remark has to be made here, and that refers to another GIMA thesis research project performed by a previous student, which had as primary scope the analysis of INSPIRE network services conformance. Aspects like delivery, portrayal, and to some extent metadata as well, were thoroughly analysed and tested against INSPIRE requirements (Sudra, 2010). This project on the other hand, is oriented more on the actual data transformation process and the end-result of that process, the INSPIRE compliant data.

Therefore, the conformance classes that will receive the most attention during conformance testing in this chapter are application schema, reference systems, data quality, and to a certain extent, delivery, as these classes are highly representative for spatial data in general.

According to the ISO 19105 (Geographic information – Conformance and testing) standard, which is the reference standard for conformance testing in the INSPIRE Generic Conceptual Model, an abstract test case must have the following structure:

- Test case identifier – a certain identification for the test (i.e. name of the test case)
- Test purpose(s) – the definition of the intended scope of the test case
- Test method – description of the testing procedure
- Reference – link to the specific standard(s) requirement(s) or any other material that may be useful during the test
- Test type – clear mention whether the test is performed against an Implementing Rule requirement or a Technical Guidance requirement of the data specification

An abstract test case will be used as the basis for generating an Executable Test Suite (ETS). As already mentioned, the different conformance classes defined will group together several test cases. In the following part of this chapter, the four conformance classes that will be tackled will receive separate sections where test cases will be defined for each of these conformance classes, followed by a test execution and results discussion. As already mentioned, the test cases will partly follow the guidelines proposed in the draft ATS template by the JRC, with comments to be made whether the test cases are clear, complete, and do indeed build up towards declaring a dataset conformant or not to the different conformance classes, and eventually to the entire data specification. Where appropriate, new test cases will be proposed, to complement and improve the quality of the conformance testing.

6.2 ATS/ETS

6.2.1 Application schema

The application schema conformance class is probably the most representative from the entire data specifications, not only for the Administrative Units theme, because it refers to the actual data model defined in INSPIRE, and whether a transformed dataset is really conformant to that data model in terms of attributes, associations, constraints, etc. It must be noted that INSPIRE doesn't restrict the creation of extensions to the original application schemas. Therefore, some data providers may define additional spatial objects and/or data types originating from domain or national data models. When that happens, the last version of the ATS draft template developed by the JRC proposes another conformance class for the extended part of the application schema, therefore clearly separating the original INSPIRE data model from possible extensions. This conformance class would check the exclusiveness of the proposed new elements (i.e. they are not already in the original schema), and if these new elements have been defined according to the INSPIRE Generic Conceptual Model. Such an extension of the original INSPIRE Administrative Units data model is not applicable and not within the scope of this research, but it is important to note that it is implementable and it will also be exposed to conformance testing.

As it was presented during the data transformation chapter, GO Publisher has a validation against the schema functionality. The validity of the entire transformation process, from the user perspective, was based on that functionality. Because this approach might not be considered very

independent (i.e. validating with the same software that the transformation was performed), a third party tool will be used as well at the end (i.e. oXygen XML Editor). Therefore, it could be assumed that once the software validates the output XML mapping against the schema, the dataset can be considered conformant against the application schema conformance class. The test cases to follow will mainly be centred on whether GO Publisher's internal validation engine can be considered reliable enough to decide whether a dataset passes or not the tests defined in the application schema conformance class.

6.2.1.1 Name test

a) Test Purpose	To ensure that the names of each instance of a spatial object type or data type specified in the Administrative Units application schema use the same designation as defined in the application schema section in the Administrative Units data specifications.
b) Test Method	Examine that the designation for an instance of a spatial object type or data type corresponds to the designation provided in the application schema section of the Administrative Units data specifications, by validating against the XSD schema with a XML validator tool (e.g. GO Publisher, oXygen XML Editor).
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

GO Publisher's capability to correctly apply this test case during its schema validation is challenged by purposely feeding it non-existent values for the spatial object type name in one of the produced GML datasets (i.e. administrative boundaries at the 2nd level in the national hierarchy – Regions). Instead of 'AdministrativeBoundary' as the name of the spatial objects, 'Boundary' is inserted. Upon validation of the modified GML data, GO Publisher pops up a message that validation failed, as it can be seen in the figure below.

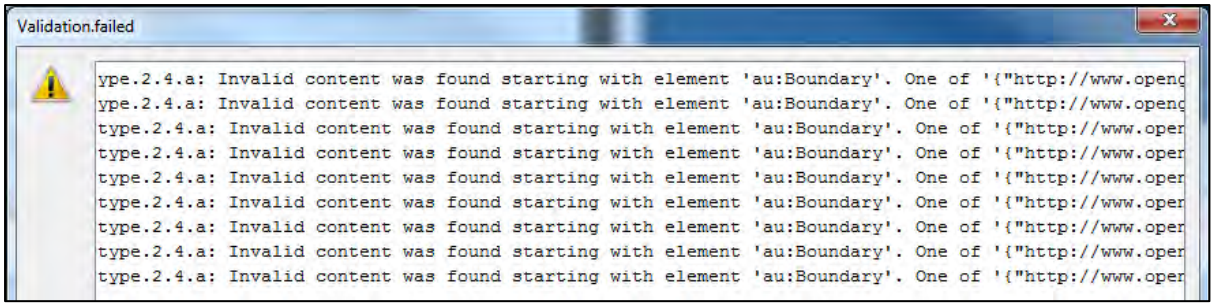


Fig. 53: Spatial object name testing in GO Publisher

There are 9 error messages, one for every spatial object with a non-valid name as the test was performed only on the Regions (level 2 in the administrative hierarchy), which contains 9 features.

6.2.1.2 Attributes/associations completeness test

a) Test Purpose	To ensure that each instance of a spatial object type or data type specified in the Administrative Units application schema includes all required attributes and association roles as defined in the application schema section in the Administrative Units data specifications.
b) Test Method	Examine all instances for the attributes and association roles with the use of a XML schema validator (e.g. GO Publisher, oXygen XML Editor). Each instance shall include all attributes and association roles as defined in the application schema section of the Administrative Units data specifications.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 2, and partly 4 and 5 regarding associations, of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

The same approach will be used here as in the previous test case, by excluding from the mapping project in GO Publisher, certain attributes or associations and apply the validation to assess the behaviour. This time the GML dataset containing the administrative units at the 2nd level in the national hierarchy (i.e. Regions) is used as the sample to be tested. The 'boundary' association and the 'country' attribute are disabled from the GO Publisher mapping project, and then each separately once at a time. The newly produced GML data doesn't validate against the schema anymore as it would've been expected, and the error log clearly specifies the element that is missing, in this case 'country'.

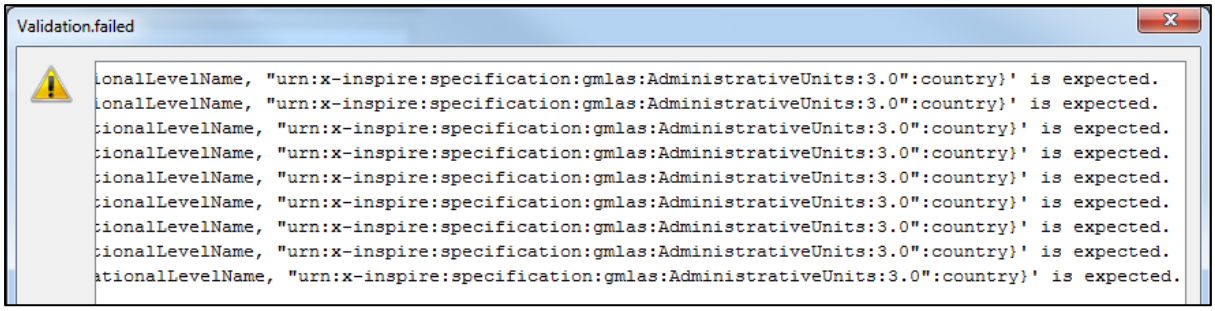


Fig. 54: Attribute completeness test

Again, the multiple errors refer to all features, as in each the 'country' attribute would be missing. It must be noted that Go Publisher's validation engine does not always notify of all the existent errors, in this case the absence of the 'boundary' association. It would alert the errors on attributes/associations/constraint in the order they are specified in the INSPIRE data model. For instance, in this case the 'country' attribute appears before the 'boundary' association in the GML data structure for any given feature. As soon as the error on the 'country' attribute is solved though, upon a new validation, it will then alert the error about the missing 'boundary' association. In some cases this tends to get rather confusing, especially if the user is not very familiar with the target

schema. It is therefore advised to make heavy use of the 'Update Preview' and 'Validate Preview' functionality to closely monitor the validity against the target schema with each change that is made to the schema mapping project. It must be noted that the validation engine does not make a difference between attributes and associations.

In essence, this test case is very similar to the previous one, at least from the behaviour of the GO Publisher validation engine 'point of view', because it doesn't make a difference if a spatial object type or data type has a name that is not specified in the Administrative Units application schema, or if an attribute/association/constraint is missing. For the validation engine both of these situations represent missing elements in the GML instance document, but with different nesting levels.

Another scenario applied in this test case was manually inserting in the GML instance document a random additional attribute (i.e. 'area'), that would for example hold the area of the administrative unit. It must be mentioned that all other required attributes/associations were in place. The additional 'area' attribute was inserted first as the last element, and then as the penultimate element in one of the features in the GML document. In both cases, schema validation failed.

```
      <au:lowerLevelUnit nilReason="Unpopulated" xsi:nil="true"/>
      <au:upperLevelUnit nilReason="Unpopulated" xsi:nil="true"/>
      <au:administeredBy nilReason="Unpopulated" xsi:nil="true"/>
      <au:area nilReason="Unpopulated" xsi:nil="true"/>
      <au:coAdminister nilReason="Unpopulated" xsi:nil="true"/>
    </au:AdministrativeUnit>
  </base:member>
```

Fig. 55: Administrative unit GML/XML fragment – attributes/associations completeness test

When inserted as the last element, the validation engine claims the absence of the 'coAdminister' association element, although this exists in the GML instance as can be seen in Figure 55. The schema validation engine expects the 'coAdminister' element to be the last, and because it is not, it treats it like it is absent. When the additional 'area' attribute is placed as the penultimate element, the schema validation expects the 'coAdminister' association element to follow after the 'administeredBy' association element, hence it fails again. Therefore the attributes/associations completeness test case is really rigid when related to GO Publisher's validation engine, as additional elements will be identified as well and make schema validation fail, although all other required elements might be in place and correctly defined. This is a positive aspect of course.

6.2.1.3 External Object Identifier test

a) Test Purpose	To ensure that the external object identifier for the unique identification of any of the spatial objects in the Administrative Units application schema, has not been changed during its life cycle of a spatial object.
b) Test Method	Compare external object identifiers in previous versions of data with the external object identifiers of the current versions of data for the same spatial object from the Administrative Units application schema
c) Reference	• Article 9(2) in the Commission Regulation (EU) No 1089/2010 • External object identifiers of all data versions
d) Test Type	IR requirement test

This test case applies in circumstances where more versions of the INSPIRE conformant dataset exist. It is not the case in this study case, but it is worth noting that this may also be implemented as an internal constraint for future version of the data. Furthermore, there is no specific reference to this implementing rule is the data specifications, although it is clearly specified in the Commission Regulation. Normally, data specification should be updated to include this additional requirement.

It must be noted that this test will probably only be possible to be performed internally at the data provider's site. One method to implement this test is to create a constraint in the source database that would check if the 'namespace' and 'localid' of the different versions of the data is the same.

6.2.1.4 Multiplicity test

a) Test Purpose	To ensure that each instance of a spatial object type, data type, attribute, and association role specified in the Administrative Units application schema, does not include fewer or more occurrences of a spatial object type, data type, attribute, and association role than specified in the feature catalogue as well as in the UML diagram for the Administrative Units theme.
b) Test Method	Examine that the number of occurrences of each spatial object type, data type, attribute and association role provided by using a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor). The numbers of occurrences for each spatial object type, data type, attribute and association role shall be compared with its multiplicity specified in the schema section of the Administrative Units data specifications.
c) Reference	• INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

Although not specifically mentioned in the Implementing Rules, the multiplicity as a requirement can be indirectly derived from the application schema and the INSPIRE Feature Catalogue. Once again, this test case is based on GO Publisher's schema validation. The intentional error is inserted in the 'nationalLevel' element this time, in the 'AdministrativeBoundary' spatial object. 'nationalLevel' accepts up to six different values according to the Administrative Units application schema. That situation would apply where a boundary section is part of all administrative boundary levels in the national hierarchy (i.e. overlapping). If a 7th value is inserted (see Figure 54 below), schema validation fails as it should, considering that instead of the 7th value, the 'legalStatus' attribute should appear.

```
<au:country>
  <gmd:Country codeList="UK" codeListValue="countryCode"/>
</au:country>
<au:nationalLevel>1stOrder</au:nationalLevel>
<au:nationalLevel>2ndOrder</au:nationalLevel>
<au:nationalLevel>3rdOrder</au:nationalLevel>
<au:nationalLevel>4thOrder</au:nationalLevel>
<au:nationalLevel>5thOrder</au:nationalLevel>
<au:nationalLevel>6thOrder</au:nationalLevel>
<au:nationalLevel>7thOrder</au:nationalLevel>
<au:legalStatus>agreed</au:legalStatus>
<au:technicalStatus>notEdgeMatched</au:technicalStatus>
<au:beginLifespanVersion nilReason="Unpopulated" xsi:nil="true"/>
<au:endLifespanVersion nilReason="Unpopulated" xsi:nil="true"/>
```

Fig. 56: GML/XML fragment – multiplicity test

Although this specific example tackled the multiplicity of an attribute value, the same logic would be applied for multiplicity testing on an association role. For example an administrative unit can be associated to maximum three instances of NUTS regions. Therefore, if a fourth association would be defined, validation would fail.

6.2.1.5 Value type test

a) Test Purpose	To ensure that all attributes or association roles use the value type specified in the Administrative Units application schema.
b) Test Method	Each provided attribute or association role is tested with a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor) to ensure its value type adheres to the value type specified in the schema section of the Administrative Units data specifications.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 2 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

In GO Publisher the value type of the different elements is automatically assigned while the mappings are created from the drop-down list. The elements that can be selected from the list have already the value type assigned to them based on the schema. For example, validation would fail if a simple text attribute is mapped to the 'name' attribute in the target schema. The validation expects to have a 'GeographicalName' data type there, which consequently contains all the other attributes of 'GeographicalName'. It is therefore impossible to assign a wrong value type to the mapped elements in GO Publisher, without having the validation engine detect it.

6.2.1.6 Constraints test

a) Test Purpose	To ensure that the instances of a spatial object type or data type specified in the Administrative Units application schema adhere to the constraints specified in the schema section of the Administrative Units data specifications.
b) Test Method	Examine all instances of data for the constraints specified for the instance's type by means of a Schematron developed in accordance to the constraints specified in the Administrative Units application schema.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 3, 10 and partly 4,5 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

As already described during the transformation methodology there are three defined constraints in the Administrative Units application schema. Two of them are referring to a common-sense rule in regard to the national administrative hierarchy, in the sense that a unit at the highest level in the administrative hierarchy cannot associate a unit at a higher level, and the units at the lowest level in the hierarchy cannot associate other units at a lower level. The third constraint refers to the 'Condominium' spatial object and the fact that it can only be associated with an administrative unit in the first level of the national hierarchy, which is basically the country level.

During data transformation, upper and lower level unit associations were defined for a limited extent, but for the purpose of this test case, some fictive upper level units were defined on the unit at the highest level in the national administrative hierarchy (i.e. country) to see check if GO Publisher validation engine is aware of the constraints. Logically, it would have to check the 'nationalLevel' attribute, and if this has the '1stOrder' value, it should give an error message if an 'upperLevelUnit' association exists.

Validation is passed, which means these constraints are not checked. It must be noted that constraints cannot be validated solely based on the provided XML schema. Therefore, the schema conformance class shouldn't be understood just as a simple validation against the XML schema, which technically, does not provide the capabilities to validate more complex constraints. Especially those that cover more than one attribute or more than one (spatial) object type. To achieve this, a rule-based constraint validation language like Schematron could be used (van Oosterom, 2006). Schematron is capable of expressing these kind of constraints, and can be used in combination with the XML schema or separately. Conceptually, INSPIRE constraints should be translated from a formal constraint language (Object Constraint Language – OCL) to a set of Schematron rules that can be then used to validate the GML document instances. OCL is a declarative language that is now part of the UML standard, and it is used for describing rules that apply to UML models. OCL supplements UML by providing expressions that have neither the ambiguities of natural language, nor the inherent difficulty of using complex mathematics. With the constraints integrated in the conceptual

model, it is a matter of automatically translating them to Schematron rules that can then be used in XML validation tools (van Oosterom, 2006).

In the context of INSPIRE the development of these Schematron rules would normally fall under the responsibility of the JRC, which can then be integrated into the various data transformation software. The version of GO Publisher used in this project (i.e. Desktop edition) does not yet support Schematron validation. According to the developers, this functionality will be implemented in a later version, as it is already available in the Enterprise edition (i.e. GO Publisher Agent). Other XML editors like oXygen, are also capable of integrating Schematron rules in the validation, and have support for creation and editing of Schematron documents.

In conclusion, automated testing of constraints is not possible unless use is made of Schematron or another constraint language. SQL assertions are powerful alternatives to Schematron for implementing constraints, but here as well it is a matter of automatically deriving the constraints from UML/OCL models (van Oosterom, 2006). For a language to be 'Schematron-like' it must be:

- a rule-based system with asserts and report;
- the asserts and reports are evaluated in a context provided by another query;
- the rules from an if-then-else chain, so that there is a lexical priority;
- the rules are combined into a higher-level abstract, in Schematron called a pattern (http://www.oxygenxml.com).

6.2.1.7 Enumeration test

a) Test Purpose	To ensure that the values of all attributes that have an enumeration type are included in the enumeration specified in the Administrative Units application schema.
b) Test Method	The value of each provided attribute with an enumeration type is tested with a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor) to ensure the value is included in the specified enumeration according to the schema section of the Administrative Units data specifications, providing that the allowed values are explicitly defined in the XSD schema.
c) Reference	• Section 4.3 of Annex II in the Commission Regulation (EU) No 1089/2010 • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

An enumeration type is a fixed list of valid identifiers of named literal values. Attributes of an enumerated type may only take values from this list (INSPIRE Thematic Working Group Administrative Units, 2010). In the Administrative Units application schema there are two enumeration data types, both related to the 'AdministrativeBoundary' spatial object (i.e. 'LegalStatusValue' and 'TechnicalStatusValue').

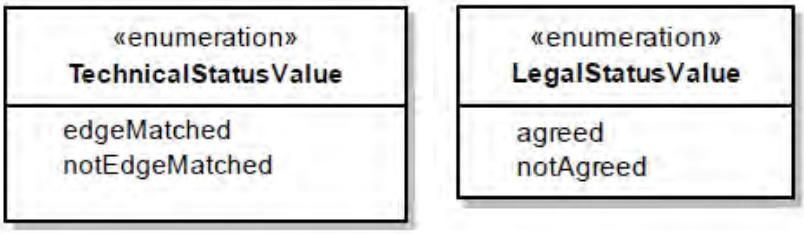


Fig. 57: Enumerations in the Administrative Units application schema

To test if GO Publisher's schema validation engine does take into consideration this test case, a different value than the ones provided by the enumerations is used in one of the produced GML datasets (i.e. instead of 'agreed', 'established' is used). As it is expected, validation fails, with an error message that specifically describes the problem.

Error:[Line: 65,Column: 51]cvc-enumeration-valid: Value 'established' is not facet-valid with respect to enumeration '[agreed, notAgreed]'. It must be a value from the enumeration.

Fig. 58: Schema validation failure upon wrong enumeration value

6.2.1.8 Code list test

a) Test Purpose	To ensure that the values of each attribute that have a code list type, takes only the values that are valid according to the code list's specification as defined in the Administrative Units data specifications.
b) Test Method	Compare the value of each provided code list type attribute with the values provided for the code list in the Administrative Units application schema.
c) Reference	• Section 4.4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Section 7 of Annex II in the Commission Regulation (EU) No 102/2011 (amendment to Commission Regulation (EU) No 1089/2010) • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

Code list is a data type very similar to the enumeration data type presented in the previous test case, being defined as a flexible enumeration that uses string values for expressing a list of potential values, and according to the INSPIRE implementing rules, *"code lists means an open enumeration that can be extended"* (European Commission, 2010). In the Administrative Units application schema there is only one code list, 'Administrative Hierarchy Level'.

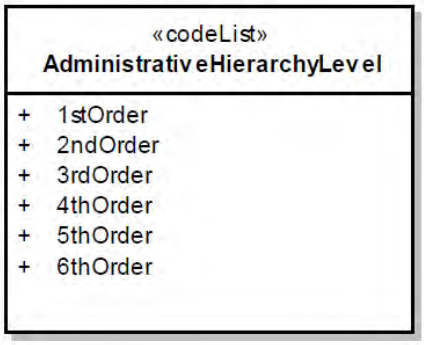


Fig. 59: Administrative Hierarchy Level code list

The INSPIRE implementing rules state that there may be code lists that cannot be extended by Member States and code lists can be extended by Member States. The attributes whose value type is one of the non-extendable code lists, must only take values from the lists specified for the code list. Although in the first version of the 'Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services' the values for the 'Administrative Hierarchy Level' code list weren't specified, this has changed with an ulterior amendment that specifically lists the allowed values (European Commission, 2011). Essentially, this means that the test case for the 'Administrative Hierarchy Level' code list will be executed in a similar manner with the enumerations from the previous test case. When randomly changing the value with any string that doesn't exist in the 'AdministrativeHierarchyLevel' code list for the 'nationalLevel' attribute, GO Publisher's schema validation still interprets the GML instance document as valid and conformant to the schema. This is a flaw of the validation process, which will probably be fixed in the future versions of the software, but it is important to note that GO Publisher's schema validation engine cannot be relied upon entirely, and some manual checks have to be performed on the code list values used.

6.2.1.9 Co-administration test

a) Test Purpose	To ensure that, where applicable, if an administrative unit is co-administered by two or more other administrative units, both 'administeredBy' and 'coAdminister' associations are used as specified in the Administrative Units application schema.
b) Test Method	Check if 'administeredBy' or by case, 'coAdminister', provide associations to the correct administrative units as defined in the application schema section of the Administrative Units data specifications.
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 6 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

As in the study case used for this project (i.e. administrative units in England) there is no known case of units administered by two or more other administrative units so the requirement doesn't really apply. If it would though, the relation between the two associations could be tested using the same Schematron-based approach as for the other constraints. In essence, this represents a constraint as well because if one of the two associations is made, the inverse one has to be made as well. If a feature has a valid association for the 'administeredBy' or 'coAdminister', then the referenced feature has to be checked for an inverse association in the other element. It is important to note that this is different from the regular associations test in Section 6.2.1.2. If any of these associations receive a void reason value, then no further action is needed.

6.2.2 Geometry/Topology

The geometry/topology structure conformance class is as important as the application schema conformance class, being as well very representative for a spatial dataset. While the application schema conformance class validates the descriptive part of the data, the data structure validates the physical aspect of the data, the geometric representation. These two parts coincide with the foundation on which spatial data in general is built upon. In the data specifications, the requirements and recommendations of both application schema and data structure are placed in the same chapter. For conformance testing there is a need to separate these components as they need a different testing approach.

There are several requirements and recommendations defined in the data specification for Administrative Units related to data structure, most of them coinciding with regular polygon geometry checks but also some specific requirements/recommendations related to administrative units and their boundaries

6.2.2.1 Administrative units geometry test

a) Test Purpose	To ensure essential polygon / multi-polygon geometry consistency is maintained.
b) Test Method	Execute essential geometric check on the administrative units polygon / multi-polygon geometry.
c) Reference	• Not specifically mentioned
d) Test Type	TG additional test

The most straight-forward method to perform geometry/topology checks on the published administrative unit GML dataset is to load it back in the database and then make use of available spatial functions. The Oracle function used to validate geometry is 'SDO_GEOM.VALIDATE_GEOMETRY_WITH_CONTEXT', which performs a consistency check for valid geometry types and returns context information if the geometry is invalid. If the geometry is valid, it will return 'TRUE'. Some of the geometry consistency elements that are checked for polygons include:

- Polygons have at least four points, which includes the point that closes the polygon.
- Polygons are not self-crossing.
- No two vertices on a line or polygon are the same.

The function returned a 'TRUE' value for all features, therefore, the administrative units polygon geometries can be considered valid.

6.2.2.2 Administrative units overlapping test

a) Test Purpose	To ensure that no administrative units established at the same level in the national administrative hierarchy overlap (i.e. their boundaries should not intersect with each other)
b) Test Method	Execute spatial functions/queries that would check if any two or more administrative units established at the same level in the national administrative hierarchy overlap
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 7 and Recommendation 3a of the AU data specifications
d) Test Type	IR requirement test

This test can be executed in the Oracle database, with the Topology Data Model extension. This is also a feature that is not available in the Express Edition, the full Oracle Spatial version being required. The Topology Data Model functionality can also be implemented as a permanent constraint in the database where the data is kept, which will permanently ensure the database doesn't store administrative units at the same level in the national administrative hierarchy that overlap. Alternatively this test can also be performed in ArcMap with the available topology functionality that allows defining several topology tests that would assure the integrity of the data.

6.2.2.3 Administrative units coverage test

a) Test Purpose	To ensure that, together, all administrative units established at the same level of national administrative hierarchy, cover the whole territory (i.e. country level)
b) Test Method	Execute spatial function/query that would check if all administrative units established at the same level of national administrative hierarchy have an equal topological relationship with the administrative unit established at the highest level (i.e. country level)
c) Reference	• Not specifically mentioned
d) Test Type	TG additional test

This test can also be implemented by the same means as the previous test case (i.e. Administrative units overlapping test), but defining of course the appropriate topological relationship. Another option is to make use of the 'SDO_EQUAL' function that checks for equal topological relationships between two datasets, one of them in this case having to be the administrative unit representing the country.

A very important observation has to be made here that refers back to the administrative geography of England that has been thoroughly described in Chapter 3 (see Figure 19 – page33). The counties

for instance, which are established at level 3 in the national administrative hierarchy, do not cover the entire extent of England. Such particular situations have to be taken into account before executing this test, as this does not necessarily mean that the test fails especially that INSPIRE doesn't claim that the units at a certain level in the administrative hierarchy have to cover the entire extent of the country. When such situations occur, it is better to drop this test in order to avoid confusion.

6.2.2.4 Administrative boundaries geometry test

a) Test Purpose	To ensure essential polyline geometry consistency is maintained.
b) Test Method	Execute essential geometric check on the administrative boundaries polyline geometry.
c) Reference	• Not specifically mentioned
d) Test Type	TG requirement test

This test is executed in a similar manner with the essential administrative units geometry test, by using the same ORACLE function, 'SDO_GEOM.VALIDATE_GEOMETRY_WITH_CONTEXT'. Some of the geometry consistency checks that are performed on polyline geometry are:

- Line strings have at least two points.
- Points on an arc are not collinear (that is, are not on a straight line) and are not the same point.

6.2.2.5 Administrative boundaries bounding test

a) Test Purpose	To ensure that the border that limits any given administrative unit shall correspond to the geometry representing the boundary of that administrative unit.
b) Test Method	Execute spatial functions/queries that would check if the border that limits the administrative units is corresponding to the geometry representing the boundaries of the administrative units.
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 8 and Recommendation 3d of the AU data specifications
d) Test Type	IR requirement test

This test refers to the topological relationship between administrative boundaries and administrative units, and in essence it is performed similarly with the administrative units overlapping test, by either using Oracle's Topology Data Model Extension, or the topology functionality in ArcMap, if the data is going to be stored in an Esri Geodatabase (i.e. File or Personal Geodatabase). It also acts as a constraint on the database and it ensures the integrity of the data permanently.

6.2.2.6 Condominium test

a) Test Purpose	To ensure that the spatial extent of any defined condominium spatial object is not part of the geometry representing the extent of any given administrative unit.
b) Test Method	Execute a spatial intersection between any defined condominium spatial object and the entire 'administrativeUnit' dataset. The result should be 'false'.
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 9 of the AU data specifications
d) Test Type	IR requirement test

Since there are no condominiums in the case study, it is not necessary to execute this test. When there would be condominiums, the same approach as in the previous test cases can be followed, by loading the 'condominium' GML instances back into the Oracle database and setting a spatial query that will verify if the 'condominium' geometry intersects with the 'administrativeUnit' geometry.

Regarding the entire geometry/topology conformance class it is possible to add further test cases that have the scope of assuring the integrity and consistency of the data, taking in consideration the topological relationship between administrative units and administrative boundaries. For example, one test case could verify there are no gaps between adjacent administrative units. In essence this is the opposite test of the one defined to check if there are is any overlapping between administrative units at the same level in the national administrative hierarchy. Regarding boundaries, one test case could also check if there are any dangles, boundaries having to always divide different administrative units.

6.2.3 Reference system

The coordinate reference system (CRS) in the INSPIRE context is bound to provide a harmonised specification for uniquely referencing spatial information because all users of spatial data need geodetic reference for the data to be in place. Therefore, the harmonisation of the coordinate reference system usage is a crucial prerequisite for the successful harmonisation of the datasets belonging to all INSPIRE themes (INSPIRE TWG CRS, 2010). This operation obviously implies a transformation from the CRS of the source data (i.e. national CRS) to the CRS of the target data model (i.e. ETRS89). CRS transformation is a complex mathematical operation which will enable coordinates in one coordinate system to be transformed into coordinates in another system, and vice-versa. There are many methods to perform this task, with varying degrees of accuracy, and this is due to the fact that CRS transformation doesn't happen to be an exact science (which is also due to the constant change in the Earth shape over time), therefore such a transformation should have an accompanying accuracy flag associated with it.

This conformance class is primarily set to check if the transformed data is making use of one of the coordinate reference systems specified in the implementing rules. What is maybe even more relevant is the quality, or the accuracy of the CRS transformation. There is no specification concerning this matter in neither the implementing rules nor the technical guidelines, thus this aspect will be emphasized more upon after the presentation of the test case.

6.2.3.1 Datum / CRS test

a) Test Purpose	To ensure that each instance of a spatial object type specified in the Administrative Units application schema is given with reference to the European Terrestrial Reference System 1989 (ETRS89) datum in areas within its geographical scope, or the International Terrestrial Reference System (ITRS) datum, or other geodetic coordinate reference system compliant with ITRS in areas that are outside the geographical scope of ETRS89.
b) Test Method	Check that each instance of a spatial object type in the Administrative Units application schema has been expressed using a coordinate systems using one of the datums specified above (i.e. in the Test Purpose).
c) Reference	• Section 1.2 and 1.3 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 12 and 13 of the AU data specifications
d) Test Type	IR requirement test

The use of the correct coordinate reference system can be checked manually of course, but it might be a good idea to implement validation of the used CRS by inlcuding Schematron rules in the application schema. In the GML structure, in the 'geometry' element, there is a 'srsName' attribute, which value is a Uniform Resource Identifier (URI). It refers to a definition of the CRS that is used to interpret the coordinates in the geometry. The CRS definition may be in a document or in an online web service. It also contains the EPSG (European Petroleum Survey Group) code. This is technically the SRID code of the used CRS in the geometry. These codes are maintained in the CRS Registry Service operated by the Oil and Gas Producers Association (OGP at http://www.epsg-registry.org). In the context of INSPIRE it might be possible to maintain the EPSG codes of the approved to use CRS in an INSPIRE CRS registry, and create a constraint that the EPSG value in the 'srsName' attribute can only be one of those that are specified in the registry. This is similar to the logic behind a code list, therefore, it would be possible to test the usage of an allowed CRS via Schematron rules. Below is a GML sample depicting the part of the instance where the CRS code is specified, in this case being 4258, which stands for the generic ETRS89 CRS, suitable for any European region; with the mention that the GML fragment refers to the geometry of the 'residenceOfAuthority' attribute. This is a point and only requires two coordinate values, being easier to illustrate here.

```
<au:geometry>
  <gml:Point srsName="urn:ogc:def:crs:EPSG::4258" gml:id="LOCAL_ID_2">
    <gml:pos>51.387890740283254 -0.2869144497634939</gml:pos>
  </gml:Point>
</au:geometry>
```

Fig. 60: GML fragment depicting 'srsName' value

As already discussed, there are no explicit accuracy aspects specified, which is also the case for the ISO 19111 (Spatial referencing by coordinates) standard that is the basis for CRS specifications in INSPIRE. In the CRS specifications it is noted that it has been developed for geographic information in general, but not for precise positioning, with the mention that specifications for an improved accuracy might be considered in the future, and that some INSPIRE themes may indeed benefit from

that (INSPIRE TWG CRS, 2010). Although this aspect is very application dependant, it might grow to become an important issue in INSPIRE in general and in the data transformation in particular. For some of the data themes, the accuracy of the CRS transformation is crucial, especially when taking into consideration cross-border data usage. Administrative Units is one of those themes, giving that it is supposed to serve as a reference for correct location of other spatial objects, and delimiting areas of competent authorities. The case of England, and the UK in general, where the study case of this research is based, tends to be somewhat less complicated because as argued in an earlier chapter, there are no neighbour countries. Therefore, the additional problems that may arise from cross-border inconsistencies, even after data was harmonised to the same data model and coordinates transformed to the same CRS, are avoided here to a certain extent. Nevertheless, there are certain aspects that are relevant and can still cause problems even within the same country. Also differences may appear even when performing the same CRS transformation but with different software.

To demonstrate that, the source data is exposed to the same CRS transformation (British National Grid to ETRS89) in FME. It must be noted that only the CRS transformation is performed, which is enough to assess the differences.

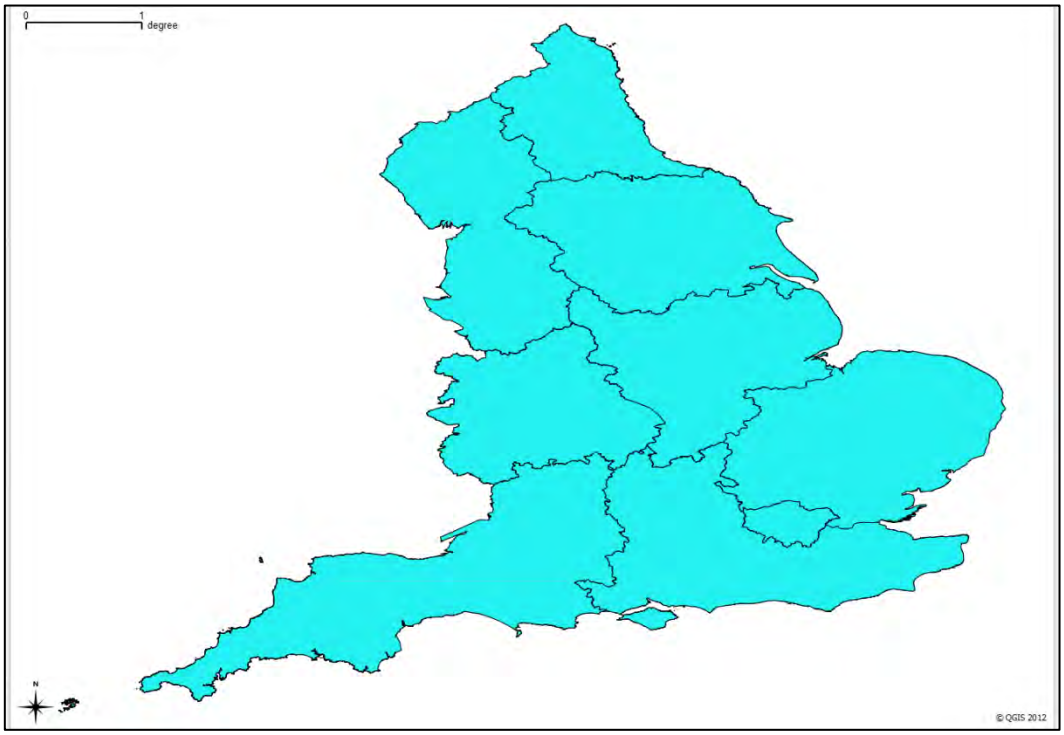


Fig 61: England Regions in ETRS89 reference system

Figure 61 above, illustrates the same dataset (i.e. England Regions) using the ETRS89 reference system and loaded twice, one version being the output from the GO Publisher transformation, and the other one only suffering the CRS transformation in FME. At the scale that it is presented, the geometry seems to be perfectly aligned, but when a spatial query is performed to check if the geometries are identical, the result was that they are not.

Fig. 62: CRS transformation differences

As it can be seen in Figure 62, there is an offset between the two datasets, which seemed to be an average of around 2.4 meters between FME and GO Publisher transformations. It is difficult to assess if that is a high or low offset because it depends how and for what scope the data is used, but it possible that for some applications it might be unacceptable, while for some this will not constitute a real problem. Either way, this is an aspect that has to be taken into account when performing data transformation that also implies CRS transformation, as is the case in INSPIRE, especially when the source data comes from different national CRS. Furthermore, an Oracle coordinate transformation function was used as well to compare differences between three options (FME, Oracle, GO Publisher), in order to have a better understanding of the variations that can occur. These are presented in Table 6 below:

Table 6: CRS transformation differences

Tools used for CRS transformation	**CRS transformation difference**
GO Publisher / FME	2.4 m
GO Publisher / Oracle	0.3 m
FME / Oracle	2.7 m

It must be noted that these figures were obtained by visually inspecting and using the line measure tool, to measure the offset of the CRS transformation for the same sample point. For a more thorough analysis it is recommended to use Root Mean Square Error (RMSE) calculation for coordinate transformation. This involves strategically selecting a number of sample points, while the differences measured between these points with each CRS transformation, are the input of the

RMSE equation, which can then provide an average offset across the entire dataset. The scope here was not to provide an exact accuracy figure, but to emphasize the variations that can occur with the CRS transformation using different tools. Nevertheless, the RMSE for CRS transformation can be the accuracy flag that is recommended to accompany the CRS transformation.

There also few other requirements would fall in the reference systems conformance class, but these will not be tackled here as most of them refer back to the CRS that the data have to be published into (i.e. the CRS to be used in View Services, identifiers for CRS, and temporal reference systems). The issue of the CRS identifiers also refers back to the proposal of validating the CRS as a constraint using Schematron rules. JRC will probably publish at a later data a CRS register that contains CRS allowed in INSPIRE that will use identifiers based on the EPSG codes. Also, the general ATS template published internally by the JRC defines test cases for all these requirements, as most of them are standard for all INSPIRE themes. For example, these would check the validity of compound CRS (when this is used), View Services CRS test, grid definition test, and grid CRS test. As it has also been specified in the introduction chapter of the thesis, the scope here to solely assess the CRS transformation to ETRS89 and analyse and understand the accuracy flag that is associated with the transformation.

Therefore, based solely on the implementing requirements, the published datasets can be declared conformant to the reference systems class, as long as the CRS used is ETRS89. It must be stressed out though, that CRS transformation comes with a certain accuracy flag, which can differ even for the same transformation on the same dataset, but using different software. This is an open issue at the moment that will have to be addressed in the future.

6.2.4 Data quality

INSPIRE does not set any specific implementing rules on data quality as it would've been expected from the chapter in the data specifications, at least not for the Administrative Units theme. This part does indeed refer to some quality elements but they fall under the metadata component. It includes a description of data quality elements and sub-elements as well as the associated data quality measures to be used to describe data related to the spatial data theme Administrative Units. The quality elements and sub-elements are:

- Completeness
    - Commission
    - Omission
- Logical consistency
    - Topological consistency
    - Conceptual consistency
- Positional accuracy
    - Absolute external positional accuracy

In the data specifications it is recommended that aggregated data quality information should ideally be collected at the level of spatial object types and included in the dataset metadata. Therefore, data quality elements become an additional and optional part of the metadata for the Administrative Units INSPIRE theme. As metadata is not within the scope of this research, it is concluded that at least for the Administrative Units theme there are no requirements regarding data quality of the transformed dataset. This as well might become an issue in the future, in situation

where a Member State might submit data that does not rise to the standards required for specific applications.

6.2.4.1 Data quality target test

In the case of other themes, where data quality requirements may be specified, the following test case is defined.

a) Test Purpose	To ensure that all data quality elements meet the specified target results in the data quality section of the data specifications.
b) Test Method	Compare the results of the data quality measure of the transformed dataset to the target proposed result specified for each data quality element specified in the data quality section of the data specifications. Results of the data quality measure have to obviously be equal or higher than the specified target result of the data quality element.
c) Reference	• Data quality section of the data specifications for the respective theme
d) Test Type	TG requirement test

6.2.5 Delivery

There are two aspects related to the delivery of INSPIRE conformant datasets: the delivery medium and the delivery format. INSPIRE Network Services represent the delivery medium, while the delivery format is the aspect that is going to be further discussed within this conformance class, since Network Services do not fall within the scope of the research. The delivery format refers to the encoding of the published datasets. The implementing rules do not have a clear requirement in this sense, but the data specifications in general, and consequently for the Administrative Units theme recommend using GML (version 3.2.1) as the encoding format.

It is important to note that GML is considered an exchange format, which can commonly be used to pass spatial data between different systems. In theory at least, GML could also be used as a native file format directly into a GIS (for processing and analysis). As it stands at the moment, GML as a spatial data exchange format is different in concept and require different approaches compared to a traditional spatial data file format, like shapefiles for instance. The regular data flow between two systems, therefore, would be that system A publishes its data from whatever format is sitting on into the GML format according to the INSPIRE data model, and system B will grab that GML format and translate it to its own data model and preferred data format.

6.2.5.1 Encoding test

a) Test Purpose	To ensure that the used encoding is conformant to the encoding provided in the delivery section of the Administrative Units data specifications.
b) Test Method	Check whether the provided encoding is conformant to the encoding for the Administrative Units application schema.
c) Reference	• Administrative Units XSD schema provided
d) Test Type	TG requirement test

Since it is assumed that the Administrative Units data specifications are compliant to the ISO 19118 (Geographic Information – Encoding) as it is required in the implementing rules, the only aspect that needs to be tested in the encoding test is conformance with the specified encoding. In the case of the GML encoding, this assumes validation against the provided XSD schema provided with the specifications. The validation can be performed with the GO Publisher validation engine, and the result is that the published GML datasets are valid against the XSD schema.

Therefore, this test case is tightly connected with the application schema conformance class, where several test cases where also relying on the validation against the schema. What this means is, that by passing the encoding test case, many of the application schema conformance class test cases are also passed. However, the difference occurs where the individual elements related to the application schema conformance class (e.g. attributes/associations) are correctly provided, but the encoding is not correct, and there are two aspects to it: XML/GML must be well formed, and valid against the application schema.

6.3 Evaluation and conclusions

This section will draw some conclusions regarding the conformance testing procedure and results, of the produced INSPIRE GML datasets. It will mainly answer the fourth research sub-question, which should then be able to blend with the conclusions from the previous chapters and provide a solid base for the general discussion of this research project and provide final conclusions, and recommendations for future research.

Conformance testing consisted of the identification of several conformance classes for which test cases where specified according to the implementing rules and requirements in the Administrative Units data specifications. In practice, conformance classes represent common requirements to all data themes, while the test cases have to be customized for each individual theme. The collection of conformance classes with their test cases represents the Abstract Test Suite (ATS) specification for the Administrative Units theme. The test cases were then executed, based on the test method recommendation, which generated the Executable Test Suite (ETS). The ETS is not necessarily a very self-contained operation, as the method to execute the defined test cases varies a lot from one conformance class to another and even within the same conformance class, depending on what has to be tested. For instance within the application schema conformance class, most of the test cases refer to the content of the produced GML datasets, and therefore are tested with the built-in GO Publisher schema validation capability, as well as with additional, external and therefore more independent, schema validation software (i.e. oXygen XML Editor).

In addition, it must be noted that the concept of conformance classes and test cases was not defined in the data specifications, which only refers to the requirement to pass the abstract test suite defined in the annex of the document. The true meaning and concept of ATS / ETS, conformance classes, and test cases had to be researched further in external resources. It might be a good idea to describe these terms in each data specification document.

As INSPIRE implementing rules and data specifications for each individual theme inevitably contain requirements and recommendations that relate to other branches of INSPIRE, like Network Services or Metadata, the process of conformance testing in this research project was focused at the elements that are highly representative for spatial data on its own, and on the results of the transformation process presented in Chapter 5 (i.e. Producing INSPIRE conformant data). Therefore, it is not possible to claim the produced GML datasets are fully INSPIRE compliant, but it is possible to claim conformance with the proposed conformance classes, which may result in a partial conformity. In addition the current version of the draft general ATS template developed by the JRC also contains template test cases that are more oriented towards checking aspects related to Network Services or Metadata, even within the conformance classes that have been approached in this chapter, especially within the reference systems and delivery conformance classes. Some test cases, even if within the scope of the actual data conformance, were simply not applicable in the Administrative Units theme because they referred to coverages.

6.3.1 Application schema

Testing the produced GML instances against the application schema conformance class was for most test cases fairly straight-forward, due to GO Publisher's internal validation engine. Some constraints could not be tested however. For this, constraint validation using Schematron could be used in the future, or as an alternative, SQL level validation procedures. At the moment GO Publisher doesn't support Schematron integration. Nevertheless, the produced GML datasets are valid schema-wise because constraints checks weren't really necessary since there were no condominiums, and the highest/lowest level constraint was not applicable because upper and lower level associations were only implemented on a limited extent (due to them being voidable), and at a level that is not the highest, nor the lowest in the administrative hierarchy. For a more independent testing the output was also validated with success in oXygen, and in general it is recommended to execute the tests in a platform as independent as possible from the one the data has been initially transformed into.

6.3.2 Geometry/Topology

In regard to the geometry/topology conformance class, tests were possible by loading the produced GML instances back in the Oracle database and using Oracle's internal spatial functions. Some of the test cases had to be executed in ArcMap, especially those referring to topological relationships between the administrative units and administrative boundaries, as the (free) Oracle database edition used in this research had limited spatial capability. One test case that was not executed refers to the condominium geometry which cannot be part of the geometry representing an administrative unit. Since there were no condominiums, and the other checks were passed, the produced datasets have passed the topology/geometry conformance class. It also must be noted that the general draft ATS that the JRC is working on at the moment does not contain a topology/geometry conformance class, which comes as a surprise since most themes will have at least some basic geometry/topology requirements that are worth checking before declaring a dataset conformant. It might seem that these checks fit better in the data quality conformance class, but in the data specifications, requirements related to data geometry/topology are clearly separated from data quality requirements, which anyway refers to optional metadata elements holding data quality information.

6.3.3 Reference systems

Solely based on the data specifications requirements, the produced GML instances are valid against the reference systems conformance class, because they are published with the ETRS89 coordinate reference system. The biggest issue of CRS transformation lies on the accuracy of the transformation, an aspect which is not tackled in the specifications. There should be a certain accuracy flag associated with every single CRS transformation, becasue results can vary a lot as it has been demonstrated. Of course accuracy may be of a higher or lower importance, depending on the scope of the data and the application the data is used in. It might be worthy taking into consideration to define more concrete requirements for coordinate transformation, which could be integrated by software providers that claim to offer INSPIRE solutions.

6.3.4 Data quality

The most ambiguous element of testing is probably data quality as there are no requirements in this sense for Administrative Units, nor in general for the other themes. Data quality chapters in the data specifications in general, not only Administrative Units, refer to descriptive information that may be added as an extension to the metadata of the transformed dataset.

6.3.5 Delivery

Finally, the delivery conformance class, which can only provide one test case relevant for this study case, as the other requirements refer to the delivery medium of the transformed datasets (i.e. Network Services), is also passed by the produced datasets after validating against the XSD schema with GO Publisher's internal validation engine and oXygen's schema validation. This is a similar test to the ones in the application schema conformance class. As already mentioned during the test case, most of the application schema conformance class test cases are passed if the encoding test is passed, the purpose of the encoding test is to avoid situation where different application schema elements are correct but the encoding is not correct.

6.3.6 Conformance and compliance

To conclude, strictly judging by INSPIRE requirements, it can be stated that the GML datasets that were the output of the data transformation in Chapter 5, are conformant to all the proposed conformance classes excluding the test cases that were not possible to execute due to software limitation, and therefore, are partially conformant to the entire INSPIRE and Administrative Units specifications; requirements for Metadata and Network Services not being considered. Despite the issues encountered during data transformation in Chapter 5, it was nevertheless possible to achieve partial technical conformance that could be verified using the described ATS/ETS conformance tests in this chapter. This conclusion is also very relevant in the context of the difference between what is legally binding, therefore legal compliance, and what is not legally binding, therefore technical conformance. Obviously, by acknowledging the difference, it doesn't mean that only the absolute minimum requirements have to be implemented, and it is strongly advised that data providers should aim for complete (technical) conformance, because it is the best way to achieve a high level of interoperability. At the same time, there might be data providers that will seriously struggle to achieve that target, and therefore the difference between legal compliance and technical conformance should be made as clear as possible, otherwise conforming to INSPIRE might be perceived as an almost impossible task.

7. Summary and conclusions

This final chapter provides an overview of the taken steps and achieved results by summarising all the observations made in the concluding sections of the other chapters in Sections 7.1 to 7.3, and will provide concise answers to the research questions in Section 7.4. Recommendations for future directions, but also suggestions regarding the encountered issues are made as well in Section 7.5.

7.1 INSPIRE and data interoperability

The INSPIRE Directive sets a framework for geographic information interoperability across Europe, specifically focusing on European Union Member States. The initiative proposes the harmonisation of various thematic datasets to standardised data models. It is not only a technically challenging task, but also policy-wise, with many political and legal aspects involved. Nevertheless, INSPIRE aims at interoperability of spatial data across Europe, being reliant on international standards and industry recognised best practices, being, conceptually at least, a big step forward and a solution to many traditional (geographical) information systems issues. Therefore, the question is how straightforward is the implementation of the INSPIRE requirements and recommendations, what are the current bottlenecks, and what can be done to make the implementation process as smooth as possible.

This thesis aimed to tackle the above aspects purely from the data model side. Although the case study is centred on one of the less complex data specifications of INSPIRE (i.e. Administrative Units), there were still several issues encountered, both in the data transformation stage as well as in the conformance testing stage. The research is based is on an extensive literature review that gave the necessary background information for the implementation of the case study, which is then followed by an in-depth study of the INSPIRE documentation, especially the data specifications document for Administrative Units. Studying the INSPIRE documentation and understanding the requirements for the implementation is considered by many a daunting task. In general, data specifications are published as a combination of documents, UML diagrams and XML schemas; therefore, solid knowledge of XML and UML is essential.

In addition, in the data specification documents there is a very fuzzy boundary between legal requirements and other technical requirements/recommendations, which is a very important aspect when it comes to judging legal compliance of a dataset. To make a clear distinction between legally binding aspects and not legally binding aspects, for this research not only data specifications had to be studied, but also the general European Commission Directive (i.e. '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'), and the European Commission Regulation in regard to spatial datasets interoperability (i.e. 'Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services') and its amendment (i.e. 'Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regard interoperability of spatial data sets and services'). These legal documents are even harder to interpret, but in conjunction with the data specifications, are a more complete source of information.

7.2 Gap analysis and data transformation

An important aspect to be taken into account in the preparation of the data transformation is related to the source data. It is crucial for the quality of the transformation to have very good knowledge about the source data. In the case study of this thesis a lot of time had to be allocated to understand the source data, in this case the issue being related to the structure of England's administrative geography which is quite irregular and not uniformly nested across all levels of national administrative hierarchy. But maybe the most important issue that can occur, and this is a general remark, is finding the semantically related attributes between the source and the target data model (i.e. schema mapping). This is why it is very important to have domain experts involved in the schema mapping process.

Of course, the success of the schema mapping and resulting data transformation is also heavily dependent on the amount of information the source data holds and to what extent this can be used to populate all required attributes in the target data model. In the case study of this thesis the source data was selected from an open data scheme of Ordnance Survey that provides the administrative units data in polygon geometry. Therefore, boundary data had to be derived from the administrative units and then processed to achieve the geometric structure defined in the specifications. These types of problems are not necessary generally encountered as the actual data provider would maybe have access to more complex (topological) models of the dataset, but there might be cases where this is a problem even for the data provider. For exactly this reason, many attributes or associations in the INSPIRE data models are voidable, which means they are not compulsory. This should give the possibility for many data providers to still be able to comply to INSPIRE, but it is also a two-edged issue because some data providers might use this approach as a shortcut to deliver conformant data. It is then a question of the real value of the produced data and of INSPIRE in general, if everybody will deliver the minimum required. At this stage though, the immediate objective of INSPIRE seems to be to get data providers to share the data first of all, even if complying at minimum level to the requirements. This is a general pattern that has also been directly observed in several INSPIRE related meetings/workshops that have been attended over the course of the research. These awareness and consultation events were generally organised or co-organised with the UK Location Council, which is the Geonovum equivalent in the UK, and had various experts and UK data providers involved, so the outcome is extremely relevant for the INSPIRE implementation in the UK.

When it comes to software tools, although the case study in the thesis relied on Snowflake Software GO Publisher, Safe software FME is also a viable alternative, referring strictly to transformation from a source data model to an INSPIRE target data model, and publishing the data into the required GML format. In essence, these two packages are also the only really viable solutions at the moment on the market to implement INSPIRE data requirements, with maybe a slight advantage for GO Publisher as it has a better integration for GML/XML, as well as database management, than FME does. Nevertheless, for additional spatial data processing, an external GIS package might be needed or a database with powerful spatial functions, like Oracle Spatial. The Express edition that was actually used during this implementation was not powerful enough.

During the data transformation several issues were encountered, most of them related to missing information in the source data. One important remark that has to be made in this respect, is that it has proved unfeasible to consider that to transform the data to the INSPIRE data model and cover all

attributes/associations in the application schema, it is enough to just have source data that contains administrative information in the case of Administrative Units. Many INSPIRE application schemas contain attributes or associations that refer to data models from different themes, so the source data will often belong to different 'domains' and consequently must be collected from different data providers. An example is the association to NUTS regions in the Administrative Units data model, which belongs to a different INSPIRE theme (i.e. Statistical Units); or the 'residenceOfAuthority' data type which requires a separate dataset as input to the transformation. It is also true that these elements, at least in the Administrative Units applications schema, are voidable, but this is an aspect that needs to be considered, especially when the aim is to be fully conformant. It therefore didn't come as a surprise when some of these aspects were not possible to implement during the case study of the thesis, or if implemented, it was done on a limited spatial extent of the data.

The second main limitation for the transformation was the lack of some information that directly concerns the data provider, like life cycle information for instance or administrative units nesting information. The latter, and also associations between administrative units and their boundaries can be derived if an efficient database topological model is defined. Again, although these elements were not implementable with the available source data, or were implemented within a small use case, they are as well voidable. Therefore, to satisfy the minimum requirements of the INSPIRE data model might not be a very complicated task but as already mentioned the value of the data will not be great either. As it stands now, it is probably fair to say that full technical conformance with the application schema requirements will come gradually, beginning with at least the minimum and growing to the complete 'interoperable data network' that INSPIRE aims to achieve, and also once data belonging to the other themes will be harmonised. It is then when the true value of INSPIRE will be achieved. This also refers to the fact that INSPIRE doesn't oblige Member States to collect new data, but when information that is required in an application schema is missing at the source, it becomes an ambiguous problem.

7.3 Conformance testing

To make the process of implementing INSPIRE requirements fully legitimate, this research also tackled the aspect of conformance testing as required by the data specifications. There are still many aspects to be decided and developed regarding conformance testing within the scope of INSPIRE data specifications, with only a general draft template of ATS available from the JRC (latest version being released internally about 10 days before the submission of this thesis – 16th July, 2012). TWGs will have to start working on ATS definition for all themes, Chapter 6 of this thesis being a relevant reference for the Administrative Units theme. Important to note again that a limitation of the conformance testing was considering only the aspects that are directly related to the data itself, excluding the ones covering aspects that belong to other INSPIRE specifications, like Network Services and Metadata. The main bottlenecks were the implementation of constraints testing and the quality of the CRS transformation. While for the testing of constraints that cannot be expressed by an XML schema, the JRC, and consequently each TWG, will be expected to provide Schematron documents in the future, the CRS transformation should be documented better in the specifications by maybe requiring a Root Mean Square Error (RMSE) calculation with each CRS transformation. It will then be when the accuracy flag that is associated with the CRS transformation will be

understood, and better informed decisions can be made on the usage of the data or the application the data is intended for.

Regarding the current version of the draft ATS template developed by the JRC, is worth mentioning that although most of the themes have specific geometry/topology requirements or recommendations, there is no conformance class defined in the template that would approach these requirements/recommendations. Most of these requirements can be considered data quality related issues, which is already a sensitive issue in INSPIRE, and the fact that they are explicitly defined in the data specification, should bring some clarifications in the ATS template. Other than that, the ATS template, and especially the latest version, seems a good starting point for the development of the more specialised ATS for the various themes.

7.4 Research questions

Based on the summaries of the previous sections, some very concise answers can provided for the research sub-questions, and consequently for the main research question.

- *What are the concepts of interoperability and data harmonisation and how do they fit in the context of INSPIRE?*

The need for interoperability is the main driving factor behind the INSPIRE framework, while data harmonisation is one of the main processes that bridges the heterogeneity between national Spatial Data Infrastructures (SDIs), as needed in INSPIRE. To achieve this goal of interoperability the data harmonisation tools and processes must be based on well-defined, internationally accepted standards.

- *How does the source data relate to the relevant INSPIRE data model and what are the preliminary bottlenecks?*

There are several inconsistencies between the source data and the relevant INSPIRE data model, which are mainly related to missing thematic attributes or geometrical/topological information in the source data model, but also requirements in the target data model that refer to additional/external data sources.

- *What are the available data harmonisation tools and how can these be used to implement INSPIRE data requirements?*

Among the list of data transformation tools, only two have been found to cope at the moment (without excessive intervention from the user – i.e. programming additional functionality) with INSPIRE data specifications: GO Publisher and FME. Data transformation is based on the three main identified steps in the literature review: schema matching, schema mapping (the most complicated to perform if there is a large gap between source data and target data model, and also a step that is almost impossible to automate), and schema trasnformation.

- *What does INSPIRE compliant/conformant mean and how can harmonised datasets be tested against INSPIRE requirements by means of an ATS/ETS?*

In INSPIRE conformance/compliance tend to mean sensibly different things, being related to legal obligations (compliance) and further technical requirements (conformance), an aspect that has a direct consequence for compliance/conformance testing. Testing is performed according to well established international standards (i.e. ISO and OGC), by defining a collection of test cases that would test individual requirements, grouped in conformance classes that refer to main aspects of the data (as also structured in data specifications), but also on good practices of maintaining data integrity/consistency. The method of testing can highly vary depending on the scope of the test case (i.e. schema validation, SQL level validation, etc.).

How can selected datasets be harmonised and declared INSPIRE compliant/conformant datasets according to the INSPIRE Data Specifications guidelines?

As it stands now, datasets should be possible to be harmonised to INSPIRE specifications gradually, at various levels, starting with the minimum requirements, which should be easier to implement and also more independent from other data sources, and in time, as data from the other INSPIRE themes is harmonised as well, the more complicated to derive attributes/associations should become easier to implement, and that is when the real value of INSPIRE will become apparent. On the other hand, although a generic methodology is implementable, conformance testing is still a sensible issue in the INSPIRE context and it is impossible at the moment to thoroughly test a dataset and declare it fully conformant, as many more aspects have to be clarified.

7.5 Recommendations

Based on the achieved results throughout the entire research project, several recommendations can be made that are not only very relevant for the Administrative Units theme, but also for the wider scope of INSPIRE data transformation and conformance testing. These could represent a starting point for further research, but also aspects to consider for future directions of INSPIRE implementation.

First of all, there is still 'room' for a more clear expression of certain aspects in the data specifications documents. This is especially true for the difference between legally binding requirements and not legally binding requirements. Furthermore, a more clear definition is needed from the very first chapter in the data specification documents about the purpose and meaning of the Abstract Test Suite (ATS), as well as Executable Test Suite (ETS), including concepts like conformance classes and test cases, giving that the very first requirement in all specifications is passing the ATS *"presented in Annex A".* Moreover, it is necessary to be more specific towards data quality aspects, and maybe even consider different requirements based on the intended use of the data and the applications it serves. This 'fuzziness' is often in contradiction with some very explicitly defined requirements for data geometry/topology. Same argument also applies towards coordinate reference system (CRS) transformation and the associated accuracy flag.

Regarding the actual data transformation process, for efficiency and correctness purposes, it is very important to have domain experts involved that have very good knowledge of the source data, and potentially, technical consultants that have very good knowledge of INSPIRE and everything that is associated with INSPIRE requirements (i.e. GML/XML, XSD schemas, UML, etc.).

Although the current general ATS template version is a good starting point for developing a given INSPIRE theme ATS, there is some confusion generated by adding test cases to a conformance class that are not applicable for all themes. It would be good idea to separate the more specific test cases from the test cases that are applicable for absolutely all themes. In addition, some of the proposed test cases are supposed to check the integrity of some generic elements in the application schema, which have been developed in the first place by the same people that developed the test cases. Also, it is really necessary to include in the ATS template a conformance class that specifically addresses geometrical/topological requirements, as most of the data specifications have explicit requirements in that sense.

As it has been demonstrated, the execution of the test cases was possible by using varied methods (XML/GML schema validation, SQL level validation, etc.). It would be recommended that a more integrated approach for the ETS is researched, giving an option that would incorporate as many test cases as possible into a single testing method/tool. In addition, it should be possible to explicitly declare conformance not only with the entire data specification, but also on a conformance class basis, because it would help with detecting non-conformant aspects of the data and also separating requirements of different INSPIRE components (i.e. Metadata, Network Services, Data Specifications).

Ultimately, as a general remark for conformance testing, in the future it might be a good direction for the JRC to collaborate closer with software providers like, Safe Software and Snowflake Software, which currently provide strong capabilities for data transformation to INSPIRE data models, and maybe create a compliance program, similar to the one offered by OGC, where various software capabilities are certified. It would be interesting to have such a programme where certain software packages can be declared INSPIRE compliant and certified in that sense. At the moment no software provider can claim that they offer INSPIRE compliant products, but they can say the software can be used to publish INSPIRE compliant data, which is a very smart marketing policy but often misunderstood by some. This ambiguity has to disappear in the future for the benefit of the entire community involved in INSPIRE.

To conclude, according to the INSPIRE timeline data transformation will become more and more of a very interesting topic in the coming years, with this being one of the last and most time consuming steps of the framework. Most of the policy and legal aspects have been implemented already in the Member States, while Metadata and Network Services have already passed the first deadlines. With the data specifications for Annex II and III almost finalised, thousands of additional data providers will soon be required to harmonise their data to these specifications, which will be the biggest challenge of INSPIRE. It is therefore very important to eliminate all the ambiguities in the requirements and have clear guidelines that cover both transformation requirements and conformance testing requirements, with a very clear distinction between legal obligations and technical requirements. In this sense, the citation below is very conclusive:

"The real success for INSPIRE will be when people don't know they are actually using it…"

Dean Hintz (Safe Software)

References

Bernard L., Kanellopoulos I., Annoni A., & Smits P., (2005), *The European geoportal—one step towards the establishment of a European Spatial Data Infrastructure*. Computers, Environment and Urban Systems, 29(1), pp. 15-31.

Bishr I., (1998), *Overcoming the Semantic and Other Barriers to GIS Interoperability*. Geographical Information Science, 12(4), pp. 299-314.

De Vries M., Giger C., Loidold M., (2007), *HUMBOLDT – State of the Art Data Harmonisation and Data Management.* Deliverable A3.5D1 of HUMBOLDT project.

De Vries M., Reitz T., Fichtinger A., Klien E., Roccatagliata E. & Zlatanova S., (2009), *HUMBOLDT Data Harmonisation Tools in Action*. Presentation given at INSPIRE Conference 2009 (15th-19th June), Rotterdam, The Netherlands.

De Vries M., Di Donato P., Ledoux H., Zlatanova S. & Laurent D., (2010), *Concept of Application-Specific Harmonised Data Models*, HUMBOLDT Project Deliverable A7.1-D2, 44 p. Available at http://www.esdi-humboldt.eu/files/1102-a7_1-d2_concept_of_application-specific_-harmonised-tud-002-final.pdf

Drafting Team 'Data Specifications', (2008a), *D2.3: Definition of Annex Themes and Scope – Version 3.0.* Available at http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/Data-Specifications/D2.3_Definition_of_Annex_Themes_and_scope_v3.0.pdf

Drafting Team 'Data Specifications', (2008b), *D2.6: Methodology for the Development of Data Specifications – Version 3.0*. Available at http://inspire.jrc.ec.europa.eu/reports/Imple-mentingRules/DataSpecifications/D2.6_v3.0.pdf

Drafting Team 'Data Specifications', (2010a), *D2.7: Guidelines for the Encoding of Spatial Data – Version 3.2*. Available at http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/-D2.7_v3.2.pdf

Drafting Team 'Data Specifications', (2010b), *D2.5: Generic Conceptual Model – Version 3.3*. Available at http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3_3.pdf

Drafting Teams 'Data Specifications', 'Network Services' ,'Metadata', (2007), *INSPIRE Technical Architecture – Overview*. Available at http://inspire.jrc.ec.europa.eu/reports/-ImplementingRules/network/INSPIRETechnicalArchitectureOverview_v1.2.pdf

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)*. Available at http://eur-lex.europa.eu/Notice.do?val=447703:-cs&lang=en&list=447703:cs,&pos=1&page=1&nbl=1&pgs=10&hwords=

European Commission, (2010), *Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services*. Available at http://eur-lex.europa.eu/Notice.do?val=545253:cs&lang=en&list=545253:cs,&pos=1&page=1&nbl=1&pgs=10&hwords=

European Commission, (2011), *Commission Regulation (EU) No 102/2011 of 4 February 2011 amending Regulation (EU) No 1089/2010 implementing Directive 2007/2/EC of the European*

Parliament and of the Council as regard interoperability of spatial data sets and services. Available at http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32011R0102-:EN:NOT

Fichtinger A., Rix J., Schäffler U., Michi I., Gone M. & Reitz T., (2011), *Data Harmonisation Put into Practice by the HUMBOLDT Project*. International Journal of Spatial Data Infrastructures Research, Vol. 6, pp. 234-260.

Friis-Christensen A., Schade S. & Peedell S., (2005), *Approaches to Solve Schema Heterogeneity at European Level*. Proceedings of the 11th EC-GI & GIS Workshop, 'ESDI: Setting the Framework' (29th June – 1st July), Alghero, Sardinia, Italy.

Goodchild M., Egenhofer M. & Fegeas R., (1997), *Interoperating GISs: Report of a Specialist Meeting held unde the Auspices of the VARENIUS Project.* Available online at: http://www.ncgia.ucsb.edu/Publications/Varenius_Reports/Interop.pdf

INSPIRE Thematic Working Group Coordinate Reference Systems and Geographical Grid Systems, (2010), *D2.8.I.1 INSPIRE Specification on Coordinate Reference Systems – Guidelines (v3.1).* Available at http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_-Specification_CRS_v3.1.pdf

INSPIRE Thematic Working Group Administrative Units, (2010), *D2.8.I.4 INSPIRE Data Specifications on Administrative Units – Version 3.0.1.* Available at http://inspire.jrc.ec.europa.eu-/documents/Data_Specifications/INSPIRE_DataSpecification_AU_v3.0.1.pdf

ISO 19105/FDIS:2000(E), (2000), *Geographic Information – Conformance and Testing*. Available at http://www.iso.org/iso/catalogue_detail.htm?csnumber=26010

ISO/DIS 19152, (2010), *Geographic Information – Land Administration Domain Model (LADM)*. Available at http://www.iso.org/iso/catalogue_detail.htm?csnumber=51206

Hemmatnia E., Van den Broecke J. & Van Raamsdonk K., (2010), *Harmonising Dutch National Geodata Conform To INSPIRE Using Combined Transformation*. International Journal of Spatial Data Infrastructures Research, Vol. 5, pp. 365-381.

Lasschuyt E. & van Hekken M., (2001), *Information Interoperability and Information Standardisation for NATO C2 – A Practical Approach*. Available at http://ftp.rta.nato.int/public//-PubFullText/RTO/MP/RTO-MP-064///MP-064-05.pdf

Manso M.-A., Wachowicz M., (2009), *GIS Design: A Review of Current Issues in Interoperability*. Geography Compass, 3(3), pp. 1105-1124.

Manso M.-A., Wachowicz M. & Bernabe M.-A., (2009), *Towards an Integrated Model of Interoperability for Spatial Data Infrastructures*. Transactions in GIS, 13(1), pp. 43-67.

Millot M. & Viganó S., (2007), *Data State of Play – Compliance Testing and Interoperability Checking*. JRC Scientifical and Technical Reports, EC-JRC-IES. Available at: http://inspire-.jrc.ec.europa.eu/reports/ImplementingRules/network/Data%20_State%20_of_Play_EUR_report.pdf

Nebert D. (ed.), (2004), *Developing Spatial Data Infrastructures: The SDI Cookbook (Version 2.0)*, GSDI Technical Working Group. Available at http://www.gsdi.org/docs2004/Cookbook/-cookbookV2.0.pdf

Nowak J., Nogueras-Iso J. & Peedell S., (2005), *Issues of Multiliguality in Creating a European SDI – The Perspective for Spatial Data Interoperability.* Paper presented at the 11th EC-GI & GIS Workshop 'ESDI: Setting the Framework' (29th June – 1st July), Alghero, Sardinia, Italy.

Portele C., Illert A., Laurent D., Borrebaek M., (2007), *Methodology & Guidelines on Use Case & Schema Development, Version 1.2.* Deliverable no. 15 [4.1.1] of the Reference Information Specifications for Europe (RISE) project. Available at http://www.euro-geographics.org/documents/RISE15MethodologyandGuidelinesV1.2.pdf

Portele C., Östman A, Skopeliti A., Kovanen J., Thomas J., Beare M., Koutroumpas M. & He X., (2011), *D12.6 – Test Report*. Available at http://www.esdin.eu/sites/esdin.eu/files/ESDIN_D12.6_-Test_report_v1.0.pdf

Open Geospatial Consortium Inc., (2007), *OpenGIS® Geography Markup Language (GML) Encoding Standard (Version 3.2.1)*. Available at http://www.opengeospatial.org/standards/gml

Ordnance Survey, (2012), *Boundary-Line™ - User Guide and Technical Specification (v1.15 – 05/2012)*. Available at http://www.ordnancesurvey.co.uk/oswebsite/docs/user-guides/boundary-line-user-guide.pdf

Östman A., Abugessaisa I., Tanzilli S., He X. & El-Mekawy M., (2009), *GeoTest: A Testing Environment for Swedish Geodata*. Paper presented at the GSDI 11 World Conference, Rotterdam, 2009. Available at http://www.gsdi.org/gsdiconf/gsdi11/papers/pdf/234.pdf

Östman A., (2010), *Data Harmonisation Concepts*. Presentation in the HUMBOLDT Training Framework, Level 1 – HUMBOLDT Knowledge. Available at http://www.gisig.it-/humboldt/training/index.php?option=com_content&view=article&id=58&Itemid=74

Rajabifard A., Williamson I. P., Holland P. & Johnstone G., (2000), *From Local to Global SDI Initiatives: A Pyramid Building Blocks*. Proceedings of the 4th GSDI Conference (13 – 15 March), Cape Town, South Africa.

Rajabifard A. & Williamson I. P., (2001), *Spatial Data Infrastructures: Concept, SDI Hierarchy and Future Directions*. Proceedings of GEOMATICS '80 Conference (29 April – 2 May), Teheran, Iran.

Shanzhen Y., Qi L. & Jicheng C., (1999), *An Interoperability GIS Model Based on the Spatial Information Infrastructure*. Proceedings of Geoinformatics '99 Conference (12 – 21 June), Ann Arbor, MI, USA.

Staub P., (2007), *A Model-Driven Web Feature Service for Enhanced Semantic Interoperability.* Proceedings of FOSS4G 2007: OSGeo Journal, 3, pp. 38-43.

Sudra P., (2010), *INSPIRE-compliant Web Services – The Case of Narew National Park, Poland*. MSc Thesis GIMA, Utrecht University – TU Delft – Wageningen UR – UT/ITC Enschede.

Tolk A., (2003), *Beyond Technical Interoperability – Introducing a Reference Model for Measures of Merit for Coalition Interoperability*. Proceedings of the 8th ICCRTS (17 – 19 June), Washington, D.C., USA.

Tóth K., Cetl V., Reznik T. & Tomas R., (2012), *Abstract Test Suite for INSPIRE Data Specifications*. Presentation given at INSPIRE Conference 2012 (23rd-27th June), Istanbul, Turkey. Available at http://inspire.jrc.ec.europa.eu/events/conferences/inspire_2012/presentations/222.pdf

Turnitsa C. & Tolk A., (2006), *Battle Management Language: A Triangle with Five sides.* Proceedings of the Simulation Interoperability Standards Organisation (SISO) Spring Simulation Interoperability Workshop (SIW) (2 – 7 April), Huntsville, AL, USA.

Van Loenen B. (2006), *Developing Geographic Information Infrastructures – The Role of Information Policies*. DUP Science, Delft.

Van Oosterom P., (2006), *Constraints in Spatial Data Models, in a Dynamic Context.* Chapter in Drummond J., Billen R., Joao E. and Forrest D. (Eds.), 'Dynamic and Mobile GIS: Investigating Changes in Space and Time', (pp. 104-137).

Villa P., Reitz T. & Gomarasca M. A., (2007), *HUMBOLDT Project: Implementing a Framework for Geo-Spatial Data Harmonisation and Moving Towards an ESDI.* GeoInformation in Europe, Proceedings of the 27th EARSeL Symposium, Millpress, pp. 29-36.

Vowles G., (2012), *OGC Standards and INSPIRE.* Presentation given at AGI Interoperability Day (22nd June), Exeter, UK. Available at http://www.agi.org.uk/storage/events/120622-InteropDay/GrahamVowles.pdf

Witschas S., (2010), *Cross-Border Mapping: Geodata, Geonames, Multilinguality and More Cartography in Central and Eastern Europe*. In G. Gartner & F. Ortag (Eds.), Cartography in Central and Eastern Europe (pp. 163-180). Vienna: Springer Berlin Heidelberg.

Websites

ArcGIS Resources – http://resources.arcgis.com

CEN/TC 287 Geographic Information – http://www.centc287.eu

Data Harmonisation Panel – http://www.dhpanel.eu/

European Committee for Standardisation – http://www.cen.eu

Infrastructure for Spatial Information in Europe (INSPIRE) – http://inspire.jrc.ec.europa.eu/

INSPIRE FOSS – http://code.google.com/p/inspire-foss/

International Organisation for Standardisation – http://www.iso.org

ISO/TC 211 Geographic information/Geomatics – http://www.isotc211.org

HUMBOLDT Project – http://www.esdi-humboldt.eu/

Office for National Statistics – http://www.ons.gov.uk

Open Geospatial Consortium (OGC) – http://www.opengeospatial.org/

Oracle – http://www.oracle.com

Ordnance Survey – http://www.ordnancesurvey.co.uk/

Safe Software – http://www.safe.com

Snowflake Software – http://www.snowflakesoftware.com

The Free Dictionary – http://www.thefreedictionary.com/

Annex A – INSPIRE Feature Catalogue (Administrative Units)

NOTE 1: The INSPIRE Feature Catalogue in this Annex is based on version 3.0.1 of the data specifications for the Administrative Units theme (03/05/2010).

NOTE 2: This Annex is mainly related to Chapter 3 (Case Study: INSPIRE Administrative Units – Gap Analysis), and especially Chapter 5 (Producing INSPIRE Compliant Data).

Spatial Object Types

AdministrativeUnit

Definition:

Unit of administration where a Member State has and/or exercises jurisdictional rights, for local, regional and national governance.

Type:

Spatial Object Type

Attribute:

Name: geometry

Definition: Geometric representation of spatial area covered by this administrative unit.

Voidable: false

Multiplicity: 1

Value type: GM_MultiSurface

Attribute:

Name: nationalCode

Definition: Thematic identifier corresponding to the national administrative codes defined in each country.

Voidable: false

Multiplicity: 1

Value type: CharacterString

Attribute:

Name: inspireId

Definition: External object identifier of the spatial object.

Description: NOTE An external object identifier is a unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object. The identifier is an identifier of the spatial object, not an identifier of the real-world phenomenon.

Voidable: false

Multiplicity: 1

Value type: Identifier (data type)

Attribute:

Name: nationalLevel

Definition: Level in the national administrative hierarchy, at which the administrative unit is established.

Voidable: false

Multiplicity: 1

Value type: AdministrativeHierarchyLevel (code list)

Values:

1stOrder	Highest level in the national administrative hierarchy (country level).
2ndOrder	2nd level in the national administrative hierarchy.
3rdOrder	3rd level in the national administrative hierarchy.

4thOrder	4th level in the national administrative hierarchy.	
5thOrder	5th level in the national administrative hierarchy.	
6thOrder	6th level in the national administrative hierarchy.	

Attribute:

Name: nationalLevelName

Definition: Name of the level in the national administrative hierarchy, at which the administrative unit is established.

Voidable: true

Multiplicity: 1..*

Value type: LocalisedCharacterString

Attribute:

Name: country

Definition: Two-character country code according to the Interinstitutional style guide published by the Publications Office of the European Union.

Voidable: false

Multiplicity: 1

Value type: CountryCode (code list)

Values:

BE	Belgium
BG	Bulgaria
CZ	Czech Republic
DK	Denmark
DE	Germany
EE	Estonia
IE	Ireland
EL	Greece
ES	Spain
FR	France
IT	Italy
CY	Cyprus
LV	Latvia
LT	Lithuania
LU	Luxembourg
HU	Hungary
MT	Malta
NL	Netherlands
AT	Austria
PL	Poland
PT	Portugal
RO	Romania
SI	Slovenia
SK	Slovakia
FI	Finland
SE	Sweden
UK	United Kingdom
HR	Croatia
TR	Turkey

Attribute:

Name: name

Definition: Official national geographical name of the administrative unit, given in several languages where required.

Voidable: false

Multiplicity:	1..*
Value type:	GeographicalName (data type)
Attribute:	
Name:	residenceOfAuthority
Definition:	Center for national or local administration.
Voidable:	true
Multiplicity:	1..*
Value type:	ResidenceOfAuthority (data type)
Attribute:	
Name:	beginLifespanVersion
Definition:	Date and time at which this version of the spatial object was inserted or changed in the spatial data set.
Voidable:	true
Multiplicity:	1
Value type:	DateTime
Attribute:	
Name:	endLifespanVersion
Definition:	Date and time at which this version of the spatial object was superseded or retired in the spatial data set.
Voidable:	true
Multiplicity:	0..1
Value type:	DateTime
Association role:	
Name:	condominium
Definition:	Condominium administered by this administrative unit
Description:	NOTE Condominiums may only exist at state level and can be administered only by administrative units at the highest level of the national administrative hierarchy (i.e. countries).
Voidable:	true
Multiplicity:	0..*
Value type:	Condominium (spatial object type)
Association role:	
Name:	boundary
Definition:	The administrative boundaries between this administrative unit and all the units adjacent to it.
Description:	NOTE Administrative boundary corresponds to the curve established between the nodes at lowest level of territory division in Member State. Thus, it does not necessairly represents boundary in political terms, but just part of it.
Voidable:	true
Multiplicity:	1..*
Value type:	AdministrativeBoundary (spatial object type)
Association role:	
Name:	NUTS
Definition:	NUTS region that topologically contains this administrative unit.
Description:	NOTE 1 NUTS regions are Territorial units for statistics defined in the framework of the Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 (see hhtp:ec.europe.eu/eurostat/ramon/nuts/home_regions_de.html). NOTE 2 Each administrative unit at lowest level is topologically covered by a certain NUTS3 region established for statistical purposes. Each NUTS3 region belongs to a specific NUTS2 region that is a part of NUTS1 region. The administrative unit at lowest level can refer the corresponding regions from all three levels: NUTS3, NUTS2, and NUTS1.
Voidable:	true
Multiplicity:	1..3
Value type:	NUTSRegion (spatial object type)
Association role:	

Name:	lowerLevelUnit
Definition:	Units established at a lower level of the national administrative hierarchy which are administered by this administrative unit.
Description:	NOTE For administrative units at the lowest level of the national hierarchy no lower level unit exists. CONSTRAINT Each administrative unit except for the lowest level units shall refer to its lower level units
Voidable:	true
Multiplicity:	0..*
Value type:	AdministrativeUnit (spatial object type)
Association role:	
Name:	upperLevelUnit
Definition:	A unit established at a higher level of national administrative hierarchy that this administrative unit administers.
Description:	NOTE Administrative units at the highest level of national hierarchy (i.e. the country) do not have upper level units. CONSTRAINT Each administrative unit at the level other than '1st order' (i.e. nationalLevel <> '1st order') shall refer their upper level unit.
Voidable:	true
Multiplicity:	0..1
Value type:	AdministrativeUnit (spatial object type)
Association role:	
Name:	administeredBy
Definition:	Administrative units established at same level of national administrative hierarchy which are co-administered by this administrative unit.
Voidable:	true
Multiplicity:	0..*
Value type:	AdministrativeUnit (spatial object type)
Association role:	
Name:	coAdminister
Definition:	A unit established at same level of national administrative hierarchy that administers this administrative unit.
Voidable:	true
Multiplicity:	0..*
Value type:	AdministrativeUnit (spatial object type)

Constraints:

AmininstrativeUnitHighestLevel: /* No unit at highest level can associate units at a higher level */ inv: self.nationalLevel = '1stOrder' implies self.upperLevelUnit->isEmpty() and self.loweLevelUnit->notEmpty()

AdministrativeUnitLowestLevel: /* No unit at lowest level can associate units at lower level and each lower level unit shall referr/ inv: self.nationalLevel = '6thOrder' implies self.lowerLevelUnit->isEmpty and self.upperLevelUnit->notEmpty

CondominiumsAtCountryLevel: /*Association role condominium applies only for administrative units which nationalLevel='1st order' (country level). */ inv: self.condominium->notEmpty implies self.nationalLevel = '1stOrder'

AdministrativeBoundary

Definition:

A line of demarcation between administrative units.

Type:

Spatial Object Type

Attribute:

Name: geometry

Definition: Geometric representation of border line.

Voidable: false

Multiplicity: 1

Value type: GM_Curve

Attribute:

Name: inspireId

Definition: External object identifier of the spatial object.

Description: NOTE An external object identifier is a unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object. The identifier is an identifier of the spatial object, not an identifier of the real-world phenomenon.

Voidable: false

Multiplicity: 1

Value type: Identifier (data type)

Attribute:

Name: country

Definition: Two-character country code according to the Inter-institutional style guide published by the Publications Office of the European Union.

Voidable: false

Multiplicity: 1

Value type: CountryCode (code list)

Values:

BE	Belgium
BG	Bulgaria
CZ	Czech Republic
DK	Denmark
DE	Germany
EE	Estonia
IE	Ireland
EL	Greece
ES	Spain
FR	France
IT	Italy
CY	Cyprus
LV	Latvia
LT	Lithuania
LU	Luxembourg
HU	Hungary
MT	Malta
NL	Netherlands
AT	Austria
PL	Poland
PT	Portugal
RO	Romania
SI	Slovenia
SK	Slovakia
FI	Finland
SE	Sweden
UK	United Kingdom
HR	Croatia
TR	Turkey

Attribute:

Name: nationalLevel

Definition: The hierarchy levels of all adjacent administrative units this boundary is part of.

Voidable: false

Multiplicity: 1..6

Value type: AdministrativeHierarchyLevel (code list)

Values:

1stOrder	Highest level in the national administrative hierarchy (country level).
2ndOrder	2nd level in the national administrative hierarchy.
3rdOrder	3rd level in the national administrative hierarchy.
4thOrder	4th level in the national administrative hierarchy.
5thOrder	5th level in the national administrative hierarchy.
6thOrder	6th level in the national administrative hierarchy.

Attribute:

Name: legalStatus

Definition: Legal status of this administrative boundary.

Description: NOTE The legal status is considered in terms of political agreement or disagreement of the administrative units separated by this boundary.

This attribute is initialised with the value "agreed".

Voidable: true

Multiplicity: 1

Value type: LegalStatusValue (enumeration)

Values:

agreed	The edge-matched boundary has been agreed between neighbouring administrative units and is stable now.
notAgreed	The edge-matched boundary has not yet been agreed between neighbouring administrative units and could be changed.

Attribute:

Name: technicalStatus

Definition: The technical status of the administrative boundary.

Description: NOTE The technical status of the boundary is considered in terms of its topological matching or not-matching with the borders of all separated administrative units. Edge-matched means that the same set of coordinates is used.

This attribute is initialised with the value "edge-matched".

Voidable: true

Multiplicity: 1

Value type: TechnicalStatusValue (enumeration)

Values:

edgeMatched	The boundaries of neighbouring administrative units have the same set of coordinates.
notEdgeMatched	The boundaries of neighbouring administrative units do not have the same set of coordinates.

Attribute:

Name: beginLifespanVersion

Definition: Date and time at which this version of the spatial object was inserted or changed in the spatial data set.

Voidable: true

Multiplicity: 1

Value type: DateTime

Attribute:

Name: endLifespanVersion

Definition: Date and time at which this version of the spatial object was superseded or retired in the spatial data set.

Voidable: true

Multiplicity: 0..1

Value type: DateTime

Association role:

Name: admUnit

Definition:	The administrative units separated by this administrative boundary.
Description:	NOTE In case of a national boundary (i.e. nationalLevel='1st order') only one national administrative unit (i.e. country) is provided.
Voidable:	true
Multiplicity:	1..*
Value type:	AdministrativeUnit (spatial object type)

Condominium

Definition:

An administrative area established independently to any national administrative division of territory and administered by two or more countries.

Description:

NOTE Condominium is not a part of any national administrative hierarchy of territory division in Member State.

Type:

Spatial Object Type

Attribute:

Name:	inspireId
Definition:	External object identifier of the spatial object.
Description:	NOTE An external object identifier is a unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object. The identifier is an identifier of the spatial object, not an identifier of the real-world phenomenon.
Voidable:	false
Multiplicity:	1
Value type:	Identifier (data type)

Attribute:

Name:	name
Definition:	Official geographical name of this condominium, given in several languages where required.
Voidable:	True
Multiplicity:	0..*
Value type:	GeographicalName (data type)

Attribute:

Name:	geometry
Definition:	Geometric representation of spatial area covered by this condominium
Voidable:	false
Multiplicity:	1
Value type:	GM_MultiSurface

Attribute:

Name:	beginLifespanVersion
Definition:	Date and time at which this version of the spatial object was inserted or changed in the spatial data set.
Voidable:	true
Multiplicity:	1
Value type:	DateTime

Attribute:

Name:	endLifespanVersion
Definition:	Date and time at which this version of the spatial object was superseded or retired in the spatial data set.
Voidable:	true

Multiplicity:	0..1
Value type:	DateTime
Association role:	
Name:	admUnit
Definition:	The administrative unit administering the condominium
Voidable:	true
Multiplicity:	1..*
Value type:	AdministrativeUnit (spatial object type)

NUTSRegion	
Definition:	
	Territorial unit for statistics defined in the framework of the Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003.
Description:	
	NOTE NUTS regions subdivide each Member State into a whole number of territorial units for statistic at NUTS1 level. Each of these is then subdivided into regions at NUTS2 level and these in turn into regions at NUTS3 level.
Type:	
	Spatial Object Type
Attribute:	
Name:	geometry
Definition:	Geometric representation of spatial area covered by this NUTS-region.
Voidable:	false
Multiplicity:	1
Value type:	GM_MultiSurface
Attribute:	
Name:	inspireId
Definition:	External object identifier of the spatial object.
Description:	NOTE An external object identifier is a unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object. The identifier is an identifier of the spatial object, not an identifier of the real-world phenomenon.
Voidable:	false
Multiplicity:	1
Value type:	Identifier (data type)
Attribute:	
Name:	NUTSCode
Definition:	Unique code of the territorial unit for statistics as defined in the framework of the Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003.
Description:	EXAMPLE A NUTScode from Denmark could be DK031.
Voidable:	false
Multiplicity:	1
Value type:	CharacterString
Attribute:	
Name:	beginLifespanVersion
Definition:	Date and time at which this version of the spatial object was inserted or changed in the spatial data set.
Voidable:	true
Multiplicity:	1
Value type:	DateTime
Attribute:	

Name:	endLifespanVersion
Definition:	Date and time at which this version of the spatial object was superseded or retired in the spatial data set.
Voidable:	true
Multiplicity:	0..1
Value type:	DateTime

Data Types

Identifier	
Definition:	
External unique object identifier published by the responsible body, which may be used by external applications to reference the spatial object.	
Description:	
NOTE1 External object identifiers are distinct from thematic object identifiers. NOTE 2 The voidable version identifier attribute is not part of the unique identifier of a spatial object and may be used to distinguish two versions of the same spatial object. NOTE 3 The unique identifier will not change during the life-time of a spatial object.	
Type:	
Data Type	
Attribute:	
Name:	localId
Definition:	A local identifier, assigned by the data provider. The local identifier is unique within the namespace, which is no other spatial object carries the same unique identifier.
Description:	NOTE It is the responsibility of the data provider to guarantee uniqueness of the local identifier within the namespace.
Voidable:	false
Multiplicity:	1
Value type:	CharacterString
Attribute:	
Name:	namespace
Definition:	Namespace uniquely identifying the data source of the spatial object.
Description:	NOTE The namespace value will be owned by the data provider of the spatial object and will be registered in the INSPIRE External Object Identifier Namespaces Register.
Voidable:	false
Multiplicity:	1
Value type:	CharacterString
Attribute:	
Name:	versionId
Definition:	The identifier of the particular version of the spatial object, with a maximum length of 25 characters. If the specification of a spatial object type with an external object identifier includes life-cycle information, the version identifier is used to distinguish between the different versions of a spatial object. Within the set of all versions of a spatial object, the version identifier is unique.
Description:	NOTE 1 The maximum length has been selected to allow for time stamps based on ISO 8601, for example, "2007-02-12T12:12:12+05:30" as the version identifier. NOTE 2 The property is void, if the spatial data set does not distinguish between different versions of the spatial object. It is missing, if the spatial object type does not support any life-cycle information.
Voidable:	true
Multiplicity:	0..1
Value type:	CharacterString
Constraints:	

Allowed characters for localId and namespace: /*The localId and the namespace shall only use the following set of characters: {”A” …”Z”, “a”…”z,””0”…”9”, “_”, “.”, “-“}, that is only letters from the Latin alphabet, digits, underscore, point, and dash are allowed.*/ inv: let allowedChar : Set {'A'..'Z', 'a'..'z', '0'..'9', '_', '.', '-'} in (namespace.element->forAll(char | allowedChar->exists(char) and localId.element->forAll(char | allowedChar->exists(char)))

ResidenceOfAuthority

Definition:

Data type representing the name and position of a residence of authority.

Type:

Data Type

Attribute:

Name: name
Definition: Name of the residence of authority.
Voidable: false
Multiplicity: 1
Value type: GeographicalName (data type)

Attribute:

Name: geometry
Definition: Position of the residence of authority.
Voidable: true
Multiplicity: 1
Value type: GM_Point

GeographicalName

Definition:

Proper noun applied to a real world entity.

Type:

Data Type

Attribute:

Name: language
Definition: Language of the name, given as a three letters code, in accordance with either ISO 639-3 or ISO 639-5.
Description: NOTE 1More precisely, this definition refers to the language used by the community that uses the name.

NOTE 2 The code "mul" for "multilingual" should not be used in general. However it can be used in rare cases like official names composed of two names in different languages. For example, "Vitoria-Gasteiz" is such a multilingual official name in Spain.

NOTE 3 Even if this attribute is "voidable" for pragmatic reasons; it is of first importance in several use cases in the multi-language context of Europe.

Voidable: true
Multiplicity: 1
Value type: CharacterString

Attribute:

Name: nativeness
Definition: Information enabling to acknowledge if the name is the one that is/was used in the area where the spatial object is situated at the instant when the name is/was in use.
Voidable: true
Multiplicity: 1

Value type: NativenessValue (code list)

Values:

endonym	Name for a geographical feature in an official or well-established language occurring in that area where the feature is situated. SOURCE [UNGEGN Glossary 2007].
exonym	Name used in a specific language for a geographical feature situated outside the area where that language is widely spoken, and differing in form from the respective endonym(s) in the area where the geographical feature is situated. SOURCE [UNGEGN Glossary 2007].

Attribute:

Name: nameStatus

Definition: Qualitative information enabling to discern which credit should be given to the name with respect to its standardisation and/or its topicality.

Description: NOTE The Geographical Names application schema does not explicitly make a preference between different names (e.g. official endonyms) of a specific real world entity. The necessary information for making the preference (e.g. the linguistic status of the administrative or geographic area in question), for a certain use case, must be obtained from other data or information sources. For example, the status of the language of the name may be known through queries on the geometries of named places against the geometry of administrative units recorded in a certain source with the language statuses information.

Voidable: true

Multiplicity: 1

Value type: NameStatusValue (code list)

Values:

official	Name in current use and officially approved or established by legislation.
standardised	Name in current use and accepted or recommended by a body assigned advisory function and/or power of decision in matters of toponymy.
historical	Historical name not in current use.
other	Current, but not official, nor approved name.

Attribute:

Name: sourceOfName

Definition: Original data source from which the geographical name is taken from and integrated in the data set providing/publishing it. For some named spatial objects it might refer again to the publishing data set if no other information is available.

Description: EXAMPLES Gazetteer, geographical names data set.

Voidable: true

Multiplicity: 1

Value type: CharacterString

Attribute:

Name: pronunciation

Definition: Proper, correct or standard (standard within the linguistic community concerned) pronunciation of the geographical name.

Description: SOURCE Adapted from [UNGEGN Manual 2006].

Voidable: true

Multiplicity: 1

Value type: PronunciationOfName (data type)

Attribute:

Name: spelling

Definition: A proper way of writing the geographical name.

Description: NOTE 1 Different spelling should only be used for names rendered in different scripts. .

NOTE 2 While a particular GeographicalName should only have one spelling in a given script, providing different spellings in the same script should be done through the provision of different geographical names associated with the same named place.

Voidable: false

Multiplicity: 1..*

Value type: SpellingOfName (data type)

Attribute:

Name: grammaticalGender

Definition: Class of nouns reflected in the behaviour of associated words.

Description: NOTE the attribute has cardinality [0..1] and is voidable, which means that:

- in case the concept of grammatical gender has no sense for a given name (i.e. the attribute is not applicable), the attribute should not be provided.
- in case the concept of grammatical gender has some sense for the name but is unknown, the attribute should be provided but *void*.

Voidable: true

Multiplicity: 0..1

Value type: GrammaticalGenderValue (code list)

Values:

masculine	Masculine grammatical gender. EXAMPLES Sena (Spanish), Schwarzwald (German).
feminine	Feminine grammatical gender. EXAMPLES Seine (French), Forêt Noire (French).
neuter	Neuter grammatical gender. EXAMPLES Zwarte Woud (Dutch), Rheinland (German).
common	'Common' grammatical gender (the merging of 'masculine' and 'feminine').

Attribute:

Name: grammaticalNumber

Definition: Grammatical category of nouns that expresses count distinctions.

Description: NOTE the attribute has cardinality [0..1] and is voidable, which means that:

- in case the concept of grammatical number has no sense for a given name (i.e. the attribute is not applicable), the attribute should not be provided.
- in case the concept of grammatical number has some sense for the name but is unknown, the attribute should be provided but *void*.

Voidable: true

Multiplicity: 0..1

Value type: GrammaticalNumberValue (code list)

Values:

singular	Singular grammatical number. EXAMPLES Danube (English), Lac du Bourget (French), Praha (Czech), Nederland (Dutch).
plural	Plural grammatical number. EXAMPLES Alps (English), Pays-Bas (French), Waddeneilanden (Dutch), Cárpatos (Spanish).
dual	Dual grammatical number.

PronunciationOfName

Definition:

Proper, correct or standard (standard within the linguistic community concerned) pronunciation of a name.

Description:

SOURCE Adapted from [UNGEGN Manual 2006].

Type:

Data Type

Attribute:

Name: pronunciationSoundLink

Definition: Proper, correct or standard (standard within the linguistic community concerned) pronunciation of a name, expressed by a link to any sound file.

Description: SOURCE Adapted from [UNGEGN Manual 2006].

Voidable: true

Multiplicity: 0..1

Value type: URI

Attribute:

Name: pronunciationIPA

Definition: Proper, correct or standard (standard within the linguistic community concerned) pronunciation

	of a name, expressed in International Phonetic Alphabet (IPA).
Description:	SOURCE Adapted from [UNGEGN Manual 2006].
Voidable:	true
Multiplicity:	0..1
Value type:	CharacterString

SpellingOfName	
Definition:	
Proper way of writing a name.	
Description:	
SOURCE Adapted from [UNGEGN Manual 2006].	
NOTE Proper spelling means the writing of a name with the correct capitalisation and the correct letters and diacritics present in an accepted standard order.	
Type:	
Data Type	
Attribute:	
Name:	text
Definition:	Way the name is written.
Voidable:	false
Multiplicity:	1
Value type:	CharacterString
Attribute:	
Name:	script
Definition:	Set of graphic symbols (for example an alphabet) employed in writing the name, expressed using the four letters codes defined in ISO 15924, where applicable.
Description:	SOURCE Adapted from [UNGEGN Glossary 2007].
	EXAMPLES Cyrillic, Greek, Roman/Latin scripts.
	NOTE 1The four letter codes for Latin (Roman), Cyrillic and Greek script are "Latn", "Cyrl" and "Grek", respectively.
	NOTE 2 In rare cases other codes could be used (for other scripts than Latin, Greek and Cyrillic). However, this should mainly apply for historical names in historical scripts.
	NOTE 3 This attribute is of first importance in the multi-scriptual context of Europe.
Voidable:	true
Multiplicity:	1
Value type:	CharacterString
Attribute:	
Name:	transliterationScheme
Definition:	Method used for the names conversion between different scripts.
Description:	SOURCE Adapted from [UNGEGN Glossary 2007].
	NOTE 1 This attribute should be filled for any transliterated spellings. If the transliteration scheme used is recorded in codelists maintained by ISO or UN, those codes should be preferred.
Voidable:	true
Multiplicity:	0..1
Value type:	CharacterString

SpatialDataSet
Definition:

Identifiable collection of spatial data.
Description: NOTE The type SpatialDataSet is offered as a pre-defined type for spatial data sets. INSPIRE application schemas may specify their own spatial data set types. It specifies three properties: an external object identifier, a container for metadata (may be void), and an association to zero or more spatial objects.
Type: Spatial Object Type
Attribute: **Name:** identifier **Definition:** Identifier of the spatial data set. **Voidable:** false **Multiplicity:** 1 **Value type:** Identifier (data type)
Attribute: **Name:** metadata **Definition:** Metadata of the spatial data set. **Voidable:** true **Multiplicity:** 1 **Value type:** MD_Metadata
Association role: **Name:** member **Definition:** The spatial objects that are members of the spatial data set **Voidable:** false **Multiplicity:** 0..* **Value type:** AbstractFeature (spatial object type)

Annex B – Administrative Units requirements / recommendations

NOTE 1: The requirements and recommendations in this Annex are based on version 3.0.1 of the data specifications for the Administrative Units theme (03/05/2010).

NOTE 2: This Annex is mainly related to Chapter 3 (Case Study: INSPIRE Administrative Units – Gap Analysis) and Chapter 5 (Producing INSPIRE Compliant Data), but especially Chapter 6 (Conformance Testing).

NOTE 3: Requirements and recommendations have been grouped according to the different chapters they are written in, in the Administrative Units data specifications.

CONFORMANCE

Requirement 1 – Any dataset claiming conformance with this INSPIRE data specification shall pass the requirements described in the abstract test suite presented in Annex A.

DATA CONTENT AND STRUCTURE

Requirement 2 – Spatial data sets related to the theme Administrative units shall be provided using the spatial object types and data types specified in the application schema in this section.

Requirement 3 – Each spatial object shall comply with all constraints specified for its spatial object type or data types used in values of its properties, respectively.

Recommendation 1 – The reason for a void value should be provided where possible using a listed value from the VoidValueReason code list to indicate the reason for the missing value.

Requirement 4 – Each instance of spatial object type AdministrativeUnit, except for the country level unit representing a Member State and co-administered units, shall refer exactly one unit at a higher level of administrative hierarchy. This correspondence shall be expressed by the "upperLevelUnit" association role of AdministrativeUnit spatial object type.

Requirement 5 – Each instance of spatial object type AdministrativeUnit, except for those at the lowest level, shall refer to their respective lower level units. This correspondence shall be expressed by the "lowerLevelUnit" association role of AdministrativeUnit spatial object type.

Requirement 6 – If an administrative unit is co-administered by two or more other administrative units the association role "administeredBy" must be used. The units co-administering this unit shall apply inverse role "coAdminister".

Recommendation 2 – The value of 'language' attribute for AdministrativeUnit.name (GeographicalName DataType) should be provided, except for the situation that the data producer does not have such information.

Requirement 7 – Administrative units at the same level of administrative hierarchy shall not conceptually share common areas.

Requirement 8 – Instances of the spatial object type AdministrativeBoundary shall correspond to the edges in the topological structure of the complete (including all levels) boundary graph.

Recommendation 3 – The following geometric and topological constraints are recommendations for this data specification:

a. Adjacent administrative units should not overlap, i.e. their boundaries should not intersect with each other.
b. There should be no gaps between adjacent administrative units.
c. Unintended gaps between administrative units due to geometrical inconsistencies are in principle not allowed. Boundaries of neighbouring administrative units shall have the same set of coordinates, within the specified resolution.
d. The border line that limits the administrative units shall correspond to the geometries representing the boundaries of this administrative unit.
e. The boundaries must not have dangles; boundaries always divide different administrative units.

Requirement 9 – The spatial extent of a condominium may not be part of the geometry representing the spatial extent of an administrative unit.

Requirement 10 – Condominiums can only be administered by administrative units at country level.

Recommendation 4 – Member States should link the Administrative Units to the NUTSRegions for each administrative unit to ensure interoperability with national and European statistical/thematic information.

Requirement 11 – The value domain of spatial properties defined in this Regulation shall be restricted to the Simple Feature spatial schema as defined by EN ISO 19125-1.

Recommendation 5 – All spatial objects should be provided at the source accuracy where possible.

Recommendation 6 – All spatial objects should have a positional accuracy of 50 meters or better.

Recommendation 7 – If life-cycle information is not maintained as part of the spatial dataset, all spatial objects belonging to this data set should provide a void value with a reason of "unknown".

REFERENCE SYSTEMS

Requirement 12 – For the coordinate reference systems used for making available the INSPIRE spatial data sets, the datum shall be the datum of the European Terrestrial Reference System 1989 (ETRS89) in areas within its geographical scope, and the datum of the International Terrestrial Reference System (ITRS) or other geodetic coordinate reference systems compliant with ITRS in areas that are outside the geographical scope of ETRS89. Compliant with the ITRS means that the system definition is based on the definition of the ITRS and there is a well-established and described relationship between both systems, according to EN ISO 19111.

Requirement 13 – INSPIRE spatial data sets shall be made available using one of the three-dimensional, two-dimensional or compound coordinate reference systems specified in the list below. Other coordinate reference systems than those listed below may only be used for regions outside of continental Europe. The geodetic codes and parameters for these coordinate reference systems shall be documented, and an identifier shall be created, according to EN ISO 19111 and ISO 19127.

Requirement 14 – For the display of the INSPIRE spatial data sets with the View Service specified in D003152/02 Draft Commission Regulation implementing Directive 2007/2/EC of the European Parliament and of the Council as regards Network Services, at least the two dimensional geodetic coordinate system shall be made available.

Requirement 15 – For referring to the non-compound coordinate reference systems listed in this Section, the identifiers listed below shall be used. For referring to a compound coordinate reference system, an identifier composed of the identifier of the horizontal component, followed by a slash (/), followed by the identifier of the vertical component, shall be used.

Requirement 16 – The Gregorian Calendar shall be used for as a reference system for date values, and the Universal Time Coordinated (UTC) or the local time including the time zone as an offset from UTC shall be used as a reference system for time values.

DATA QUALITY

Recommendation 8 – Aggregated data quality information should ideally be collected at the level of spatial object types and included in the dataset (series) metadata.

DATASET-LEVEL METADATA

Requirement 17 – The metadata describing a spatial data set or a spatial data set series related to the theme Administrative units shall comprise the metadata elements required by Regulation 1205/2008/EC (implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata) for spatial datasets and spatial dataset series (Table 6) as well as the metadata elements specified in Table 7.

Recommendation 9 – The metadata describing a spatial data set or a spatial data set series related to the theme Administrative units should comprise the theme-specific metadata elements specified in Table 8.

Recommendation 10 – In order to report conceptual consistency with this INSPIRE data specification, the Conformity metadata element should be used. The value of Conformant should be used for the Degree element only if the dataset passes all the requirements described in the abstract test suite presented in Annex A. The Specification element should be given as follows:

- title: “INSPIRE Data Specification on Administrative units –Guidelines”
- date:
    - dateType: publication
    - date: 2010-04-26

Recommendation 11 – Apart from describing the process history, if feasible within a free text, the overall quality of the dataset (series) should be included in the Lineage metadata element. This statement should contain any quality information required for interoperability and/or valuable for use and evaluation of the data set (series).

DELIVERY

Requirement 18 – Data conformant to this INSPIRE data specification shall be made available through an INSPIRE network service.

Requirement 19 – All information that is required by a calling application to be able to retrieve the data through the used network service shall be made available in accordance with the requirements defined in the Implementing Rules on Network Services.

Requirement 20 – Data conformant to the application schema Administrative units shall be encoded using the encoding specified in section 9.2.1.1.

PORTRAYAL

Requirement 21 – If an INSPIRE view service supports the portrayal of data related to the theme Administrative units, it shall provide the layers specified in this section.

Requirement 22 – If an INSPIRE view network service supports the portrayal of spatial data sets corresponding to the spatial data theme Administrative units, it shall support the default styles specified in the tables in this section. If no user-defined style is specified in a portrayal request for a specific layer to an INSPIRE view service, the default style specified in this section for that layer shall be used.

Annex C – INSPIRE GML Samples

NOTE 1: The GML samples contain one feature each from the 'AdministrativeUnit' and 'AdministrativeBoundary' spatial objects from the Administrative Units application schema.

NOTE 2: The geometry elements do not contain the full sets of coordinates as this would've occupied tens of pages, especially for an administrative unit. Dots are place between the very first and very last coordinates of the geometry object.

AdministrativeUnit


```
<?xml version="1.0" encoding="UTF-8"?><!--Created by GO Publisher Desktop
2.1.2 Build 28125 from 2012-03-22 11:26--><!--Snowflake Software Ltd.
(http://www.snowflakesoftware.com)-->
<base:SpatialDataSet xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:au="urn:x-
inspire:specification:gmlas:AdministrativeUnits:3.0" xmlns:su-
nuts="http://inspire.jrc.ec.europa.eu/schemas/su-nuts/2.0" xmlns:su-
vector="http://inspire.jrc.ec.europa.eu/schemas/su-vector/2.0" xmlns:su-
core="http://inspire.jrc.ec.europa.eu/schemas/su-core/2.0"
xmlns:hh="http://inspire.jrc.ec.europa.eu/schemas/hh/2.0"
xmlns:ef="http://inspire.jrc.ec.europa.eu/schemas/ef/2.0"
xmlns:base2="http://inspire.jrc.ec.europa.eu/schemas/base2/0.1"
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gts="http://www.isotc211.org/2005/gts"
xmlns:gsr="http://www.isotc211.org/2005/gsr"
xmlns:om="http://www.opengis.net/om/2.0" xmlns:base="urn:x-
inspire:specification:gmlas:BaseTypes:3.2"
xmlns:sd="http://inspire.jrc.ec.europa.eu/schemas/sd/2.0"
xmlns:gmlcov="http://www.opengis.net/gmlcov/1.0"
xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:gn="urn:x-
inspire:specification:gmlas:GeographicalNames:3.0"
xsi:schemaLocation="urn:x-
inspire:specification:gmlas:AdministrativeUnits:3.0 AdministrativeUnits.xsd
urn:x-inspire:specification:gmlas:BaseTypes:3.2 BaseTypes.xsd"
gml:id="UK.INSPIRE.AU.adm.unit">
  <base:identifier>
    <base:Identifier>
      <base:localId>adm.unit</base:localId>
      <base:namespace>UK.INSPIRE.AU</base:namespace>
    </base:Identifier>
  </base:identifier>
  <base:metadata nilReason="Unknown" xsi:nil="true"/>
  <base:member>
    <au:AdministrativeUnit gml:id="UK.INSPIRE.AU.adm.unit_50448">
      <au:geometry>
        <gml:MultiSurface srsName="urn:ogc:def:crs:EPSG::4258"
gml:id="LOCAL_ID_0">
          <gml:surfaceMember>
            <gml:Polygon srsName="urn:ogc:def:crs:EPSG::4258"
gml:id="LOCAL_ID_1">
              <gml:exterior>
                <gml:LinearRing>
```

```
                  <gml:posList srsDimension="2"
count="1267">51.32900799519856 -0.330678992765062 51.3289258380587 -
0.33071929605568934 51.32885262791949 -0.3307564035645403 51.32862843961237
-0.3306741156047602 51.32840512958921 -0.33059036116203366
………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………
51.329267760197176 -0.3302920516708573 51.32920558217105 -
0.33040914131198534 51.32914581410541 -0.33050604767327185
51.329085024475745 -0.3305943785584255 51.32900799519856 -
0.330678992765062</gml:posList>
                </gml:LinearRing>
              </gml:exterior>
            </gml:Polygon>
          </gml:surfaceMember>
        </gml:MultiSurface>
      </au:geometry>
      <au:nationalCode>E09000021</au:nationalCode>
      <au:inspireId>
        <base:Identifier>
          <base:localId>50448</base:localId>
          <base:namespace>UK.INSPIRE.AU.adm.unit</base:namespace>
        </base:Identifier>
      </au:inspireId>
      <au:nationalLevel>4thOrder</au:nationalLevel>
      <au:nationalLevelName>
        <gmd:LocalisedCharacterString locale="ENG">London
Borough</gmd:LocalisedCharacterString>
      </au:nationalLevelName>
      <au:country>
        <gmd:Country codeList="countryCode" codeListValue="UK"/>
      </au:country>
      <au:name>
        <gn:GeographicalName>
          <gn:language>ENG</gn:language>
          <gn:nativeness nilReason="Unpopulated" xsi:nil="true"/>
          <gn:nameStatus nilReason="Unpopulated" xsi:nil="true"/>
          <gn:sourceOfName nilReason="Unpopulated" xsi:nil="true"/>
          <gn:pronunciation>
            <gn:PronunciationOfName>
              <gn:pronunciationSoundLink nilReason="Unpopulated"
xsi:nil="true"/>
              <gn:pronunciationIPA nilReason="Unpopulated" xsi:nil="true"/>
            </gn:PronunciationOfName>
          </gn:pronunciation>
          <gn:spelling>
            <gn:SpellingOfName>
              <gn:text>Kingston upon Thames London Boro</gn:text>
              <gn:script nilReason="Unpopulated" xsi:nil="true"/>
              <gn:transliterationScheme nilReason="Unpopulated"
xsi:nil="true"/>
            </gn:SpellingOfName>
          </gn:spelling>
          <gn:grammaticalGender nilReason="Unpopulated" xsi:nil="true"/>
          <gn:grammaticalNumber nilReason="Unpopulated" xsi:nil="true"/>
        </gn:GeographicalName>
      </au:name>
      <au:residenceOfAuthority>
        <au:ResidenceOfAuthority>
          <au:name>
            <gn:GeographicalName>
              <gn:language>ENG</gn:language>
              <gn:nativeness nilReason="Unpopulated" xsi:nil="true"/>
```

```
              <gn:nameStatus nilReason="Unpopulated" xsi:nil="true"/>
              <gn:sourceOfName nilReason="Unpopulated" xsi:nil="true"/>
              <gn:pronunciation>
                <gn:PronunciationOfName>
                  <gn:pronunciationSoundLink nilReason="Unpopulated"
xsi:nil="true"/>
                  <gn:pronunciationIPA nilReason="Unpopulated"
xsi:nil="true"/>
                </gn:PronunciationOfName>
              </gn:pronunciation>
              <gn:spelling>
                <gn:SpellingOfName>
                  <gn:text>Royal Borough of Kingston upon Thames</gn:text>
                  <gn:script nilReason="Unpopulated" xsi:nil="true"/>
                  <gn:transliterationScheme nilReason="Unpopulated"
xsi:nil="true"/>
                </gn:SpellingOfName>
              </gn:spelling>
              <gn:grammaticalGender nilReason="Unpopulated"
xsi:nil="true"/>
              <gn:grammaticalNumber nilReason="Unpopulated"
xsi:nil="true"/>
            </gn:GeographicalName>
          </au:name>
          <au:geometry>
            <gml:Point srsName="urn:ogc:def:crs:EPSG::4258"
gml:id="LOCAL_ID_2">
              <gml:pos>51.387890740283254 -0.2869144497634939</gml:pos>
            </gml:Point>
          </au:geometry>
        </au:ResidenceOfAuthority>
      </au:residenceOfAuthority>
      <au:beginLifespanVersion nilReason="Unknown" xsi:nil="true"/>
      <au:endLifespanVersion nilReason="Unknown" xsi:nil="true"/>
      <au:NUTS nilReason="Unpopulated" xsi:nil="true"/>
      <au:boundary nilReason="Unpopulated" xsi:nil="true"/>
      <au:lowerLevelUnit nilReason="Unpopulated" xsi:nil="true"/>
      <au:upperLevelUnit xlink:href="UK.INSPIRE.AU.adm.unit_117537"
xlink:title="Greater London"/>
      <au:administeredBy nilReason="Unpopulated" xsi:nil="true"/>
      <au:coAdminister nilReason="Unpopulated" xsi:nil="true"/>
    </au:AdministrativeUnit>
  </base:member>
</base:SpatialDataSet>
```


AdministrativeBoundary


```
<?xml version="1.0" encoding="UTF-8"?><!--Created by GO Publisher Desktop
2.1.2 Build 28125 from 2012-03-22 11:26--><!--Snowflake Software Ltd.
(http://www.snowflakesoftware.com)-->
<base:SpatialDataSet xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:au="urn:x-
inspire:specification:gmlas:AdministrativeUnits:3.0" xmlns:su-
nuts="http://inspire.jrc.ec.europa.eu/schemas/su-nuts/2.0" xmlns:su-
vector="http://inspire.jrc.ec.europa.eu/schemas/su-vector/2.0" xmlns:su-
core="http://inspire.jrc.ec.europa.eu/schemas/su-core/2.0"
xmlns:hh="http://inspire.jrc.ec.europa.eu/schemas/hh/2.0"
xmlns:ef="http://inspire.jrc.ec.europa.eu/schemas/ef/2.0"
xmlns:base2="http://inspire.jrc.ec.europa.eu/schemas/base2/0.1"
```

```
xmlns:gml="http://www.opengis.net/gml/3.2"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:gmd="http://www.isotc211.org/2005/gmd"
xmlns:gco="http://www.isotc211.org/2005/gco"
xmlns:gss="http://www.isotc211.org/2005/gss"
xmlns:gts="http://www.isotc211.org/2005/gts"
xmlns:gsr="http://www.isotc211.org/2005/gsr"
xmlns:om="http://www.opengis.net/om/2.0" xmlns:base="urn:x-
inspire:specification:gmlas:BaseTypes:3.2"
xmlns:sd="http://inspire.jrc.ec.europa.eu/schemas/sd/2.0"
xmlns:gmlcov="http://www.opengis.net/gmlcov/1.0"
xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:gn="urn:x-
inspire:specification:gmlas:GeographicalNames:3.0"
xsi:schemaLocation="urn:x-
inspire:specification:gmlas:AdministrativeUnits:3.0 AdministrativeUnits.xsd
urn:x-inspire:specification:gmlas:BaseTypes:3.2 BaseTypes.xsd"
gml:id="UK.INSPIRE.AU.adm.boundary">
  <base:identifier>
    <base:Identifier>
      <base:localId>adm.boundary</base:localId>
      <base:namespace>UK.INSPIRE.AU</base:namespace>
    </base:Identifier>
  </base:identifier>
  <base:metadata nilReason="Unpopulated" xsi:nil="true"/>
  <base:member>
    <au:AdministrativeBoundary
gml:id="UK.INSPIRE.AU.adm.boundary_LOCAL_ID_0">
      <au:geometry>
        <gml:LineString srsName="urn:ogc:def:crs:EPSG::4258"
gml:id="LOCAL_ID_2">
          <gml:posList srsDimension="2" count="103">51.34019450691697 -
0.14430068235337837 51.34029541679426 -0.14408554600715215 51.3404030049217
-0.143781119249269 51.34049829405208 -0.14355184970446125 51.34055048892083
-0.14344062030323265
.....................................................................................
51.34549818392768 -0.11890689698283241 51.34560095923909 -
0.11881078814598821 51.34578436456755 -0.11863957940883894
51.345619323313926 -0.11838642501675128 51.34566823341178 -
0.11728877540238894</gml:posList>
        </gml:LineString>
      </au:geometry>
      <au:inspireId>
        <base:Identifier>
          <base:localId>LOCAL_ID_1</base:localId>
          <base:namespace>UK.INSPIRE.AU.adm.boundary</base:namespace>
        </base:Identifier>
      </au:inspireId>
      <au:country>
        <gmd:Country codeList="countryCode" codeListValue="UK"/>
      </au:country>
      <au:nationalLevel>4thOrder</au:nationalLevel>
      <au:nationalLevel>5thOrder</au:nationalLevel>
      <au:legalStatus>agreed</au:legalStatus>
      <au:technicalStatus>notEdgeMatched</au:technicalStatus>
      <au:beginLifespanVersion nilReason="Unknown" xsi:nil="true"/>
      <au:endLifespanVersion nilReason="Unknown" xsi:nil="true"/>
      <au:admUnit nilReason="Unpopulated" xsi:nil="true"/>
    </au:AdministrativeBoundary>
  </base:member>
</base:SpatialDataSet>
```

Annex D – Abstract Test Suite for the Administrative Units theme

NOTE: - Basic text as provided by the JRC in the latest ATS template version (v.1.0-16/07/2012)
- Extensions/changes/additions as required for the Administrative Units theme

1. Application Schema Conformance Class

1.1 Name test

a) Test Purpose	To ensure that the names of each instance of a spatial object type or data type specified in the Administrative Units application schema use the same designation as defined in the application schema section in the Administrative Units data specifications.
b) Test Method	Examine that the designation for an instance of a spatial object type or data type corresponds to the designation provided in the application schema section of the Administrative Units data specifications, by validating against the XSD schema with a XML validator tool (e.g. GO Publisher, oXygen XML Editor).
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.2 Attributes/associations completeness test

a) Test Purpose	To ensure that each instance of a spatial object type or data type specified in the Administrative Units application schema includes all required attributes and association roles as defined in the application schema section in the Administrative Units data specifications.
b) Test Method	Examine all instances for the attributes and association roles with the use of a XML schema validator (e.g. GO Publisher, oXygen XML Editor). Each instance shall include all attributes and association roles as defined in the application schema section of the Administrative Units data specifications.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 2, and partly 4 and 5 regarding associations, of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.3 External Object Identifier test

a) Test Purpose	To ensure that the external object identifier for the unique identification of any of the spatial objects in the Administrative Units application schema, has not been changed during its life cycle of a spatial object.
b) Test Method	Compare external object identifiers in previous versions of data with the external object identifiers of the current versions of data for the same spatial object from the Administrative Units application schema
c) Reference	• Article 9(2) in the Commission Regulation (EU) No 1089/2010 • External object identifiers of all data versions
d) Test Type	IR requirement test

1.4 Multiplicity test

a) Test Purpose	To ensure that each instance of a spatial object type, data type, attribute, and association role specified in the Administrative Units application schema, does not include fewer or more occurrences of a spatial object type, data type, attribute, and association role than specified in the feature catalogue as well as in the UML diagram for the Administrative Units theme.
b) Test Method	Examine that the number of occurrences of each spatial object type, data type, attribute and association role provided by using a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor). The numbers of occurrences for each spatial object type, data type, attribute and association role shall be compared with its multiplicity specified in the schema section of the Administrative Units data specifications.
c) Reference	• INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.5 Value type test

a) Test Purpose	To ensure that all attributes or association roles use the value type specified in the Administrative Units application schema.
b) Test Method	Each provided attribute or association role is tested with a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor) to ensure its value type adheres to the value type specified in the schema section of the Administrative Units data specifications.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 2 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.6 Constraints test

a) Test Purpose	To ensure that the instances of a spatial object type or data type specified in the Administrative Units application schema adhere to the constraints specified in the schema section of the Administrative Units data specifications.
b) Test Method	Examine all instances of data for the constraints specified for the instance's type by means of a Schematron developed in accordance to the constraints specified in the Administrative Units application schema.
c) Reference	• Section 4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 3, 10 and partly 4,5 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.7 Enumeration test

a) Test Purpose	To ensure that the values of all attributes that have an enumeration type are included in the enumeration specified in the Administrative Units application schema.
b) Test Method	The value of each provided attribute with an enumeration type is tested with a XML schema validator tool (e.g. GO Publisher, oXygen XML Editor) to ensure the value is included in the specified enumeration according to the schema section of the Administrative Units data specifications, providing that the allowed values are explicitly defined in the XSD schema.
c) Reference	• Section 4.3 of Annex II in the Commission Regulation (EU) No 1089/2010 • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

1.8 Code list test

a) Test Purpose	To ensure that the values of each attribute that have a code list type, takes only the values that are valid according to the code list's specification as defined in the Administrative Units data specifications.
b) Test Method	Compare the value of each provided code list type attribute with the values provided for the code list in the Administrative Units application schema.
c) Reference	• Section 4.4 of Annex II in the Commission Regulation (EU) No 1089/2010 • Section 7 of Annex II in the Commission Regulation (EU) No 102/2011 (amendment to Commission Regulation (EU) No 1089/2010) • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.

d) Test Type	IR requirement test

1.9 Co-administration test

a) Test Purpose	To ensure that, where applicable, if an administrative unit is co-administered by two or more other administrative units, both 'administeredBy' and 'coAdminister' associations are used as specified in the Administrative Units application schema.
b) Test Method	Check if 'administeredBy' or by case, 'coAdminister', provide associations to the correct administrative units as defined in the application schema section of the Administrative Units data specifications.
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 6 of the AU data specifications • INSPIRE Feature Catalogue and UML diagram of the Administrative Units application schema.
d) Test Type	IR requirement test

2. Geometry/Topology Conformance Class

2.1 Administrative units geometry test

a) Test Purpose	To ensure essential polygon / multi-polygon geometry consistency is maintained.
b) Test Method	Execute essential geometric check on the administrative units polygon / multi-polygon geometry.
c) Reference	• Not specifically mentioned
d) Test Type	TG additional test

2.2 Administrative units overlapping test

a) Test Purpose	To ensure that no administrative units established at the same level in the national administrative hierarchy overlap (i.e. their boundaries should not intersect with each other)
b) Test Method	Execute spatial functions/queries that would check if any two or more administrative units established at the same level in the national administrative hierarchy overlap
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 7 and Recommendation 3a of the AU data specifications
d) Test Type	IR requirement test

2.3 Administrative units coverage test

a) Test Purpose	To ensure that, together, all administrative units established at the same level of national administrative hierarchy, cover the whole territory (i.e. country level)
b) Test Method	Execute spatial function/query that would check if all administrative units established at the same level of national administrative hierarchy have an equal topological relationship with the administrative unit established at the highest level (i.e. country level)
c) Reference	• Not specifically mentioned
d) Test Type	TG additional test

2.4 Administrative boundaries geometry test

a) Test Purpose	To ensure essential polyline geometry consistency is maintained.
b) Test Method	Execute essential geometric check on the administrative boundaries polyline geometry.
c) Reference	• Not specifically mentioned
d) Test Type	TG additional test

2.5 Administrative boundaries bounding test

a) Test Purpose	To ensure that the border that limits any given administrative unit shall correspond to the geometry representing the boundary of that administrative unit.
b) Test Method	Execute spatial functions/queries that would check if the border that limits the administrative units is corresponding to the geometry representing the boundaries of the administrative units.
c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 8 and Recommendation 3d of the AU data specifications
d) Test Type	IR requirement test

2.6 Condominium test

a) Test Purpose	To ensure that the spatial extent of any defined condominium spatial object is not part of the geometry representing the extent of any given administrative unit.
b) Test Method	Execute a spatial intersection between any defined condominium spatial object and the entire 'administrativeUnit' dataset. The result should be 'false'.

c) Reference	• Section 4.5 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 9 of the AU data specifications
d) Test Type	IR requirement test

3. Reference Systems Conformance Class

3.1 Datum / CRS test

a) Test Purpose	To ensure that each instance of a spatial object type specified in the Administrative Units application schema is given with reference to the European Terrestrial Reference System 1989 (ETRS89) datum in areas within its geographical scope, or the International Terrestrial Reference System (ITRS) datum, or other geodetic coordinate reference system compliant with ITRS in areas that are outside the geographical scope of ETRS89.
b) Test Method	Check that each instance of a spatial object type in the Administrative Units application schema has been expressed using a coordinate systems using one of the datums specified above (i.e. in the Test Purpose).
c) Reference	• Section 1.2 and 1.3 of Annex II in the Commission Regulation (EU) No 1089/2010 • Requirement 12 and 13 of the AU data specifications
d) Test Type	IR requirement test

4. Data quality Conformance Class

4.1 Data quality target test

a) Test Purpose	To ensure that all data quality elements meet the specified target results in the data quality section of the data specifications.
b) Test Method	Compare the results of the data quality measure of the transformed dataset to the target proposed result specified for each data quality element specified in the data quality section of the data specifications. Results of the data quality measure have to obviously be equal or higher than the specified target result of the data quality element.
c) Reference	• Data quality section of the data specifications for the respective theme
d) Test Type	TG requirement test

5. Delievery Conformance Class

5.1 Encoding test

a) Test Purpose	To ensure that the used encoding is conformant to the encoding provided in the delivery section of the Administrative Units data specifications.
b) Test Method	Check whether the provided encoding is conformant to the encoding for the Administrative Units application schema.
c) Reference	• Administrative Units XSD schema provided
d) Test Type	TG requirement test