Master Thesis in Geographical Information Science nr 85

Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules

Luka Jovičić

2018
Department of
Physical Geography and Ecosystem Science
Centre for Geographical Information Systems
Lund University
Sölvegatan 12
S-223 62 Lund
Sweden

Luka Jovičić (2018). Spatial data harmonisation in regional context in accordance with the INSPIRE implementing rules
Master degree thesis, 30 credits in Master in Geographical Information Science (GIS)
Department of Physical Geography and Ecosystem Science, Lund University

Spatial data harmonisation in regional context in accordance with the INSPIRE implementing rules

---

Luka Jovičić

Master thesis, 30 credits, in Geographical Information Sciences

Supervisors

Lars Harrie
Dept of Physical Geography and Ecosystem Science
Faculty of Science, Lund University

Vlado Cetl
Faculty of Geodesy, Chair of Spatial Information Management
University of Zagreb

Abstract

Spatial data seamless exchange and interoperable usage has become a necessity in efficient data management and competitive positioning in the European Union. Conceptual and technical framework for the spatial data and services interoperability is specified within the EU INSPIRE Directive. The Directive provides flexible and modular structure, giving the opportunity for customisation of the data specifications and usage. From the data publisher level to the European spatial data infrastructure, this opened the question of disharmony of the spatial data structure and sharing. Arisen challenges in data harmonisation process are thus subject of interest for different formalisation approaches.

This study approaches the spatial data harmonisation process focusing on the area of Western Balkans, the region of Europe with countries that have similar interest for implementation of the INSPIRE Directive. With the main aim to propose the improvement to regional data harmonisation process, the study (1) analyses the INSPIRE data harmonisation process, (2) assesses critical factors of the process in the region and (3) tests the implementation of the INSPIRE data model formalised in accordance with user needs. INSPIRE Theme Geology is in focus in terms of its practical application to each of the objectives, throughout the thesis.

The INSPIRE data harmonisation process is analysed in order to better comprehend the process, determine standards for approach to harmonisation and provide reference to best practice cases as the source material for data publishers' capacities building. Critical factors of the harmonisation process are assessed through semi-structured questionnaire and the responses from the competent representatives of each of the countries of the Western Balkans region. Finally, the outcomes of the INSPIRE defined harmonisation process and user needs are formalised and implemented on a practical example, a geological dataset.

Results of the study present the structure and formalisation concepts of the INSPIRE data model, its extensibility, means for securing interoperability and standardised approach in defining data model elements. The responses to the questionnaire have shown that, on a regional level, spatial data managers have made certain progress towards compliance and are familiar with the Directive, but also still lack a coordinated approach and implementation guidance. Aside from the low capacities, the respondents' view is that due to the current state of the data structures, harmonisation is a highly complex process and a goal that is difficult to reach. The user needs and data model structure characteristics of the regional geology dataset that were obtained, were integrated in the formal description of the source and target INSPIRE data model. The concept required structuring the source model to meet both INSPIRE and local requirements. Source data schema and data itself are successfully transformed into the target data model and validated against the INSPIRE requirements.

The general aim was reached by implementing the INSPIRE data harmonisation with fulfilling the main objectives – creating market-oriented, interoperable and accessible dataset, meeting national legal requirements towards the geological data management and increasing efficiency of data usage. Further application of the developed approach is seen as the implementation methodology for other INSPIRE themes and other geographical regions.

Keywords: Geography, GIS, INSPIRE, data harmonisation, spatial data, Western Balkans, geology

Contents

List of Figures

List of Tables

1. Introduction

Spatial character of data plays an important role in the decision-making process within public institutions. A multidisciplinary approach in the fields of natural resources management, environmental risk assessment or infrastructural objects planning imply the need for wider use of spatial data across industrial domains, with various data sources and more utilisation possibilities. During the beginning of the 2000s, the situation in this domain showed severe limitations in spatial data utilisation in policies related to environmental issues (INSPIRE, 2003).

At the European Union level, the limitations were tackled by the framework of environmental spatial information sharing among public institutions, called European Spatial Data Infrastructure (SDI). It is formalised as the INSPIRE framework Directive for sharing geospatial information in the European Union (The European Parliament and the Council of the European Union, 2007) enabling discovery, access and data exchange from various sources, for various needs without limitation in an effective and flexible way. Starting from the fact that the data owners and providers come from diverse industry fields, working with different procedures, standards and at different quality levels, as well as the fact that INSPIRE does not enforce change of provider's data nor data model, the essence of having functional SDI is a seamless use of data. To achieve such use, INSPIRE defines implementing rules which enable provider's data harmonisation with the INSPIRE model. In the geospatial domain this approach is known as the interoperability concept. As defined by the INSPIRE Directive, interoperability 'means the possibility for spatial data sets 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 data sets and services is enhanced' (The European Parliament and the Council of the European Union, 2007).

Each organisation that implements the INSPIRE Directive can have their own approach and be unique in their view of reaching harmonisation. Across Europe, a number of different approaches and implementation models have been practiced and various experience levels were thereby reached. The data acquired in this fashion are harmonised on the EU level, in conformance with the INSPIRE requirements. Yet, specifics and consequently disharmony are still present on more local levels of data provider and user cooperation. These difficulties in reaching interoperability at local, national and regional levels come from the need for datasets and services that have a structure different from the INSPIRE specifications. Thus, striving towards interoperable and more open data on higher levels, brought out the importance of dealing with spatial data disharmony on lower levels. In recent years, various scientific projects, research and publications have been focusing on the topic of data harmonisation. For example the Humboldt (Villa et al., 2012) and the Plan4All (Neuschmid et al., 2010) are projects discussed in several scientific studies like: Hedefalk & Östman (2011); Tomas et al. (2015); Crompvoets et al. (2010); Tóth et al. (2012). These studies emphasise different aspects of the approach to spatial data harmonisation. The concepts these papers provide are discussed further in this thesis.

This study discusses the aforementioned constellation focusing on a group of countries with mutual characteristics of interest for implementation of the INSPIRE Directive. The area can be referred to as the Western Balkans region, as used in the EU international affairs correspondence (Commission of the European Communities, 2008). The Western Balkans in this sense relates to the EU enlargement process initiative and consists of Albania, Bosnia and Herzegovina, the former Yugoslav Republic of

Macedonia (FYR Macedonia), Montenegro, Serbia, Kosovo* and Croatia (until joining the EU). This group has another historically tight related member – Slovenia, since all of the countries (except Albania) used to be part of former Yugoslavia. These countries share mutual historic development through which they built common political, infrastructural and social background of governance. Such development reflected institutions' organisation, standards and information exchange as key factors for spatial data management (Cetl, Tóth & Smits, 2014). In this context, the study will focus and further refer to these countries as the Western Balkans region (including Croatia and Slovenia) in analysing ways of data harmonisation and the possibility for formalisation to better meet institutional and INSPIRE requirements towards seamless data sharing goals. The second, more formal reason for focusing on the chosen area is that all Western Balkans countries acknowledge and are in various stages of incorporating the INSPIRE Directive in their laws, thus accepting the concepts and obligations, as well as recognising the benefits of implementing such an infrastructure (Ogrizovic, 2013).

1.1. Problem statement

On top of the shared development history of the governmental institutions within the Western Balkans countries, adoption of the legislature for the INSPIRE harmonisation process shows common challenges and builds very slowly in comparison with other countries (Aleksic, 2013). The first of these challenges can be formulated as the difference in structuring spatial data as the basis of supply of the required information. Spatial data as such is considered to have the required formats, to be in an up-to-date state and in conformity with the international standards as recommended by the INSPIRE implementing rules. The second important challenge is legislation, which needs to support data availability and accessibility. This is not the case in countries where data, products and services specifications date from decades ago, causing confrontation with the current requirements for data exchange, security issues and standards applied. The third challenge is the non-existence of supply and use of spatial data through web services, in a standardised structure and in an interoperable manner. This requirement brings out the question of technical capacities and know-how within the institutions.

The presented constellation reveals data structuring, national legislation and institutional capacities as critical factors in the endeavour for utilising data harmonisation within the Western Balkans. In addressing these challenges, the question is how can the INSPIRE harmonisation process serve as the basis for seamless data exchange in the Western Balkans region?

1.2. Aim

The main motive of this study is to contribute to the qualitative approach to spatial data infrastructure development in Western Balkans, through the research of the INSPIRE Directive implementation. In light of that, the thesis general aim is to propose an improvement to the regional data harmonisation process through a technical approach for formalising the INSPIRE data model. Specific objectives for reaching the general aim were (1) studying the INSPIRE data harmonisation process, (2) analysis of the critical factors of the process in a regional context and (3) testing the implementation of the INSPIRE harmonisation process covering one INSPIRE spatial data theme. The results of these objectives provide the approach to developing a formalised INSPIRE model for the referred INSPIRE theme in a regional context.

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244/1999 and the ICJ Opinion on the Kosovo declaration of independence

1.3. Methodology

This thesis examines the stated problem firstly by studying the implementation of the INSPIRE data harmonisation, focusing on geology as a chosen thematic domain and the related INSPIRE implementing rules. Such approach was taken in order to retain the standards as a basis, to address other levels of cooperation (national and regional SDI), to review the underlying concepts of the harmonisation process and review examples from practice. Theoretical aspects are followed by study of identified critical regional factors of the harmonisation process within the referential organisations throughout Western Balkans. Analysis of referential institutions' approaches to the harmonisation process is presented through a questionnaire. Afterwards a practical example within one Western Balkans national SDI stakeholder institution (FYR Macedonian Geological Survey) is studied. For this purpose, test data is being harmonised, following the general INSPIRE implementing rules for the theme Geology. The final output represents the developed approach to formalisation of the data harmonisation for INSPIRE theme Geology in order to improve regional INSPIRE implementation and spatial data management.

1.4. Expected results

The main contribution of the research is the approach to formalisation of the INSPIRE data model for seamless harmonisation process of the INSPIRE theme Geology, in the Western Balkans region. On a higher level, with applicable standards applied and following the proposed methodology, an equivalent approach could be developed for other applicable INSPIRE themes and other geographical regions.

1.5. Disposition

The study is formed with the following structure: In Chapter 1 the idea and the approach are formulated by describing the subject area, problem definition, aim and the methodology in order to achieve the envisioned results. Following that, Chapter 2 gives the theoretical background of the research and constellation in the field subject area, by analysis of the core domains of interest for the thesis. These domains are the legislative, the standards and experiences in INSPIRE Directive data harmonisation for the theme Geology on European and regional - Western Balkans levels. The theoretical background forms the base for Chapter 3 and the first study which analyses regional critical factors for the implementation of the INSPIRE data harmonisation. This was achieved by design and dissemination of a questionnaire which resulted in description of the critical factors. The second study is described in Chapter 4, covering the INSPIRE data harmonisation process practical implementation in a national SDI context. The results of the practical segments of the work are discussed within the Chapter 5 which reflects on the INSPIRE data harmonisation by evaluation of the two studies results and reaching the main aim of the thesis. In the end, Chapter 6 reviews the outcomes of the taken approach and the contribution to the INSPIRE data harmonisation on the national and the regional level.

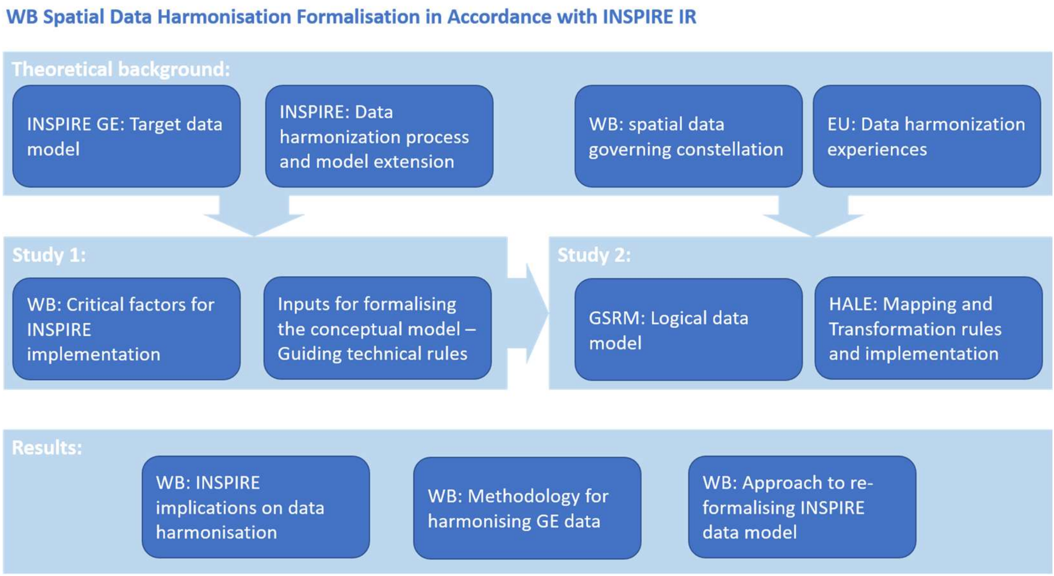


Figure 1-1 Work composition flow

WB: *Western Balkans region,* ***INSPIRE IR:*** *INSPIRE Implementing Rules,* ***INSPIRE GE:*** *INSPIRE Geology Theme,* ***GSRM:*** *The Geological Survey of the Republic of Macedonia,* ***HALE:*** *Humboldt Alignment Editor software*

2. Background

The environment in which spatial data harmonisation process occurs is described firstly from the legal aspect in the section *INSPIRE Directive and Implementing Rules*, which discusses the approach developed from the needs on the European level. Secondly, the section *Data harmonisation process – the INSPIRE Theme Geology* focuses on one of the INSPIRE Directive themes, describing its structure and key concepts. Along with the following section *INSPIRE model formalisation rules,* these sections form the basis for the following studies later in this thesis. The formalised way of performing the existing process and approaches to the extension process are discussed in the next section - *Underlying concepts of INSPIRE data model formalisation*. Following this top down approach, the regional context is presented in the section *Western Balkans countries commons in context of spatial data govern*, putting in focus the area of interest, describing its specifics and key elements that influence the existing harmonisation process. Finally, scenarios in the section *Overview of the experiences in data harmonisation process*, assess practical implementations of the previously discussed aspects of the harmonisation process. In this manner, the envisioned goals are given the base, means and comparability to be assessed in the following chapters.

2.1. INSPIRE Directive and Implementing Rules

The INSPIRE (*Infrastructure for SPatial InfoRmation in Europe*) Directive takes an approach towards enabling spatial data availability, sharing and better use on the EU level. The Directive represents a framework for establishing the European spatial data infrastructure. The environmental policies and related activities are primarily in the thematic focus. INSPIRE defines an infrastructure architecture and relations of its elements with the aim of enabling seamless discovery, access and data exchange from various sources, for various needs without limitation in an effective and flexible way. (The European Parliament and the Council of the European Union, 2007). The main principles behind INSPIRE are (European Commission, 2017a):

- 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.

INSPIRE divides spatial data, activities and related themes into the three annexes covering 34 spatial themes in total. INSPIRE technical architecture provides the framework for the spatial data themes using the Implementing Rules (IR) for (Figure 2-1): (1) Metadata (which refers to the information describing data and services), (2) Data specifications (focusing on the data harmonisation based on their content and spatial component), (3) Network services (describing the technology and standards enabling spatial data and metadata availability), (4) Data and service sharing (consider terms of sharing including those referring to spatial data accessibility), (5) Spatial data services (technical specification for the harmonised services defined by the INSPIRE) and (6) Monitoring and reporting (in the context of the INSPIRE implementation requirements).

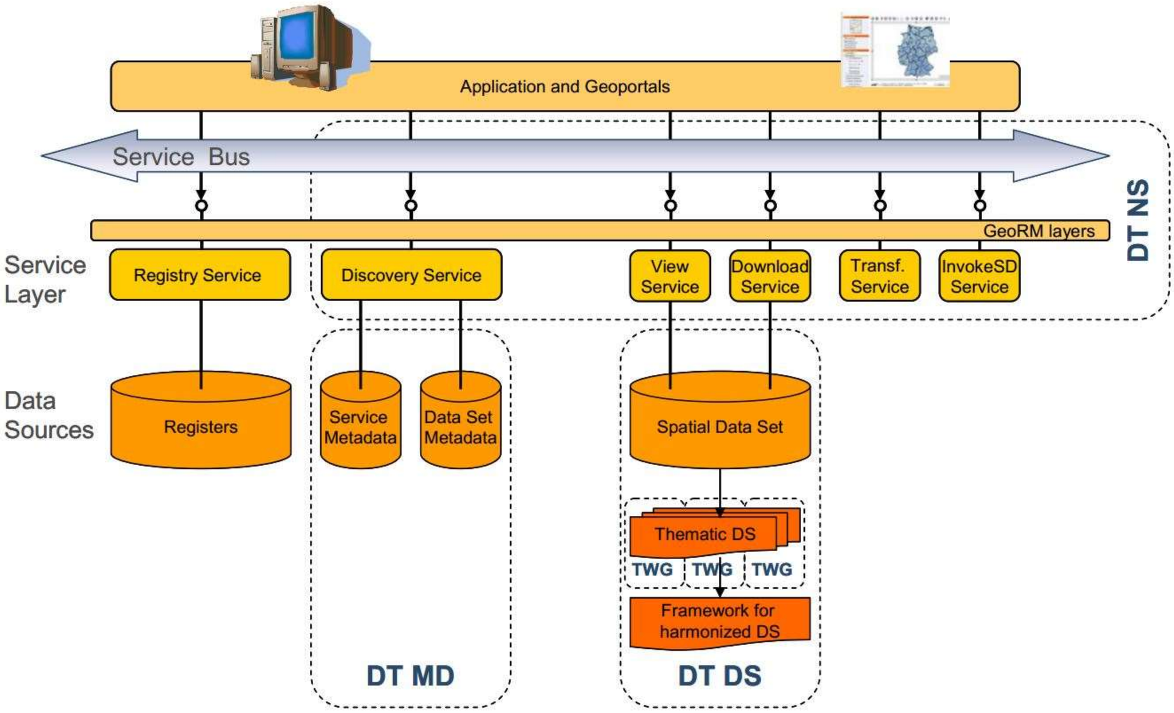


Figure 2-1 INSPIRE technical architecture overview (Drafting Teams 'Data Specification' 'Network Services' 'Metadata', 2007)

The IR specify the process of INSPIRE implementation as a whole and are mandatory for all member states.

The IR on Metadata specify managing rules, structure and contents of the datasets and services information describing the INSPIRE themes. These were to be provided in required form until the end of 2013 (Figure 2-2).

IR Data specifications formalise requirements towards data models in order to define each of the INSPIRE themes specifics for interoperable usage. Corresponding documents that concretise conceptual IR are Technical Guidelines which are not obligatory like IR but specify ways and means of IR implementation as well as support to data providers in the data provision. This is the core concept for securing interoperable usage of spatial data across Europe. In that sense, Data specifications are the focus IR of this thesis and are discussed in more detail. The INSPIRE roadmap sets the ground for the implementation of new and extensively restructured data sets interoperability rules for the Annex I themes during 2013 while Annex II and III were to be available by the end of 2015. Spatial datasets still in use, depending on the Annex they belong, are to be available in the timeframe from the end of 2017 up to end of 2020.

Network services IR treat common interfaces for web services, enabling client applications and users to interact on an EU wide level. Web services refer to discovery, view, download, transformation and invoke services. According to the INSPIRE roadmap, network services (except invoke services) were to be implemented by the end of 2013 (Figure 2-2).

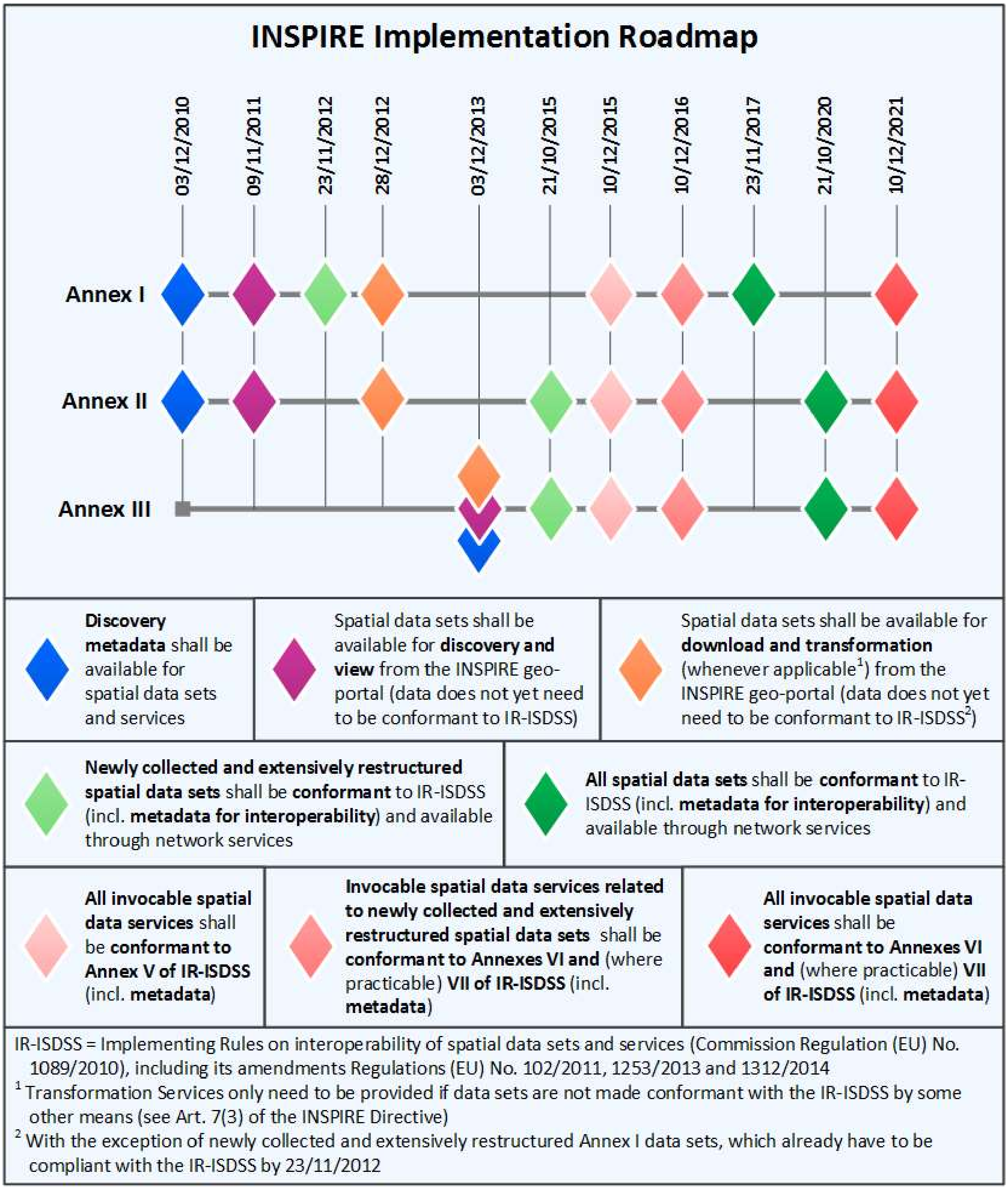


Figure 2-2 Inspire implementation roadmap (European Commission, 2015)

Data and service sharing is another subject of the INSPIRE IR, which is described as harmonisation terms which enable accessibility and usability of data and services. These are defined on a conceptual level, leaving concretisation to the individual national regulating bodies. Their full implementation was supposed to be provided by the end of 2013.

Spatial data service IR regulate interoperability of services themselves. This term considers seamless communication, execution and data transfer among services. This group of IRs are considered as advanced services and are out of the scope of this thesis but are mentioned in order to show the big picture of the whole INSPIRE process. These are to be implemented in steps from the end of 2015 until 2021.

Finally, the IR on Monitoring and reporting describe requirements on advance and quality proofing of the INSPIRE implementation progress at the Member states' level, on an annual basis. These refer to a number of implementation indicators statuses, inventory of included datasets and services, usage of the spatial data infrastructure and its performance.

The IRs are decomposed further within the Technical guidance documents which describe possible ways of IR implementation. It is specified that a localised approach is favoured, respecting national needs and regulations in Europe. Observing the contents of the IRs, their conceptual character can be divided into regulation of the spatial data technical environment and regulation of the spatial data itself. In that sense, Data Specification IR represent a core component in enabling interoperability and harmonisation.

2.2. Data harmonisation process – the INSPIRE Theme Geology

This section presents the INSPIRE Data Specification IR related to the theme Geology as the concretisation of the previously described spatial data harmonisation framework. The focus is on the data model structure and relation to the stakeholder needs. The core of the data model are the application schemas which formalise the EU level target data model.

The INSPIRE Theme Geology IR relates to stakeholder needs by firstly defining the scope, purpose, use cases, model limitations and further development in order to provide better comprehension of the technical framework for geologists. In that sense, for the purpose of providing interoperable geological information, requirements from spatial data providers refer to geological materials characteristics (composition, structure, age, etc.), groundwater information as well as geomorphology of the rocks, boreholes and geophysics. Moreover, geological data goes broader than just Theme Geology and also covers sections of some other INSPIRE themes (Mineral Resources, Natural Risk Zones, Soil, Energy Resources). Thus, the field of geology includes several data models which are also further referenced by other themes. Although the thematic scope is clearly defined, in practice and in logical data modelling, this situation causes significant complexity of expressing geological objects and relations.

Conceptual data schema was developed by an international group of geology experts. The schema is based on the complex data model GeoSciML (more details on this model in section 2.4.1), which enables description of broader geological characteristics, thus allowing for model extension according to user needs.

Apart from this group of requirements from data producers, the IR involve some general elements and thematic specifics that contribute to interoperable data usage. These refer to cardinality, domain values, constraints etc. Using standardised *Unified Modelling Language* (UML) enables data model specification, automatic processing, encoding, querying and updating. UML is further explained in section 2.4.2. These general elements are covered either by the INSPIRE data model stereotypes or specific concepts like identifiers and geometric representations shown in *Appendix A - Main UML stereotypes in INSPIRE Geology schema.*

Data specification for the theme Geology defines geological, hydrogeological and geophysical models. Of interest for this thesis is the geological model and the approach to its formal description is presented in *Appendix B - Theme Geology Conceptual data model*.

The schema diagram and the description of its elements in the Feature Catalogue show that elements of the INSPIRE model are well documented in terms of their type, explanation, comprehension, cardinality and adhering concept. In this way geologists are given a general model as a target in building INSPIRE compliant geological data. This is also the basis for building up a model that suits regional needs.

2.3. INSPIRE model formalisation rules

There is a number of approaches which provide the main concepts of INSPIRE model formalisation. All of these are based on the INSPIRE Generic conceptual model (GCM) and develop according to specifics that refer to the different user requirements. The following sections discuss firstly the GCM, after which several examples are given in order to present the conceptual approach to the INSPIRE model formalisation.

2.3.1. INSPIRE Generic conceptual model

Reflecting on the process of enabling data and services interoperability and usage within the spatial data infrastructures in Europe, the question of the approach to processing different thematic datasets in an interoperable manner is raised. INSPIRE thus took the approach of defining conceptual framework, which provides a repeatable data specification development methodology and general provisions for the data specification process which is valid for all spatial data themes (Tóth et al. 2012). Such a framework is known as the Generic conceptual model (INSPIRE Drafting Team 'Data Specifications', 2013) describing the concepts of data modelling and data specification development. It describes elements of the data specifications that can refer to any of the thematic groups as well as their relations and interdependencies (Figure 2-3).

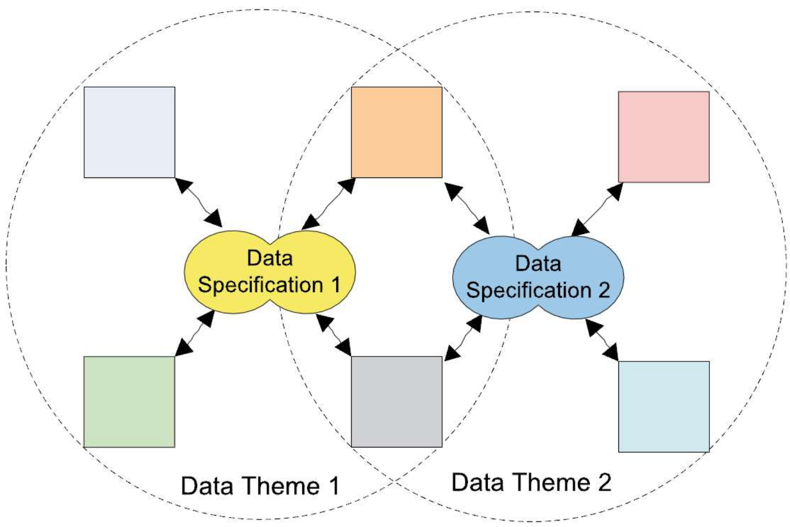


Figure 2-3 Cross theme interoperability (Tóth et al. 2012,p 25)

The Generic conceptual model is on a higher level of abstraction in comparison to IR, describing their elements in conceptual level and providing means of their formalisation. This document is also a first stop in extending the basic INSPIRE data models, supporting and covering the extensions approach through the definition of core principles, where (INSPIRE Drafting Team "Data Specifications" 2013, p 128.):

Extending an INSPIRE data specification would imply at a minimum that:
- the extension does not change anything in the INSPIRE data specification but normatively references it with all its requirements
- the extension does not add a requirement that breaks any requirement of the INSPIRE data specification

However, the extension may, for example, do any of the following:
- add new application schemas importing INSPIRE or other schemas as needed
- add new types and new constraints in your own application schemas
- extend INSPIRE code lists as long as the INSPIRE data specification does not identify the code list as a centrally managed, non-extensible code list
- add additional portrayal rules

In addition to these general rules that are mainly implied by the rules of UML, further harmonisation will be achieved, if the extensions conform to all requirements of this document and the document —*Guidelines for the encoding of spatial data*.

The INSPIRE Generic Conceptual Model contains all definitions for the data models' formalisation, known as the INSPIRE UML profile. Each INSPIRE application schema has to be defined with UML version 2 according to ISO 19103 and ISO 19109. The INSPIRE Generic Conceptual Model furthermore defines which stereotypes that are allowed to be used in the INSPIRE UML Model. Most of the stereotypes have already been defined by other standards and are reused here (Mihaljević, 2011). In section 2.4.2, the role and characteristics of UML for the process of model extension are explained more thoroughly.

2.3.2. Data transformation solutions project - Extending INSPIRE Data Specifications

A spin-off of the Fraunhofer Research Institute, *wetransform,* focuses on data transformation and development software solutions in this niche. Based on INSPIRE, ISO and OGC as spatial standards, they developed HALE – an open source desktop software for data transformation used around Europe as one of the most comprehensive solutions offered. With the support of the European Commission's Directorate-General for Environment, the Joint Research Centre and the European Environmental Agency, *Extending INSPIRE Data Specifications* project was conducted by experts in the field of modelling and implementing the INSPIRE models from the data specifications. During 2016, the *wetransform* project resulted in an inventory of the current INSPIRE Model Extensions, the extension pattern catalogue and an end-to-end tutorial project. It recognises the INSPIRE Generic conceptual model, thus the focus and the contribution to the model formalisation approach is the methodology of the model extension.

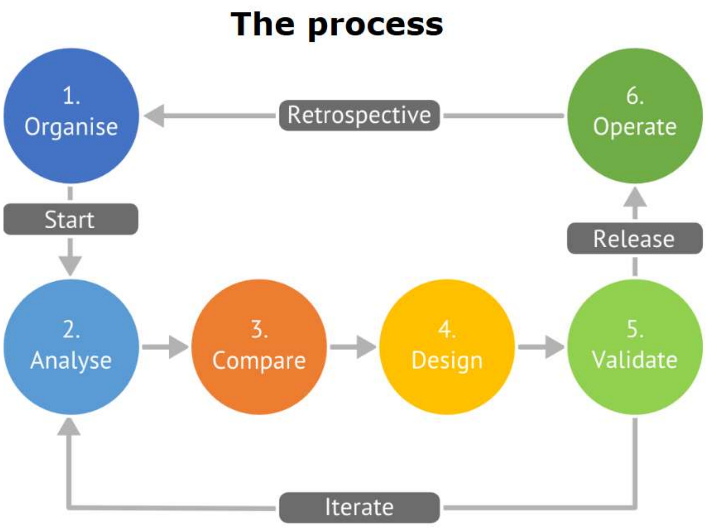


Figure 2-4 Geonovum study on INSPIRE model extending, Process flow (Grothe, Bulens & Reitz, 2016)

Methodology firstly considers implementing the model by defining the work process, which serves as the input for the analysis phase. The analysis phase requires identification and studying of the requirements in terms of work processes and spatial data users, evaluation of local models and INSPIRE specifications. The next step is the comparison of the requirements between the existing and the INSPIRE specified models. The core phase is the extended model design which follows the described rules and builds upon the INSPIRE Generic conceptual model, links the required new classes to INSPIRE using UML concepts (Aggregation, Inheritance, Association, Composition, Multiple Inheritance) and adds new classes' properties, whose linkage to INSPIRE properties is guided by the UML approach on defining constraints and code lists. Finally, the process of model design needs to be validated to test the compatibility with INSPIRE Data Specifications. Implementation of the designed model is performed using available data model transformation software. On the implementation level, testing and validation is done in regards to the target platform and required data usage. Finally, model extension can be deployed for public testing and, when needed, the whole process can iterate until the model reaches an acceptable level.

2.3.3. ELF project approach in formalising the INSPIRE data model

The European Commission funded European Location Framework (ELF) project addressed the need for more detailed data in the cross-border areas in Europe, focusing on the existing spatial products and underpinned legacy requirements, like the EuroGeographics (the European National Mapping, Cadastral and Land Registry Authorities) products - EuroBoundaryMap, EuroRegionalMap, and European Commission Eurostat needs. The project formalised activities by building the technical infrastructure with a single access point to pan-European services. Of interest for this thesis is the ELF infrastructure component of the ELF data specifications (ELF DS), which describes the conceptual data model for creating harmonised pan-European reference data.

ELF DS follows the INSPIRE recommendations, namely INSPIRE Data Specifications and data model concepts defined within GCM. On the other hand, it incorporates user requirements through formalisation of the multi-scale compliant model which contains schemas covering themes from all three of the INSPIRE Annexes. The main principle of retaining INSPIRE compliancy requires a defining data model development approach, building it up through standardised modelling guidelines and finally implementing it through the ELF data model and application schemas. The approach is depicted in the Table 2-1.

Table 2-1 Adopted from (Hopfstock, 2016) *Table 2-1 – ELF extensions*

Identification of:	**Extensions**	**Data Specification**	**Data Model**
corresponding concepts between INSPIRE and ELF model	Common part of the two models	Data Quality conformance criteria validation	Specification is implemented through Simple inheritance / specialisation concepts.
		Establishing data capture criteria	n/a
concepts present in INSPIRE but missing in existing data	Restricted part	Application schema not considered for ELF	
		Feature types ignored	
		Data types, attributes, associations	constraint
concepts present in existing data but missing in INSPIRE	Extended part	New theme	Add application schema
		New feature types	Add feature type
		New data types or attributes	Add data type or attribute
		New associations	Add association

Developing of the ELF data specification took into consideration various needs of the stakeholders, like respecting different levels of detail, user requirements survey results, existing data specifications constraints etc. The Specification forms a framework, for which implementation is formalised within the ELF Modelling guidelines. The guidelines also describe the process of the INSPIRE data model extension. Matching concepts between schemas are analysed and in accordance to the described approach in the ELF DS, following the rule of inheriting INSPIRE concepts characteristics as needed, like adding optional attributes missing in INSPIRE, adding constraints to ensure ELF requirements, defining new feature types not present in INSPIRE, adding items in code lists, adding optional associations and implementing other applicable INSPIRE GCM conventions.

ELF project presented another way of realisation of the adopted concepts for developing the INSPIRE model following INSPIRE GCM. The key aspect of the Project approach is defining user needs and focusing on their formalisation in the process of model building in compliance with INSPIRE GCM.

2.3.4. The GeoSmartCity

The GeoSmartCity (GSC) project started with the aim to contribute to added value applications and services development using geographical open data from the cities towards the Smart Cities concept (European Research Institutes' initiative towards low-carbon Europe). The project aimed at building up a framework for utilising available data through the creation of pilot applications and services. It draws on the consolidated standards INSPIRE represents. Its orientation towards an interoperable environment, enabling various data providers and users to cooperate, has led to solutions like GeoSmartCity Data Catalogue, Data Portal, Client Side API and support resources like Validation Service, Codelists Manager etc. (Reitz, 2016)

In regards to data modelling, GSC firstly used a template model requirements file from the users, then did the comparison with INSPIRE Data Specifications, aligning requirements grouped into common classes, after which, the needed extensions in accordance to the acquired requirements were created, along with schema maintenance and revision instructions, and finally validation of the results through available software tools.

The process of extending the data model considered adding new attributes which were guided by the principle of inheriting INSPIRE feature types characteristics. Also, new feature types were introduced to the model in order to formalise concepts exceeding the ones covered by the INSPIRE schema. The same was done in relation to code lists which were not present in the INSPIRE model. As an automation approach, open source software tool Re3gistry was used to manage code lists and values. (Martirano, Morrone & Vinci, 2016)

Using an equivalent approach as the previously presented INSPIRE data model extension examples, the GSC project, being in line with the INSPIRE Generic Conceptual Model, emphasized the core of formalising the INSPIRE model. Namely, as INSPIRE Data Specification Technical Guidelines are not mandatory but optional and the example case of Implementing Rules implementation, there are drawbacks in their usage. The first, general issue is that these guidelines are just a draft data model, and second, it stems from the need for major changes in order to meet GSC requirements. Thus, the GSC approach in extending the INSPIRE model rests on using schemas that are included in the legally binding INSPIRE Implementing Rules, so that INSPIRE core schemas conformity is secured, as well as the fact that model development starts from a stable point specified in the Implementing Rules.

2.4. Underlying concepts of INSPIRE data model formalisation

Previous sections discussed the process of INSPIRE data model implementation and formalisation. Data model development process phases cover defining a source data model and requirements, understanding their relation to the target – INSPIRE data model specification, schema and data transformation, and finally, validation of the resulting schema and dataset.

Basis for spatial data harmonisation on a conceptual level is secured using standards as formalisation of structural and semantic levels of interoperability. In the geospatial domain, these standards are provided by the International Standardisation Organisation (ISO), the Open Geospatial Consortium (OGC), and the European Committee for Standardisation (CEN). They regulate methods, tools and services for managing spatial data in digital form for the use among various users, systems and locations. Implementation of the extended data model using these standards is made available with software tools. Firstly, software enables notation of data structure and schemas in a standardised and interoperable way. Further, schema and data transformation are subject of several software solutions that base their concepts and offered results on the common spatial standards. Finally, as a validation of the process outcomes and conformance to required standards, formal rules exist to which data and services need to comply. These are defined by the standardisation organisations but are available as automated processes only upto a certain degree. The following subsections present these main characteristics of the data model extension process.

2.4.1. Standards regulating geological spatial data structure - GeoSciML

INSPIRE Directive describes the way of reaching an interoperable state of data and services, through standards prescribed by the ISO and the OGC, incorporated in conceptual data models contained in the IR. For the theme Geology, the data model used is GeoSciML, generally described in Richard & CGI (2007). It represents the OGC adopted standard framework for geoscientific data encoding based on Geography Markup Language as standard for representation of features and geometry. GeoSciML started as a model for capturing geologic maps and observations (boreholes) information. It is used as the data interchange format, which can be added on to existing data systems, thus avoiding restructuring producers' current data formats.

Explaining the GeoSciML and examples of its usage within INSPIRE through the EU project OneGeology-Europe (Laxton, Serrano & Tellez-Arenas, 2010) present an implementation of the standard for the geoscience information exchange. GeoSciML formalisation through UML concepts and underlying ISO standards shows the core of the data model. Thus, it allows expressing complex use cases and extension of the model to meet the provider requirements, while retaining an interoperability level at the same time. Even the INSPIRE data model for the theme Geology represents a simplified version of the GeoSciML segments (INSPIRE Thematic Working Group Geology 2011, p.286).

Figure 2-5 gives an overview of the GeologicFeature context diagram, expressed in UML as part of the application schema for GeoSciML 4.1. Core elements are present in the INSPIRE schema but hold more abstract concepts as well as links to different vocabularies. This thesis does not perform a more detailed analysis, since this is out of its scope. Of interest is simply to emphasise that the INSPIRE data model itself is an example of a more generalised model extension, retaining a standardised and interoperable structure, which remains compliant with underlying standards after modifications.

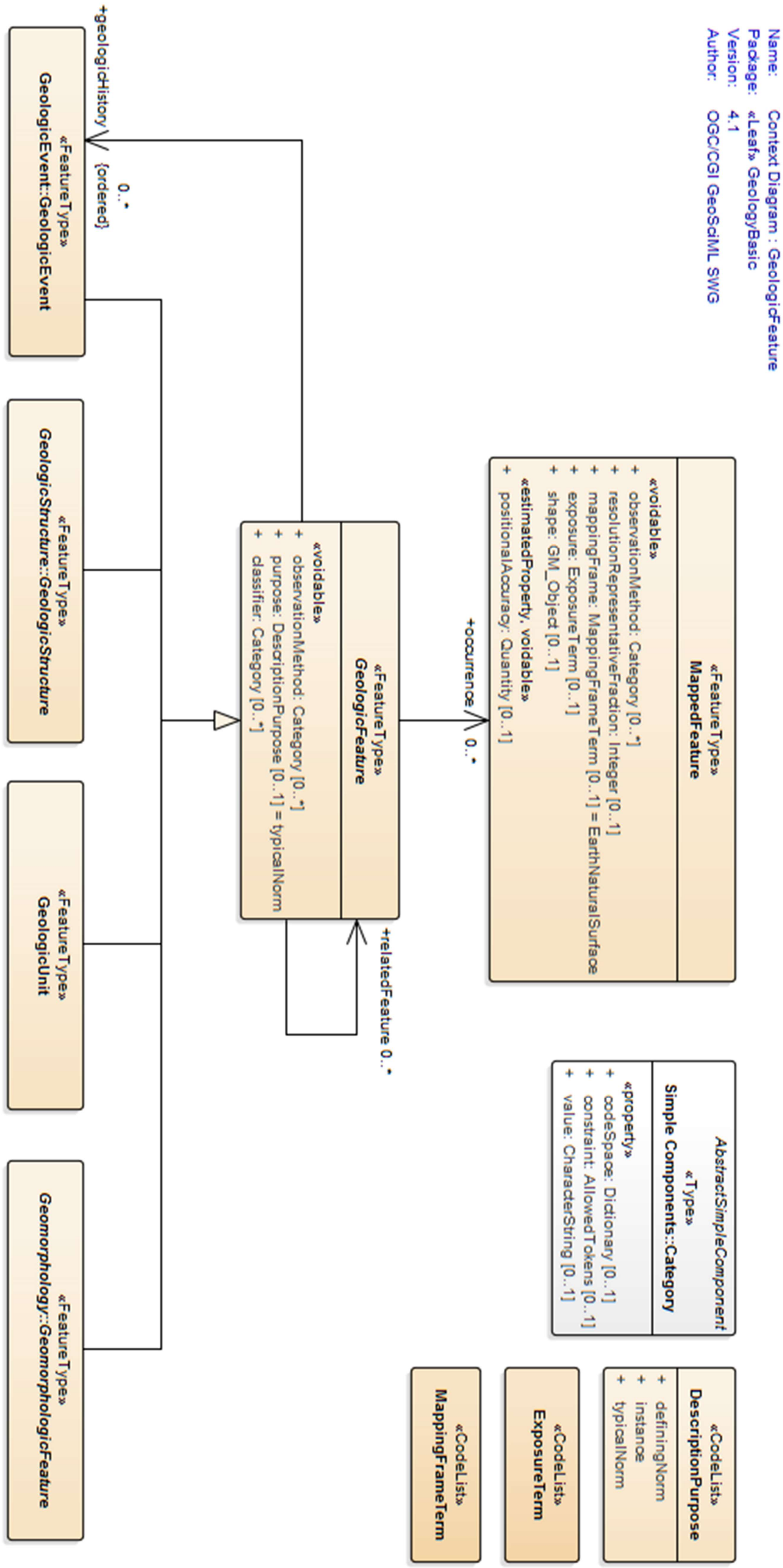


Figure 2-5 GeoSciML application schema segment – GeologicFeatue (IUGS Commission for the Management and Application of Geoscience Information, 2013)

2.4.2. Data model formal description

INSPIRE data model Concepts and user requirements notation is defined as the INSPIRE UML profile. Unified Modelling Language (UML) is a modelling language designed for general purposes in software engineering. It conforms to a standard created and managed by the Object Management Group. UML syntax is a set of graphic notation techniques. It is used to specify, visualise, modify, construct and document the details of an object-oriented system under development. UML profiles provide means of adopting UML to certain areas of application. This is accomplished by extension mechanisms which allow adding new elements to the syntax in a way that facilitates the application of UML on a domain of interest, while avoiding the contradiction with the standard semantics. Profile is a collection of extension mechanisms (stereotypes, tag definitions, and constraints) that are applied to specific model elements (Classes, Attributes, Operations, and Activities). (Alhir, 2002)

Different UML profiles exist and are used to formally describe data models. On EU level ISO19103 ("Geographic information — Conceptual schema language") UML profile is used. Furthermore, this model, used as INSPIRE UML profile is not conformant to the core UML specification, but extends it with additional elements. As it was mentioned earlier, GCM defines stereotypes used in the INSPIRE UML model. Thus, it carries UML syntax, but builds up new semantics on top of it.

In terms of harmonising regional spatial thematic data this means that there is a need for tools that support managing the INSPIRE UML model in order to formalise and extend the model according to user requirement. Furthermore, tools for encoding the model are needed to enable the transfer process and services for accessing data, namely for the support of GML for INSPIRE and specifically GeoSciML in case of the geology theme. Enterprise Architect, which was used in the examples described in section 2.3 on model extension, was proven as a compliant tool for model description in that sense.

2.4.3. Schema and data transformation tools

Data harmonisation is a process of making data conformant with the referent characteristics that the target data model specifies. As a concept, Tamash (2012, p.12) defines harmonisation as a process that handles: syntax (that represents the data format that is covered by the referential standards), structure (meaning mapping of the source to the target data model schema) and semantics (as the meaning of the concepts in defined context, which relates to referential vocabularies).

Core phase of the harmonisation process is defining the structure, i.e. mapping of the data model schema. Östman & Abugessaisa (2014) define schema mapping as a three-phase process. The first phase is the adaptation of the source schema (which entails the identification of the semantically related objects by comparison of attributes, their meaning and representation), the second is schema mapping (finding transformation rules like data conversion, merge, split etc.) and the third phase is the schema transformation (extracting data from the source schema, transformation according to the mapping rules and loading into the target schema). Additional requirements towards the source data model and data set consider: specification of the coordinate reference system, measurement units and grid for data representation in a uniform manner; domain values consistency, which requires data and data type level conformance with the domain defined values; conceptual consistency, or conformance with the conceptual data schema on spatial objects and spatial object types level; metadata completeness. The stated transformation processes are to various extent automated through the use of the software tools that conform to the standards and requirements INSPIRE prescribes.

Available tools are discussed in Tamash (2012) and classified according to their scope of use, functionalities and usability. Core functionalities relate to the harmonisation process phases handling, including spatial data geometric and attribute management. These include, for example, coordinate systems reprojection, spatial transformation, managing data inconsistencies, merging data, data quality evaluation, formats transformation. For this thesis, robust and widely used software tool with INSPIRE user community and open source platform - Humboldt HALE is chosen.

The Humboldt HALE software resulted out of the EU Humboldt project (Villa et al., 2012), focused on the integration of the EU level spatial data spread across domains and organisations. The project aimed to contribute to a higher level implementation of the European Spatial Data Infrastructure. Of interest for this thesis are the project activities in related to the technical enabling of the data harmonisation process. Project outputs in this sense provided the means for enabling documenting and harmonisation of organisations' spatial data through: applications for specification of data conceptual schemas and schema transformation; services which enable interoperable use of data in conformance with INSPIRE requirements; transformation services for harmonisation purposes. Figure 2-6 depicts the structure of the Humboldt framework technical environment. All components are published as open public licences (GNU Lesser General Public License version 3 (LGPL v3)) and are freely available.

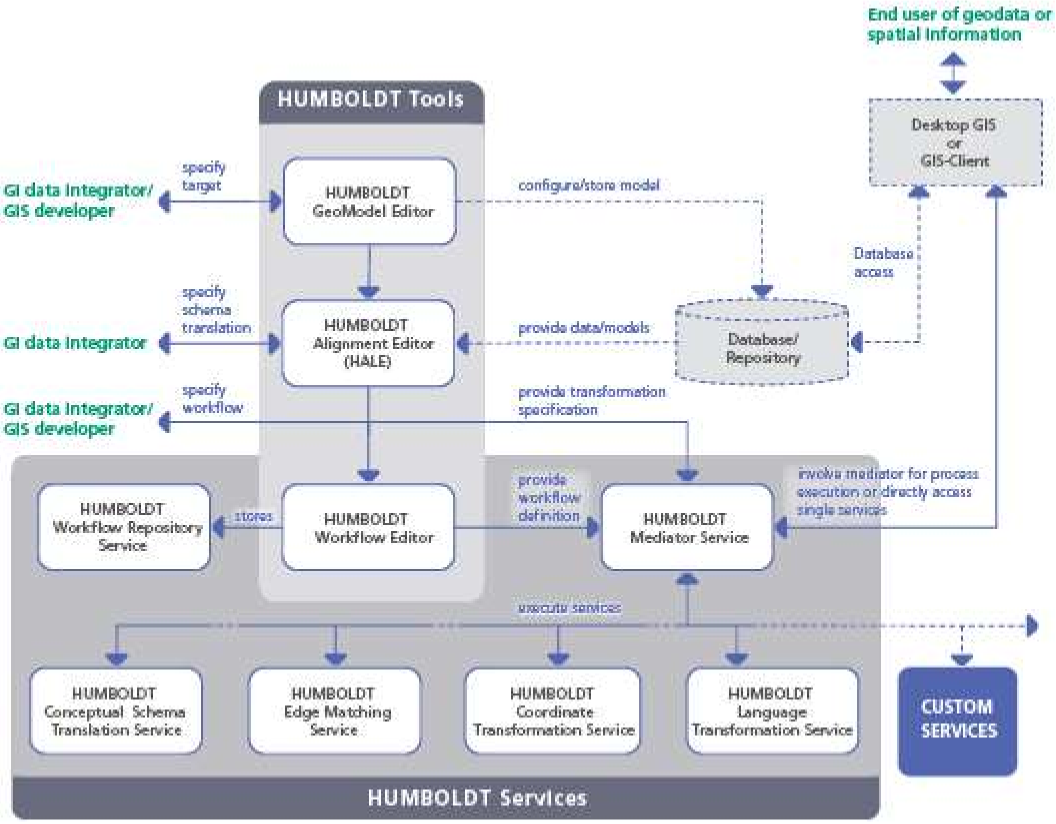


Figure 2-6 The HUMBOLDT Framework (The HUMBOLDT project, 2010)

Humboldt HALE stands for HUMBOLDT Alignment Editor and represents basically a tool for schema transformation. Other project tools cover modelling of schemas, processes, implementation and analysis of data transformation as well as transformation services etc. HALE Schema transformation tool enables mapping between different conceptual schemas through the capability of creating logical and semantically valid connections. These connections serve as the basis for data transformation.

HALE has an intuitive interface, a well documented and straight forward way of implementing functionalities. Its strengths are availability, compliance with INSPIRE, its spectrum of functionalities and intuitive usage.

2.4.4. Validation of the data model

Schema and spatial data transformation are the main output of the harmonisation process. Thus, these are the subject of the validation against the requirements of the INSPIRE and underlying standards. This, in turn, defines the “what” and the “how to convey” regarding the harmonisation process quality control.

On a conceptual level INSPIRE GCM treats this topic in the segment Conformance testing as the interoperability quality control step. Namely, validation considers testing data specification requirements that refer to each conceptual data model class of the referential spatial theme. The formal procedure for testing conformance level is defined in the INSPIRE context with the Abstract Test Suite (ATS), which defines test cases. ATS is the formal basis for creating Executable Test Suite (ETS) as the concretisation of the ATS parameters which can be automated.

Tools used for formalisation of the data model, structuring and transformation into target data model are still being developed in accordance with INSPIRE requirements. During the research on this thesis, the European Commission Joint Research Centre released the official version of the INSPIRE validator, developed under ARE3NA, Action 1.17 of the ISA Programme (JRC, 2017).

Checking the conformance level as well as the extension of the target model implies that the resulting data set needs to be validated using the ATS. Depending on the spatial data theme available, there are automated testing tools like OGC Validator Test Suite for GML version 3.2. This tool can, at least partially, test the application schema, while complete testing requires further manual use of ATS/ETS. For example, for the theme geology, the OGC Validator Test Suite offers automated one third of the applicable tests.

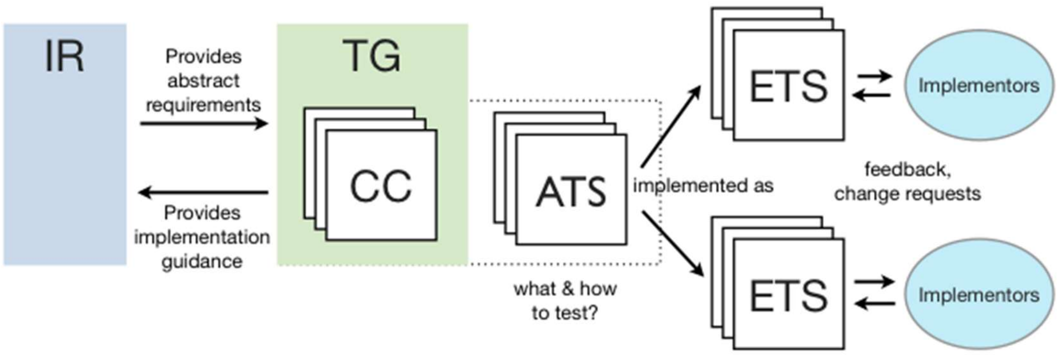


Figure 2-7 Schematic on relations of INSPIRE IR and TG requirements and the Abstract and Executable tests

“CC” stands for conformance class (Rinne, 2014)

Following the formal rules and using available software capabilities, INSPIRE conformance for the treated theme is validated against the target application schema, data formats and structure within the HALE software. Conformance to semantic and spatial requirements which ATS specify are further tested either with the OGC Validation Tool or by manual reference to the ATS.

2.5. Western Balkans countries commons in context of spatial data governance

Introductory part of this thesis presented the positions of the Western Balkans (WB) countries. Here, their characteristics in relation to the INSPIRE Directive implementation is discussed as the background for considering a common interoperable spatial data model. Within the WB countries there are EU member states (Slovenia and Croatia), candidate countries (Albania, FYR Macedonia, Montenegro and Serbia) while others are potential candidates with a strategic orientation towards EU (Figure 2-8).

Figure 2-8 Western Balkans region countries

In relation to their EU membership status, obligations and activities in implementing INSPIRE differ from one WB country to the next. Member states are obliged to follow the INSPIRE roadmap and are have therefore reached a more advanced status of the Directive implementation. In FYR Macedonia, the INSPIRE Directive is already a part of the national legislation, while in other countries this process is started on a strategic level along with bringing the national development programmes into alignment with EU legislation as part of the EU integration process. The EU integration process is seen as the initiator of the activities in the domain of public data management in the WB region.

On the other hand, the process of political and economic transition of the region during the 1990s from socialistic to liberal market economy caused huge changes in domains like real property ownership (Cetl et al., 2013) from public to private owned property. This caused the involvement of large projects and international institutions' loans in order to foster the economic progress of the WB countries. In such a constellation, as the first and experienced beneficiary and the most advanced institutions in the data managing domain, the National Mapping and Cadastral Agencies (NMCA) emerged as the drivers of the national spatial data infrastructures (NSDI).

Support to the process of NSDI development primarily comes from the EU pre-accession funds through implementation of projects like INSPIRATION – Spatial Data Infrastructure in the Western Balkans (Aleksic, 2013). INSPIRATION promoted and worked on establishing a means of coordination of NSDI implementations in order to meet the INSPIRE Directive requirements in the WB region. NMCAs played the central role here and showed the importance of an integral approach. Knowing the limitations of the resources in the WB countries and common institutional history, standards, organisational, infrastructural and market position, regional cooperation become a key factor of synergetic development and dispersion of activities for the mutual benefit (Cetl, Tóth & Smits, 2014). Observing the key segments of the NSDI in the WB region, background for building the common model was set up. Within the INSPIRATION project, the analysis of INSPIRE domains is conveyed following the INSPIRE reporting requirements methodology. Namely, major aspects of NSDI: Legal issues and funding, Coordination and organisational issues, Spatial data sets, Metadata, Network services, Interoperability and standardisation, as well as Use and efficiency were analysed for the WB region. Table 2-2 presents the outcome of the analysis in the form of a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis table.

Table 2-2 SWOT analysis of the status of INSPIRE in the WB countries, adopted from (Cetl, Tóth & Smits, 2014)

Strengths	Weaknesses
Organisational framework Legal framework Existing reforms (e.g. LAS) NSDI strategies NMCAs capacity NMCAs cooperation in region (exchange of the experiences and lessons learnt)	Absence of funding models **Lack of data sharing** Licensing **Lack of interoperability** **Usage of standards** Metadata catalogues Cost/benefit analysis is missing
Opportunities	**Threats**
INSPIRATION project Stronger involvement in the different INSPIRE bodies (e.g. Maintenance and Implementation) Involvement of broader spatial data interesting community (private sector) Accession of HR to EU (reuse of existing best practices) Different funding opportunities (IPA, donors, etc.) Joint projects (cross-border cooperation)	Lack of capacity on national level (other NSDI stakeholders) Political changes (lack of political support) Lack of funding

The conclusion on the constellation of the INSPIRE Directive implementation in the WB and so far performed activities refer to the strategic and normative positioning of the Directive in the national approaches. Thus, the first important step, recognition and orientation towards adoption of the Directive has been done. Furthermore, there are current projects and active opportunities for funding and infrastructural strengthening of the WB NSDIs. These actualise the moment for establishing standardisation approach and establishing methodology in technical approach to data managing.

On the other hand, some of the main weaknesses consider the absence of standardisation, problems in data exchange and lack of interoperability. In the long term, this situation is threatened by the possibility for change of the political direction which may sifgnificantly influence underdeveloped and dependent NSDIs. This thesis addresses the described situation by providing a common approach in data structuring as the model for interoperable exchange, considering a standardised approach and wider data usage. The thesis focuses on one of the themes to justify the methodology and opens the possibilities for better NSDI development region-wide.

2.6. Overview of the experiences in data harmonisation process

Having defined the conceptual approach as well as the requirements that INSPIRE specifies, data providers across Europe have made significant progress in implementing the INSPIRE Directive. Constellation and issues in WB show which segments of INSPIRE are essential to be addressed first. This thesis already highlighted the key segments and discussed the model building concepts. This section gives examples of positive practice as the input for the practical realisation of the INSPIRE data model building and implementation concepts in Western Balkans.

As one of the Europe's most mature SDI, the Swedish environmental geodata and metadata compliance with INSPIRE specifications on the example of the European project Nature-SDIplus developed data is described in Hedefalk & Östman (2011). The goal was to examine all phases of the process of creating vendor independent and at the same time valid structure and syntax according to INSPIRE requirements. The paper presents a developed process bearing in mind costs of the process operations. As depicted in Figure 2-9, the existing data model for the source data was extracted in order to perform transformations (spatial and non-spatial) to specified INSPIRE formats, after which the resulting sets were validated and loaded into the target INSPIRE compliant database (Extract-Transform-Load or the ETL model). Data from the harmonised target model can be published and used further in an interoperable manner.

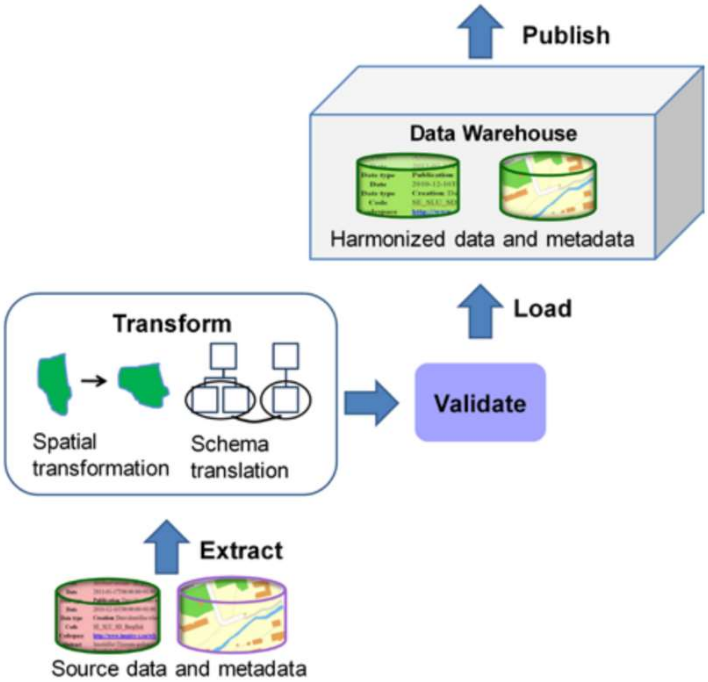


Figure 2-9 Overall harmonization approach (Hedefalk & Östman 2011, p. 32)

Core of the process that is of interest for this thesis is the transformation and data harmonisation. As emphasised in the paper, harmonisation considers resolving syntax, structure and semantics mismatch between compared data models. This process required comparison and identification of the elements from both data models. Further on, operations for the translation of one model to the other were defined, after which the actual transformation was executed.

As Hedefalk & Östman (2011) note, different phases of the harmonisation process require and have available automated solutions. The downside is the impossibility to have a fully automated process and consequently there is always a need for activities which demand manual work or decrease data quality. These are identified as the more costly processes. Thus, the divergence of the source data and the possibilities of the tools for automated processing (like coordinate systems transformation, merging, renaming, filtering etc.) determine the complexity of the transformation process. Hedefalk & Östman (2011) further discuss other phases of reaching interoperability, like validation of the schema and publishing various web-services in a standardised manner, which are out of scope for this thesis.

Experience gained in building up Swedish environmental data showed conceptual agreement on the INSPIRE guidelines for data model building as well as previously discussed theoretical concepts of model extension. What is important to mention is the practical implications that arise from the adopted process of data harmonisation. In that sense, extensive analysis of the source data, processes needed to transform the data and tools for performing the activities are the key segments for data harmonisation.

Another example of the spatial data harmonisation is the geological data in Portugal that was led by the need for more efficient and INSPIRE compliant data structure in digital geological maps production. The process described in Pereira et al. (2013) considered restructuring of the national model and transformation to extended INSPIRE theme geology model within the data producer's - *Laboratório Nacional de Energia e Geologia* (LNEG) - information system. The approach was as depicted in Figure 2-10.

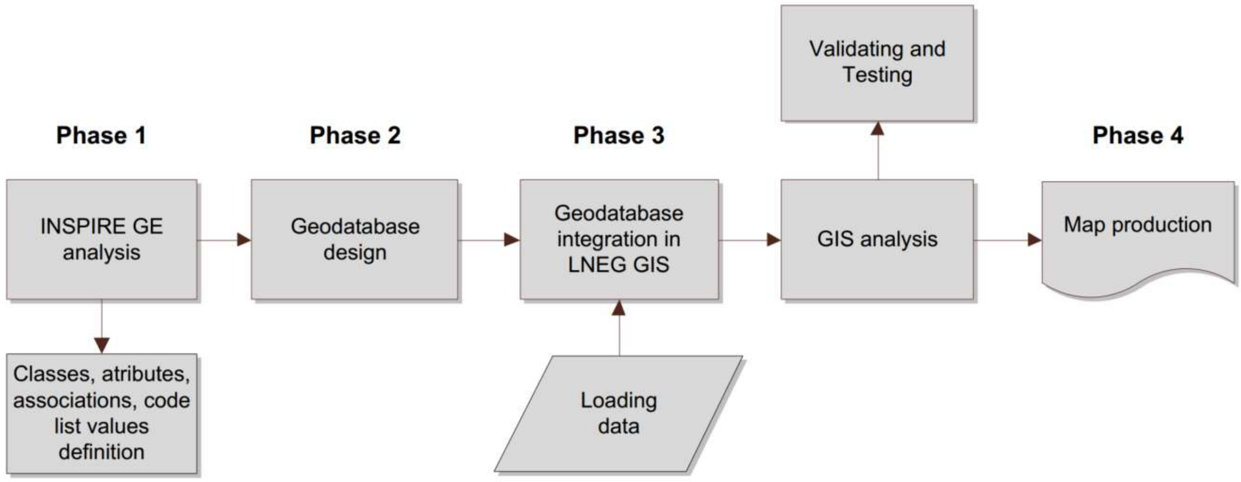


Figure 2-10 Methodological workflow (Pereira et al. 2013, p. 131)

In order to have full INSPIRE compliance and satisfy Portuguese geological survey requirements, the underlying data needed structural change (Figure 2-10, Phase 1). The INSPIRE model needed extension in terms of modifying existing object classes, excluding certain elements and adding new object classes, code lists and attribute values (Figure 2-10, Phase 2). LNEG approach focused also on the integration of the new model with the existing production process (Figure 2-10, Phase 3). At that stage, the system was ready for production phase, GIS analytics, validation and map creation (Figure 2-10, Phase 4).

For this example the key segment was Phase 2. The Geodatabase design considered use of the database design software tool Enterprise Architect software from Sparx Systems in which the INSPIRE theme geology core model (described earlier in section 2.2) application schema was loaded and edited. Selected main types of the schema were loaded, coded value domains were also added and the coordinate system, compliant with the INSPIRE requirements, was defined. On top of the imported classes, a new class was added as the representation of the important features for the national geological maps comprehension. Beside the INSPIRE model *ShearDisplacementStructure* feature class, a new class *Displacement* was added to provide further information on the faults as the types belonging to the *ShearDisplacementStructure*. Several other class associations, attribute values and domain types were created. All modifications were altered using GeoSciML rules in accordance with the INSPIRE recommendations. Extended model was thus ready for data deployment from the existing model. Before the mapping between models took place, modifications in the source model were made in order to enable the transformation process in a more automated manner. Namely, a single table containing data about the geological map was split into two classes with type associations between them. Semantic harmonisation further required identification of the corresponding terms between Portuguese and English.

Following the described principles of INSPIRE model extension, experience from Portugal showed how specific needs for retaining essence of the national model, conformance with the existing GIS and taking most of the automation process were addressed.

Other similar experiences reveal projects for reaching data compliance with INSPIRE (like in GEORZ Lab & Research department, 2011) which considered creating fully compliant data through transformation of the existing data and storing it as XML in order to provide seamless and fast data publishing and serving, without intermediate processes between source data and targeted data model.

In another example Hemmatnia et al. (2010) in the ESDIN project, a transitional database was used to enable automated transformation from the source model to a target model, keeping source model as the operational and unchanged, but satisfying the target model requirements too. The cost of such an approach is the time and additional processing by the described structure every time data is requested by the service.

Practise of the mature SDIs throughout Europe conceptually take the same approach, following the INSPIRE GCM principles and extending the model upon the underlying standards. In case of the geology theme, GeoSciML is the language for formalising the spatial data. On the more local level, within the projects' scope or national domain of interest, equally important model requirements are the needs of the data users and providers. Practice confirms the described theoretical approach.

The shortcomings in such an approach is the reverse proportionality between the institutional level of spatial data development and the efforts needed for data harmonisation. Thus, in case of the environments where capacities, data structure and resources are limited, the importance of developing a harmonisation approach becomes even more important. With the analysis and presentation of the state of the art approach in the INSPIRE implementation domain, for the theme geology, along with the regional specifics in the previous sections, the core elements of the harmonisation process are defined and the framework for geological data exchange is set. Furthermore, concepts of extension of the model are presented, giving the methodology and means for INSPIRE harmonisation example practical implementation. On this basis, the defined concepts and approach represent the background for the exploration of the key factors for the regional model formalisation within Study 1. The subsequent Study 2 builds upon the first with a practical example of INSPIRE data harmonisation implementation.

3. Study 1 - Analysis of critical factors of the regional INSPIRE harmonisation model

This chapter describes the questionnaire that addressed the regional SDI stakeholders responses with the aim to identify their maturity, needs and specifics regarding the implementation of the INSPIRE harmonisation model. Domain of research is narrowed down to INSPIRE theme Geology. The questionnaire examined the stakeholder data structuring approach and expectations, level of INSPIRE Directive implementation and experience with data harmonisation in order to form the basis for the development of a regional model in accordance with the INSPIRE requirements.

3.1. Methods

The questionnaire covers previously identified critical factors of the harmonisation process, namely spatial data, legislature in spatial data management and institutional capacities. Spatial data refers to supplying the required information, in required format, in up to date state, conforming to international standards, as recommended by the INSPIRE implementing rules, implying certified data quality status. Legislature is treated in terms of regulating data availability and accessibility in the state of the art manner. Institutional capacities are covered with questions on the capabilities for supply and use of spatial data through web services, in standardised structure and in an interoperable manner.

The study was conducted by means of a questionnaire based on several examples of surveys from other INSPIRE implementations: the situation and perspectives of the GI market in Europe (Humboldt Consortium, 2011), public consultation on the implementation of the INSPIRE Directive in Serbia (Aleksic, 2013), exploration of the technical prerequisites for serving geological models (The Geological Surveys of Europe, 2014), Mid-term evaluation report on INSPIRE implementation (European Commission & European Environmental Agency, 2014) and surveying various INSPIRE models implementation and extending (wetransform, 2016). Example surveys focus on various regions, spatial themes and implementation scopes. Thus, relevant questionnaire' segments and questions were selected, modified and expanded to fit this thesis specifics.

3.1.1. Overview and questionnaire

The questionnaire was structured to examine the identified key areas for INSPIRE implementation within geological authority organisations in the Western Balkans region. Thus, it covered data quality, legislation and institutional capacity as areas of interest divided into four thematic segments: INSPIRE harmonisation needs, harmonised data status, INSPIRE harmonisation implementation and optional part on extension of the INSPIRE harmonisation models.

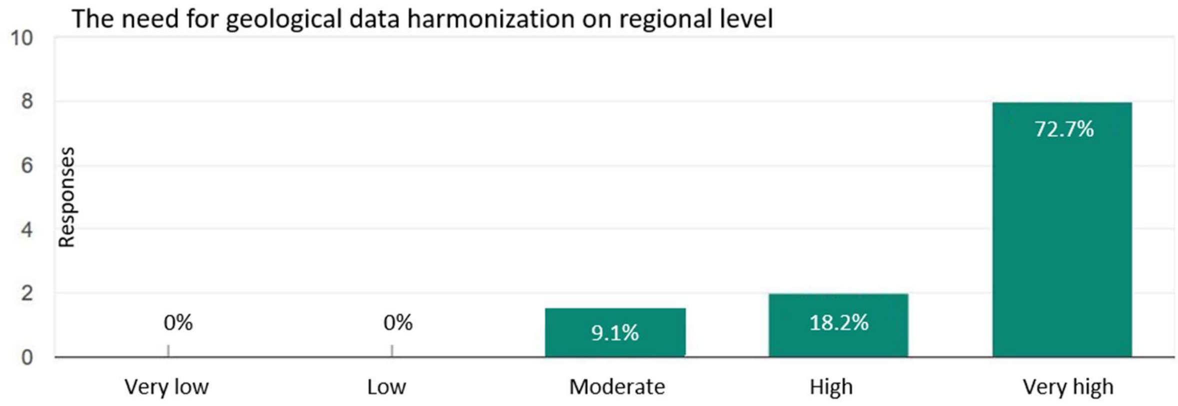
The first part of the questionnaire referred to the actual needs for harmonised spatial data according to the INSPIRE Directive. It reflected current and planned usage of the spatial data within geology as the thematic scope. Further, it referred the capacities for meeting the needs and envisioned benefits the harmonisation process ought to bring. The first part of the questionnaire focused on the position of the existing regulations that cover geological data usage. The second part considered the level of harmonisation of the spatial data within the organisation. It reflected the performed activities, legislation adopted and consumption of the INSPIRE conformant data. Next part of the questionnaire analysed the organisations' approach to the process of spatial data harmonisation according to INSPIRE. Main focus was the institutional capacities and, in that relation, the status of the data and regulations the organisation faces in terms of work on the actual harmonisation. The last part focused on the organisations that already have experience in INSPIRE data harmonisation implementation and modification. Thus, it was optional and intended to provide the answers on the existing activities already taken in the direction of extending the INSPIRE model. At the end, questionnaire provided space for information on the questionnaire participantsand their affiliation, expertise and contact information.

The technical realisation and dissemination of the questionnaire was done using Google Forms (Appendix C – Questionnaire) and delivered to the Western Balkans SDI stakeholder institutions' representatives on GIS management positions through direct contact (Appendix D – Institutions contacted for participation in the questionnaire). Review of INSPIRE implementation approaches in WB countries was based on the answered questionnaires in period between May 20th to June 7th 2017. The questionnaire was answered by respondents from nine stakeholders, each from one of eight Western Balkans countries, while in Bosnia and Herzegovina there was one representative for each of the two federal parts of the country. Stakeholders either belong to the official national geological authority institutions (55%) or referential academic institutions (45%). All of the respondents are directly engaged in GIS development and data harmonisation as professional geologists with backgrounds in engineering geology, hydrogeology and geo-hazard management. A majority of them are at least in project management positions, while their involvement in the geological data management refers to spatial analysis, processing, harmonisation, utilisation of referential tools and services. This structure of the respondents' profile corresponds to the targeted group for the questionnaire and provides relevant subjects for gathering information on the harmonisation in the WB region. The results were reviewed and discussed in light of general outcomes, three critical factors of the INSPIRE implementation in the region (spatial data, legislation and capacities) and finally from the actual model extending experiences.

3.2. Results of the questionnaire

As the first inference, it can be taken from the questionnaire results that harmonisation of the geological data in the WB region is seen as a highly important necessity. Namely, six of thennine respondents evaluated harmonisation as the highest ranked need, while the remaining three stated that the need is either high or moderate. The background of this need lies in the responses given on the demand for interoperable spatial data in numerous geological domains, namely base geological maps, managing mineral resources, water resources, boreholes, structural geological maps, landslides and cross border cooperation. Another argument for the harmonisation need is the respondents' orientation towards reaching and managing interoperability as a continuous effort of a dedicated department within the geological authorities in over 50% of the answers, while remaining answers were dispersed in seeing interoperability management as the necessity that should be addressed as a periodical cost that could be covered by the external resources or internal campaign work (Figure 3-1).

Three of the nine respondents agree that the implementation of INSPIRE is well coordinated region-wide, while the majority of the remaining six respondents disagree. The coordination of the INSPIRE implementation on the regional level is evaluated as a positive trend among the EU member states of WB countries and respondents from universities, while geological authorities and EU candidate countries are much more reserved or unsatisfied with this coordination.



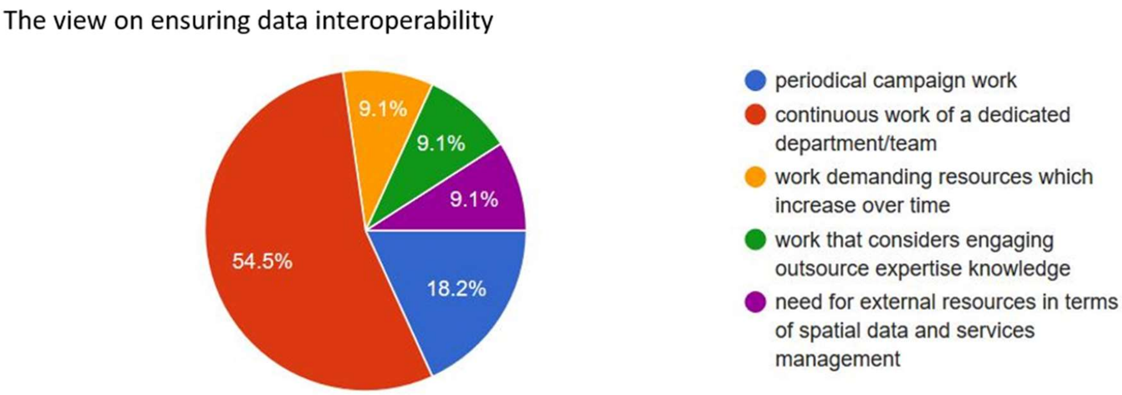


Figure 3-1 Stakeholders' need for harmonisation

Spatial data status

Being one of the key factors, spatial data is seen equally by four of the respondents as highly important (while remaining five agree that it is just important) in terms of decreasing redundancy, technical standards usage, data accessibility, management and efficiency growth. The most recognised value of the harmonised geological data for the geological authorities is the added value for the private business sector. This implies focus of the geological authorities on market orientation and user needs as a sign of positive trend towards INSPIRE recommended wider usage, better accessibility and availability of data.

At the same time, application of the INSPIRE harmonisation is seen as an extensive and time-consuming data processing and restructuring process. In addition, the current data specification quality is emphasised as one of the important drawbacks of the data harmonisation process.

Indicator of the INSPIRE implementation level is the status of the different INSPIRE specifications among the respondents' institutions. Hence, according to the INSPIRE implementation roadmap (European Commission, 2015), higher priority milestones (metadata, data structure, discovery and view services) are in 50% of the answers seen as at least partially implemented, while more advanced requirements (like download services) are met only in a couple of cases (Figure 3-2). Accessing spatial data from third parties is in 80% of the cases either hardly accessible or covered through traditional exchange channels (physical digital mediums or spatial files exchange via internet) while only a couple of examples exists where geoportal or web services are used for obtaining third party data (Slovenia and Croatia).

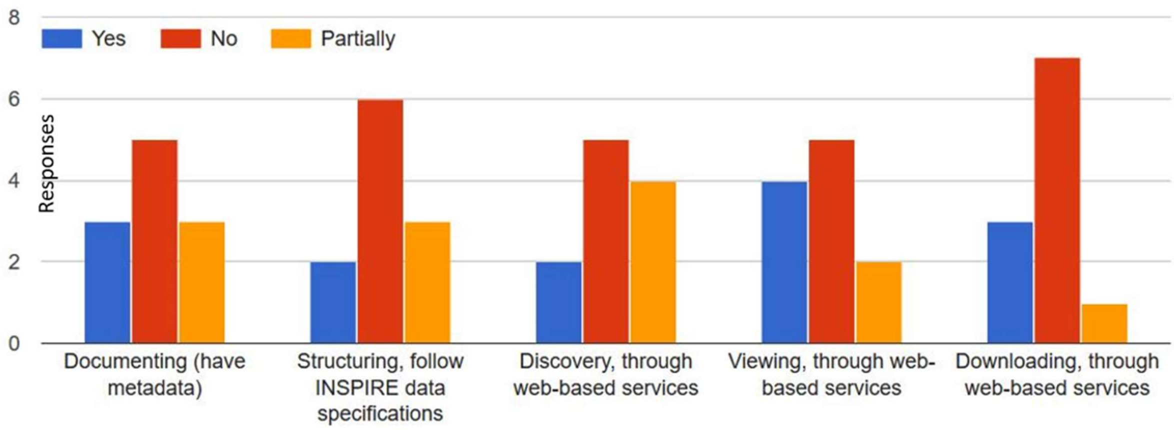


Figure 3-2 Domains and level of following INSPIRE specifications in stakeholder institutions

So far experiences in relation to the difficulties in managing spatial data show the biggest challenges in the inconsistencies of the metadata and data itself, followed by the data models and the obstacles in integrated use of different datasets, their conversion and semantic synchronisation. The background of such a situation is identified to be the data models' inconsistencies along with the spatial integrity issues of the data. This directly formulates the need of the regional institutions for data models' seamless usage and integration, as one of the INSPIRE specifications main goals.

Legislative status

Seven of the nine respondents agree that one of the biggest drawbacks INSPIRE implementation brings is the asynchronization of the current geological legislature with the INSPIRE requirements. The emphasis is on the data management, data accessibility restrictions and data and service sharing licencing models.

Identified documents which treat the structure and methodology of producing basic geological maps were produced long before INSPIRE introduction. Furthermore, these are not formalised for the digital usage either and are often outdated regarding the INSPIRE conceptual demands for describing data structure and management. Thus, the relation of the current legislation with the INSPIRE concepts is in large majority of cases (over 80%) very low or even non-existent (Figure 3-3).

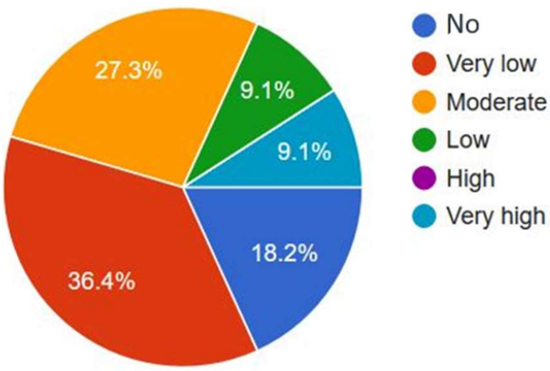


Figure 3-3 INSPIRE legal coverage in Western Balkans countries

Capacities status

In terms of capacities of the regional institutions managing geological data, data harmonisation means also the need for additional resources and skills development. Furthermore, coordination and support on national levels as an important segment of increasing institutional capacities is seen as the most underdeveloped segment.

One of the more indicative signs of the need for further capacity building is the diversity among the answers on the question whether INSPIRE implementation would cause data redundancy as a problem in its management. This issue can be explained with the legislation on current data managing which does not correspond to INSPIRE recommendations. In that direction, the need for knowledge transfer for the approach in long term spatial data management is seen as a valuable contribution to this situation.

Four of the nine institutions that took part in the survey actually participated in projects implementing INSPIRE recommendations. Those who did, took part in EU projects that treat this domain like, Minerals4EU, eENVplus, EGDI etc. This implies the need for experience in practical implementation and usage of the harmonised data and services, as well expert level comprehension on spatial data structuring and management.

The initiated processes of adopting the INSPIRE Directive in the WB region also caused three of nine institutions to at least plan implementation with their own capacities, while the other six have not actually decided on this question (Figure 3-4). Such a constellation is shown in more detail when observed from the aspect of software tools used for the spatial data harmonisation. Firstly, these tools are present in less than half of the institutions, and secondly the dispersity among the existing ones is very high (practically each institution relies on different software).

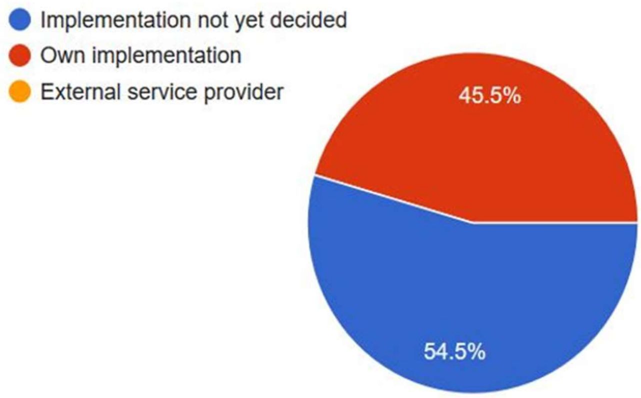


Figure 3-4 Stakeholders' approach in implementing harmonisation

From the stakeholders' point of view INSPIRE harmonisation process adoption and implementation quality is in the first place seen as a matter of capacity building, then as budget issues and finally as technical standards implementation.

Experiences in re-formalising the INSPIRE model

The survey showed that almost half of the respondents are by INSPIRE standards expert or professional users, while one third is considered as an occasional user. In addition, about two thirds of respondents were involved in some way with work on INSPIRE data specifications, although less than 20% was engaged in creating data models. Further insight into work with the INSPIRE data models was provided knowing that 50% were engaged in the creation of the INSPIRE compliant data model either for research purposes, internal use or as a legal obligation. At the same time, over two thirds of respondents at least plan on developing models referring to INSPIRE data specifications (Figure 3-5). Four data models from different countries were identified within the questionnaire responses and their characteristics were analysed.

Examining the approach to the model extension it is notable that most often modifications refer to new properties for existing data types. In addition, usually *Codelists* are expanded or new ones are created. Less often the case is to create new data types or to take away some elements. The creation and maintenance of the data model is applied using various software tools (like HALE and ESRI ArcGIS). All of the created models cover at least the national level and spread to regional coverage of main geological domains (lithology, hydrology, stratigraphy, land cover).

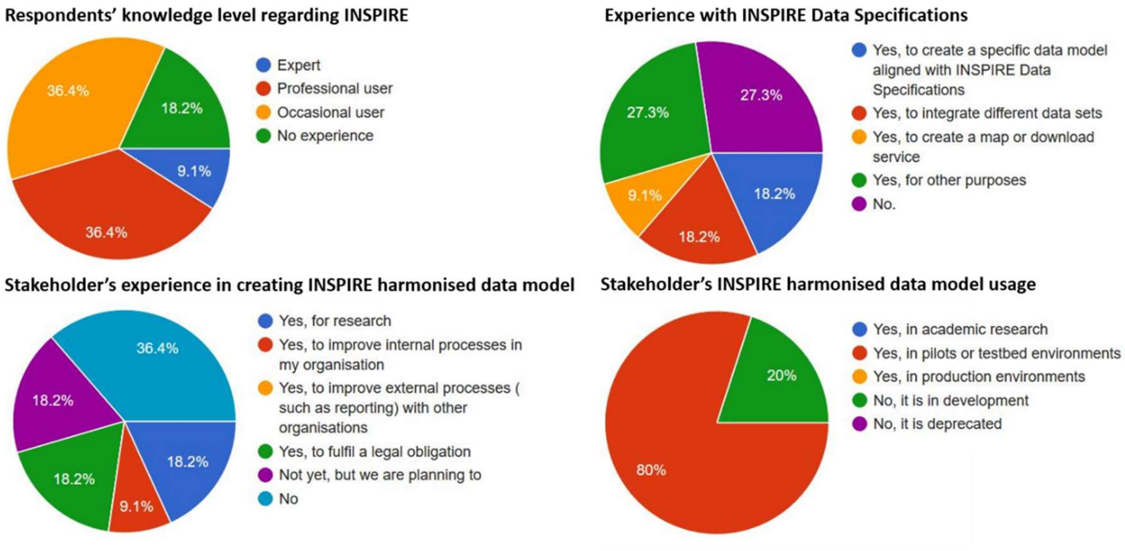


Figure 3-5 Stakeholders' experience in implementing INSPIRE Directive

On the other hand, the access to and availability of the data is not that uniform. Four out of nine institutions use web services to serve data to some extent either directly or through their portals, while others still do not have regulated harmonised data delivery process in an interoperable manner. This reflects the situation in which three of the four answers showed that extended data model and data harmonised in such a way is used as the pilot or testbed environment while the one model is still being developed. Furthermore, data model licencing is on a very low level having only one of four covered by licence.

Insight into the developed models structure was possible only on a conceptual level for the two models, while others were not publicly available. Available models are the results of projects BEWARE (Ministry of Mining and Energy of the Republic of Serbia, 2015) and eENVplus (GISIG, 2013) pilot application. While the data model from the BEWARE project gives only an overview of the model creation and usage, the eENVplus developed data model gives a much deeper view of its structure. For further reference, readers can refer to technical documents and training materials available from (eENVplus Consortium, 2014).

Regarding the issues in model design, respondents in most cases had only minor problems with the harmonisation process underlying standards, extension methodology, modelling languages knowledge and were more concerned about physical resources (people, funding) and support. Yet, implementation of the data model showed that issues arise in the phase of practical model and data transformation. On a more general level, the impression is that the lack of human resources is evident and data usage is something that more effort needs to be put in.

Implemented local models show that INSPIRE schemas do cover the core of the local models too, while customisation is needed in the sense of classification of local data types and their attributes. Bearing in mind the scientific, testing and in-development nature of the existing models, the question raised is about the suitability of the regulatory acts that treat the field of geology in the region and their requirements. Current legislation is not flexible towards the need for geological data to be widely available. Thus, available models can be seen as milestones in the development of more comprehensive and advanced regional models.

3.2.1. Western Balkans advance in INSPIRE Theme Geology

Spatial data harmonisation is seen region-wide not just as an obligation but a necessity and an important economy and market driver, although present, stronger engagement in regional INSPIRE implementation coordination is needed, especially in non-EU countries. Initial phases of the INSPIRE harmonisation are being implemented, having obstacles in available resources, data restructuring and transformation, as well as with download services implementation. The drawback that lies in the background of such a constellation is the legislature in the geology field which is outdated and limits the desired interoperable approach. Furthermore, experts working in the field can hardly cover the sheer number of involved stakeholders and market needs. This is shown in the implemented INSPIRE models extended for the local needs. Namely, they fully follow INSPIRE and in addition introduce basic customisation concepts like classification of described objects domain values or language issues.

It can be said that there is regional consensus on the need for a model that could be used and implemented as an example to motivate and improve adoption of the harmonisation process for better regional cooperation, influence legislative evolution, contribute to capacities development and further regional INSPIRE extended model development.

This paper shapes its *Study 2 - Implementation of INSPIRE data harmonisation* using insights and conclusions obtained from the questionnaire in Study 1, giving contribution to the identified key aspects of the regional INSPIRE model implementation process. It goes into more detail covering the approach to the INSPIRE model implementation, its elements, development, modifications and presents a use case to validate the approach.

4. Study 2 - Implementation of INSPIRE data harmonisation

This chapter shows the implementation of the discussed theoretical concepts of INSPIRE harmonisation process, respecting the best practice in INSPIRE model formalising principles and the results of the regional needs structured within the questionnaire from Study 1. Figure 4-1 shows the flow of the geological dataset harmonisation process.

Firstly, the source of the geological dataset model, Basic geological map of the Geological Survey of the Republic of Macedonia (GSRM), is described in order to depict its formalising as the logical data model. Technical framework and the conclusions from Study 1 were considered for structuring the logical data model and further on in formalising the INSPIRE GE target data model. Re-formalised target data model was structured in relation to the requirements by INSPIRE GCM and the source data model characteristics, bearing in mind lossless transformation and conformance with INSPIRE standards. Processes of model structuring and transformation are represented in detail in the Appendix E – Conceptual data model of the basic geological map, Appendix F – Formalising the logical data model for the Geological map of FYR Macedonia in scale 1:500,000 and Appendix G – Source to target model mapping and transformation using HALE. The results of the Study 2 are a harmonised geological dataset and a review of findings from implementing the INSPIRE Data Specifications locally.

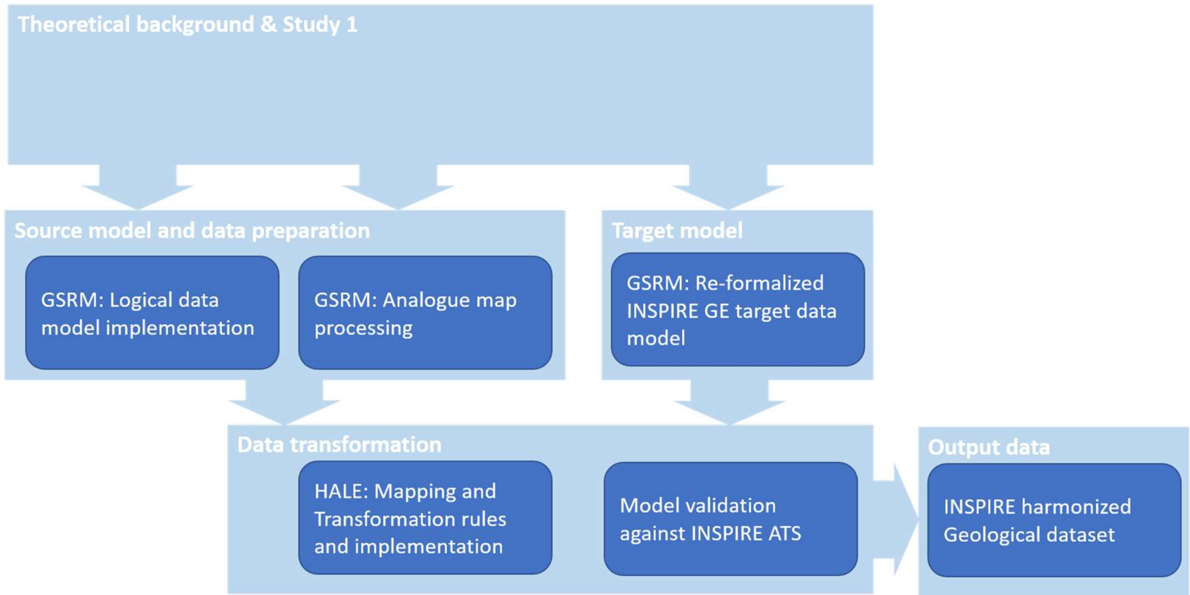


Figure 4-1 Geological Dataset Harmonising Process

4.1. Methods

The workflow of Study 2 as represented in the Figure 4-1 firstly analyses the model of the GSRM geological dataset in scale 1:500,000 using the Guiding technical rules of making the basic geological map (Ministerstvo za Ekonomija, 2007). It is described as a conceptual model of the geological map. The analysis led to structuring the dataset for thelogical data model and the approach to map digitalisation and data pre-processing. Further on, the logical data model and the digitalisation concepts were implemented as the logical data model and its spatial representation, which are referred to as the source model. As the target model in its core represents INSPIRE Data Specification application schema, it is formalised in light of the local needs and the limitations that the source model brings, bearing in mind the INSPIRE requirements. These were the key inputs for the harmonisation process implemented using the HALE software for design, mapping, transformation and validation of harmonised spatial data. The resulting harmonised dataset was validated and tested within the GIS client software, followed by the discussion on the gap between the source and the target model.

4.1.1. Source model and data preparation

The results of Study 1 imply the need for the interoperability of the managed geological data as well as the technological readiness for the process. At the same time, the low capacity and the regulatory constraints form the framework in which the process should be performed. This chapter introduces the example of a national institution and the characteristics of the dataset harmonised in the INSPIRE compliant manner.

The Geological Survey of the Republic of Macedonia is the institution put in charge of managing geological data on the national level. The GSRM is also an active stakeholder in the national SDI and delegates one NSDI Council member. Its orientation towards GIS implementation and usage of geological data brought the GSRM in focus of this thesis as the subject for the study. Geological data management considers applying regulations in the fields of mineral resources management, digital maps production and publishing, supervising geological surveying works and taking part in national policy making regarding spatial planning, environmental protection policies, agriculture land utilisation etc. In this sense, it is important for the GSRM to have structured, available and accessible data in order to keep efficiency and effectiveness at the necessary level.

Basic geological map of the FYR Macedonia is regulated by the Guiding technical rules of making the basic geological map (Ministerstvo za Ekonomija, 2007). The same guideline is a standardised document that had regulated creating maps in the whole WB region since the 1960's. Thus, it represents the basis for geological maps production region-wide. The Guideline treats the structure of the basic geological map and represents the basis for producing geological maps in the scale 1:50,000, which were generalised for the creation of the maps in scale 1: 100,000, 1: 200,000 and 1: 500,000. Figure 4-2 shows the coverage of the produced maps in scale 1:100,000.

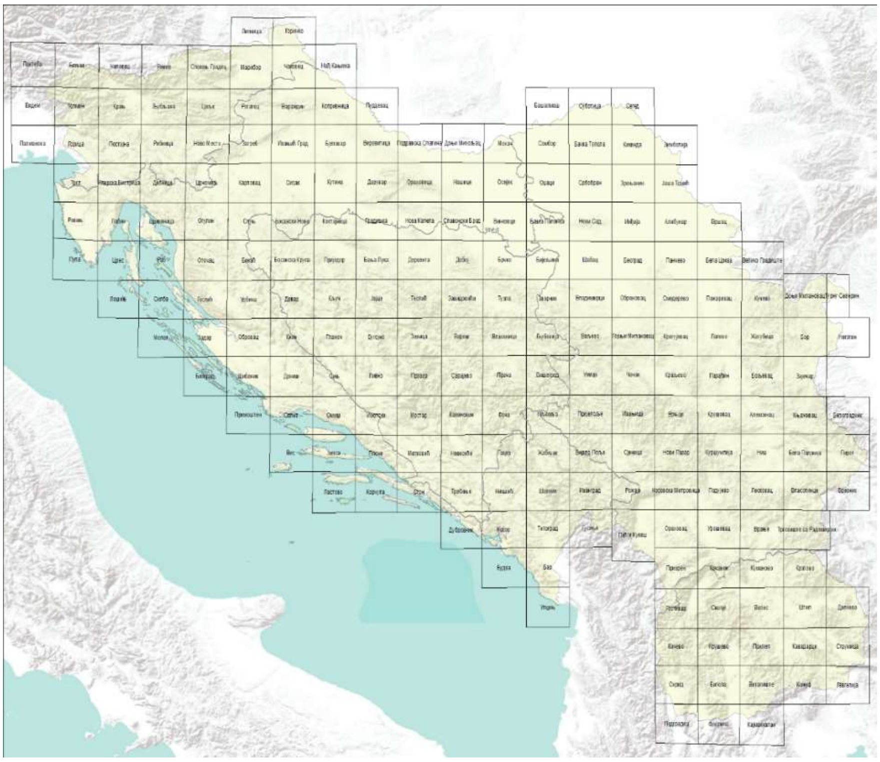


Figure 4-2 Sheets coverage of the basic geological map in scale 1:100,000 of the former Yugoslavia (Geologicharka, 2012)

The guideline describes the process of acquiring data, map structure and formal rules for producing the map. Although it was produced in the mid-20th century the guideline or its revised versions are still the basis of national regulations for geological maps. For the purposes of this thesis, the guideline is used as the formal description of the dataset structure since in FYR Macedonia it is the only formal legal act covering map production. It is described in *Appendix E – Conceptual data model of the basic geological* map.

Conceptual data model for the basic geological dataset represents a complex data structure, describing all surveyed geological information, as well as their interpretation. Although this approach results in rich maps, their usage is firstly limited by the need for expert interpretation of the dataset elements. For example, rules for mapping bedrock age are dependent on the bedrock type and can be represented in one case with the polygon colour and in another with the symbol. This makes wider usage of the maps, their flexibility in making thematic maps and digital representation complex. Furthermore, a universal approach in acquiring data made it too general for the thematic maps and specific needs in some regions (hydrographical objects or mineral resources) and limited data synchronisation and exchange with neighbouring countries. Being prepared for production of analogue maps, the guideline lacks a standardised approach in formalising elements like relationships, domain values, representations. These specifics make the basic geological map structure an informative and thorough source, but imply the need for further analysis and geological interpretation in order to have all elements of the logical data model further defined for its digital representation.

In several countries of the WB region there was certain progress in this direction, like creating a geological vocabulary along with terms relations (Rudarsko-geološki fakultet, 2011) or decomposing maps in widely used forms, like formation maps or age maps in (GeoZS, 2017). Yet, these examples do not represent a unified approach.

Scope of this study and the available resources influenced the choice of the source dataset to be the basic geological map in scale 1:500,000. Its structure and dataset elements were defined in consultations with the GSRM. In this way, the structure of the basic geological dataset that is uniform on a regional level could be followed and, moreover, it could be included in the practical example of the data harmonisation process.

Pre-processing the basic geological map in scale 1:500,000

The original form of the source dataset is the paper map - Geological map of FYR Macedonia and southeast part of Serbia in scale 1:500,000. Its pre-processing considered scanning, geo-referencing to FYR Macedonian national coordinate system and digitising using ESRI ArcGIS Desktop 10.2 software. The result is presented in Figure 4-3. Datasets were provided in the form of .shp files, ESRI spatial format for describing and representing data, keeping geometric, attribute and coordinate referential system information.

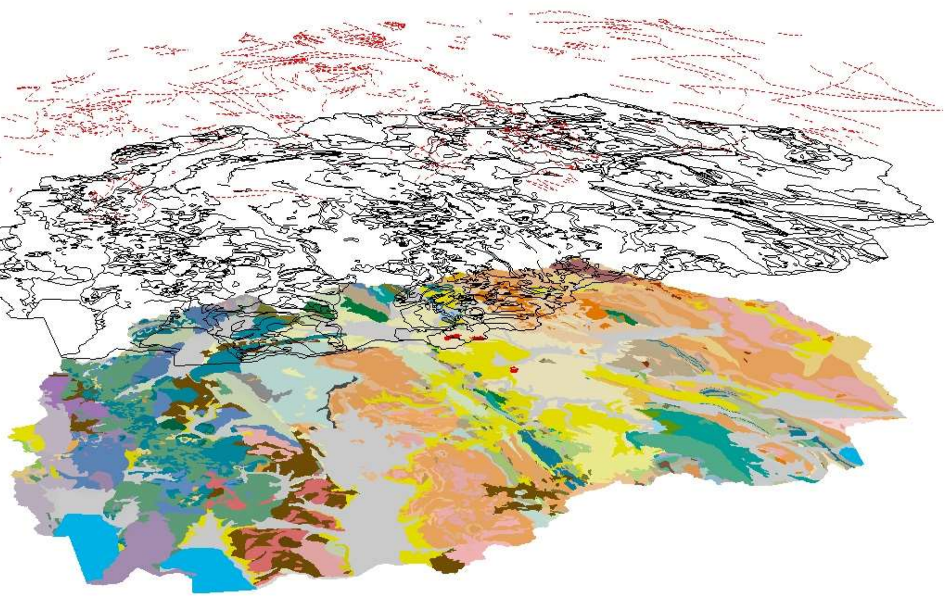


Figure 4-3 Layers of the Geological dataset of FYR Macedonia digitised and visualised in ESRI ArcGIS for Desktop (Jovičić, 2016)

Further processing consisted of modelling the data structure according to the map contents, the Guiding technical rules of making the basic geological map (Ministerstvo za Ekonomija, 2007) and interpretation with the geologists from the GSRM. The process is described in the Appendix F – Formalising the logical data model for the Geological map of FYR Macedonia in scale 1:500,000.

The most time- and resources consuming activity was caused by the unstandardized description of map elements, which were thus hardly comparable to other classifications like INSPIRE. This problem was resolved by assessing geological maps of different scale and themes, further interpretation of the elements in consultation with geologists and generalisation of elements' values. In that manner, the structure of map elements as well as requirements of the national guideline were retained and methodology of the harmonisation could be implemented as planned. It is important to note that the results obtained were thorough enough for this study, however, for the complete national model, further work in defining geological elements would be needed.

4.1.2. Target model and data transformation

With the formalisation of the source data model, the previous section provided input for the transformation process described on a conceptual level in the *Appendix B - Theme Geology Conceptual data model* and section *2.3.2 Data transformation solutions project - Extending INSPIRE Data Specifications*. Since Study 1 showed strong needs but very basic progress level in implementation, the target model provided within the INSPIRE Data Specification on Geology is used as the frame for introducing a regional approach to data harmonisation and interoperability. INSPIRE Data Specification on Geology is re-formalised to fit the purpose of the local model and to fulfil the INSPIRE requirements. Described in detail in *Appendix G – Source to target model mapping and transformation using HALE*, this study defines the approach and implementation flow to harmonisation.

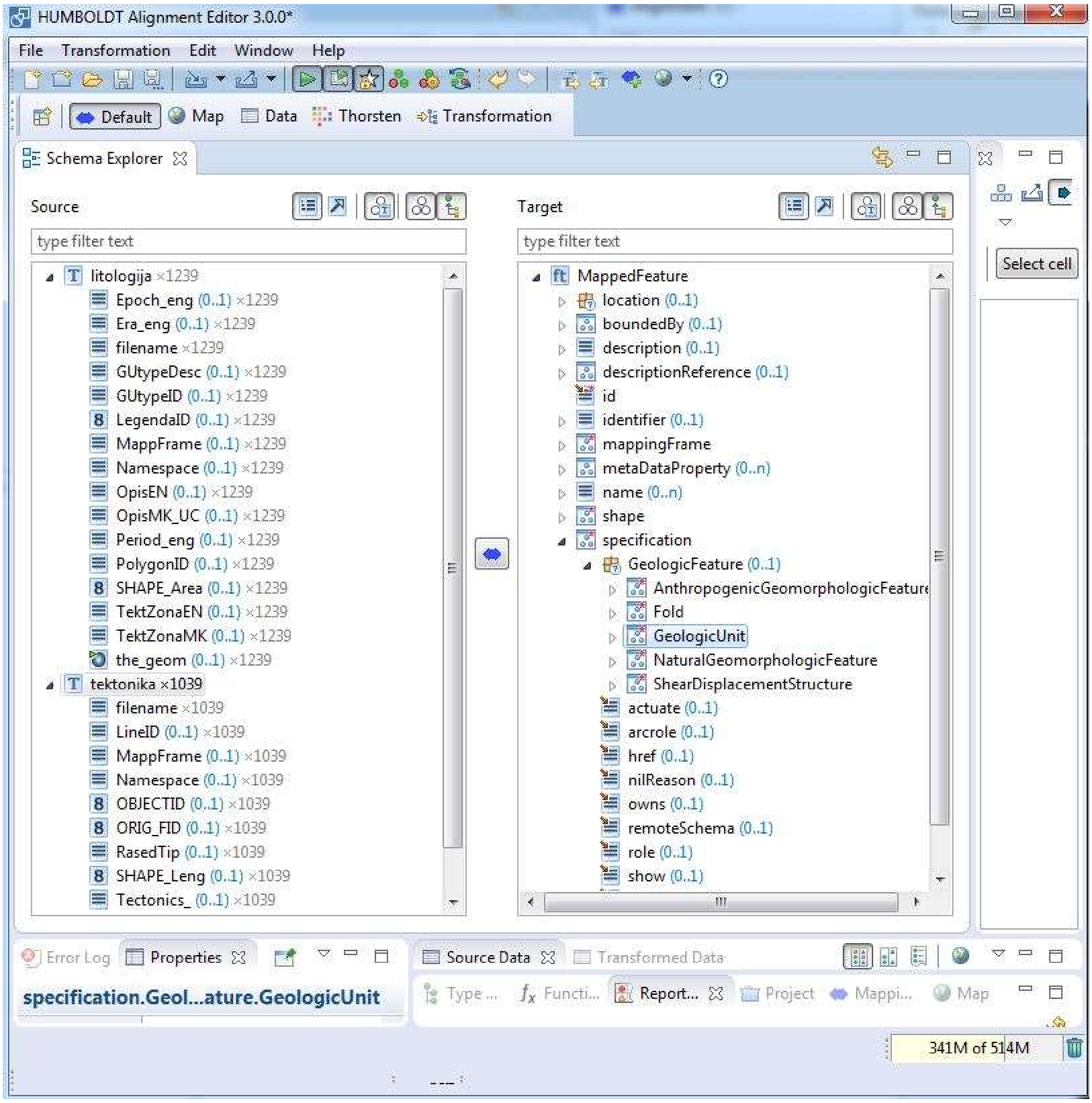


Figure 4-4 Source and target data model schemas described in Appendix G

In comparison to the target model scheme shown in Figure 4-4, the source model contains two feature classes along with the referent geographic and visual representations of dataset elements, while the target model contains six feature classes only for the core data model. Moreover, the target model feature classes contain fields describing geometry, identifiers and other generic elements, which need to be derived from the source model, but to keep quality and semantics of the source model. These were the initial requirements for the mapping and transformation.

Discussed in section 2.4.3, the software tool HALE is considered as an adequate utility for data harmonisation process implementation, namely model schema description and obtaining the required data structure through the definition of mapping rules, data transformation and validation of the performed process. On the logical level, the mapping process is considered as the definition of rules for different types of managing data characteristics (geometric, spatial, attribute etc.) in order to meet the target schema element requirements. Within HALE, data preparation gave the source model more flexibility and potential for the automated process of transformation, without devaluing source data or the existing data management process. In addition to core requirement of data structure conformance, the related IR refer to description, access, delivery and visualisation of the spatial data sets. HALE supports managing these segments too. The mapping and transformation process is described in detail in Appendix G – Source to target model mapping and transformation using HALE.

In addition to the core requirement of data structure conformance, related IR refer to description, access, delivery and visualisation of the spatial data sets. HALE supports managing these segments too. Here, only the visualisation will be briefly discussed for the purpose of data harmonisation process validation. During the mapping and transformation process, HALE gives overview of source and target models data representation. Figure 4-5 depicts this relation.

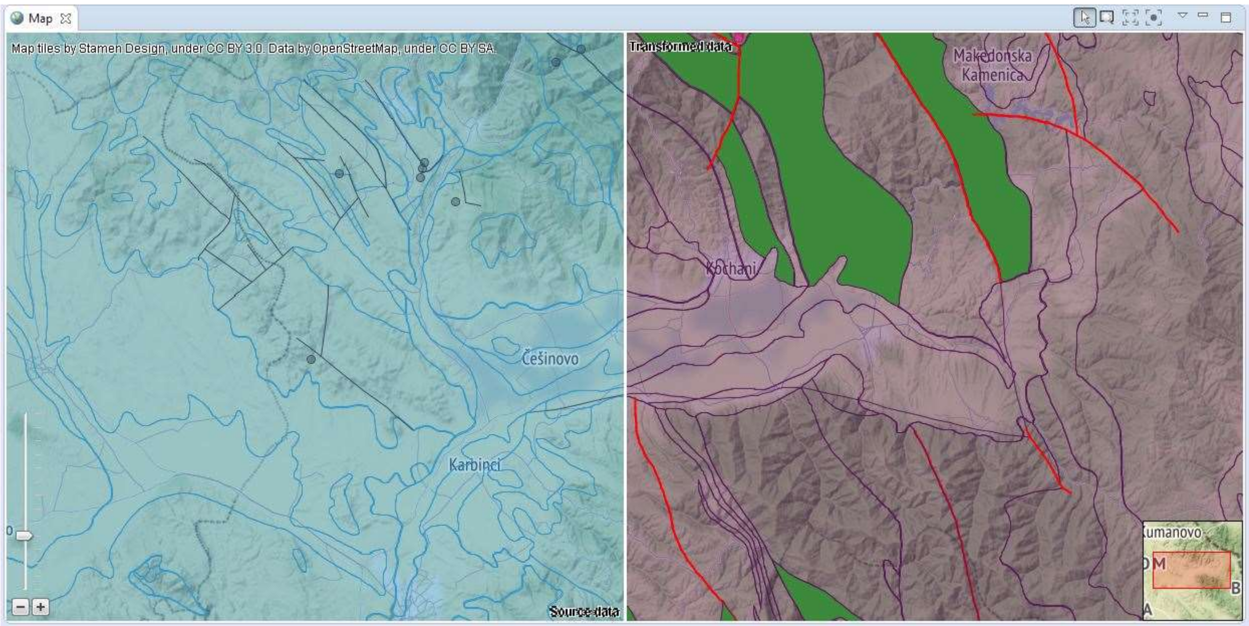


Figure 4-5 Visualisation of the source data and on-the-fly representation of the transformed data

HALE supports XML Styled Layer Description as standardised way of representing spatial data. SLD can be defined for the simple examples directly in HALE, or for the purpose of more complex visualisation, it can be imported. Regarding INSPIRE geology theme, within the data specification one segment treats representation of the harmonised dataset elements. In the image above using slider in the middle visual check of the source and target data can be made. Left side represents source dataset with automatically assigned symbology laid over *Open Street Map* as basemap. On the right, transformed data is shown, where according to INSPIRE tectonic units are represented as red lines, while other elements should have been coloured according to lithological unit they represent. Due to limitation of HALE in processing imported SLD for the complex datasets, symbology could not be completely verified within HALE. It was tested afterwards within other GIS software where transformed INSPIRE compliant GML data was imported, as shown in the following section.

4.2. Results

The mapping process and the transformation from the source to the target schema produce a dataset harmonised with the requirements of the target model. Furthermore, these processes are a formalisation of the automation of the data transformation. In this manner, the source model schema, as a representation of the data producer's regulations in data processing, becomes structured for much broader use. The study in this thesis presented the example of harmonising a geological dataset according to INSPIRE requirements. The used software showed the availability and usability of tools for data processing.

Formal dataset validation is conducted exclusively with the INSPIRE IR Abstract Test Suite (ATS) for the referential theme. ATS as described in section 2.4.4 treats the target dataset by testing its conformance to INSPIRE specification. It treats the application schema, reference systems, data consistency, metadata, information accessibility, data delivery, data portrayal and technical guidelines. In order to have an INSPIRE conformant dataset, it needs to pass all of the tests. Since metadata, data access and delivery are out of scope of this thesis, ATS for the application schema and coordinate reference system were tested. HALE software offers validation of application schema ATS while coordinate reference system is validated manually using GIS client software. HALE tools for validation refer to on the fly test, as well as target model conformity test upon the transformation.

During the mapping and transformation process, HALE enables validation of the defined processes on the fly using the Report List, where any error or warning can be seen and further addressed in order to have the resulting schema as required by chosen specification. Furthermore, data consistency is being checked in the background in accordance with the mapping rules. Apart from the online check which gives valuable insight, Instance Validation as part of the Reporting tool gives detail insight in errors occurred during transformation and testing conformance with the specified target schema. Image below represents example of lack of reference of source data to the target model in class *MappedFeature*. It was resolved with mapping of all mandatory fields from source to target schema. Second warning shows multiple occurrence of values which should contain unique identifier. Solution to this problem was recreating of the unique identifier for the whole dataset.

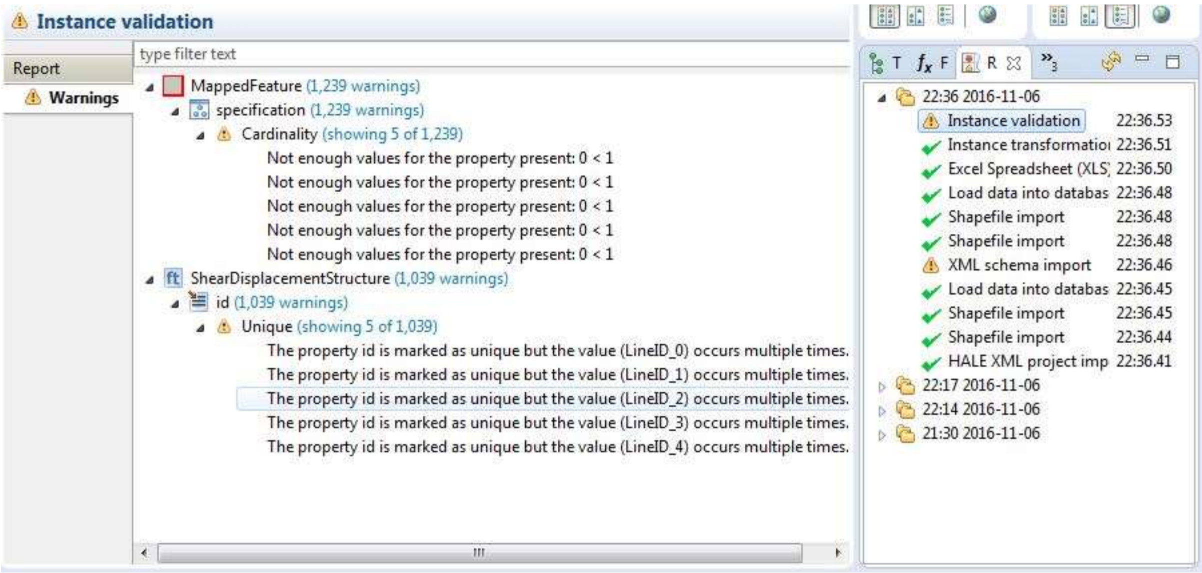


Figure 4-6 Instance validation error example in HALE

Upon specifying all mapping rules and performing data transformation, validation of the created dataset in specified target schema can be performed. Dataset can be inspected in comparable tabular view of source and target datasets. Figure 4-7 shows structure and example values of the source and target datasets as well as the report list with validated data instances (without errors or warnings which occurred during the process and treated as shown in Figure 4-6).

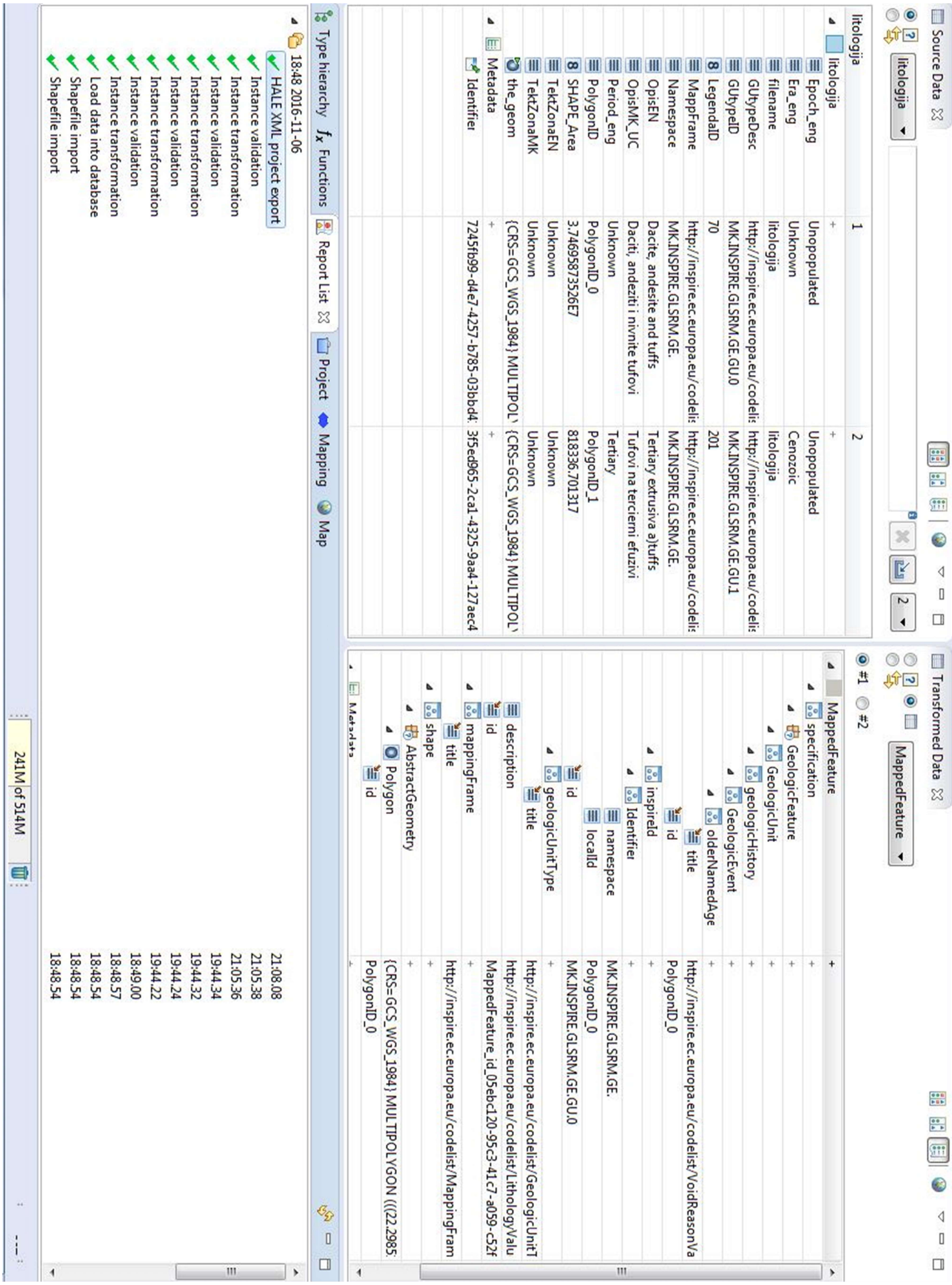


Figure 4-7 Comparison of the source and target schema values and the validated elements status list

Result of the transformation process is being saved in a GML file with standardised structure of the resulting dataset in conformance with the INSPIRE requirements, which is secured by usage of the INSPIRE schema as the target schema. Along with the file for visualisation of the data (SLD) dataset can be used for exchange in interoperable manner on different GIS platforms. The exported GML file is validated using HALE XML validator.

As the practical test of the resulting harmonised data, output harmonised dataset in GML format and the output SLD symbology description file were imported in QGIS, as one of the widest-spread open source GIS tools (Figure 4-8).

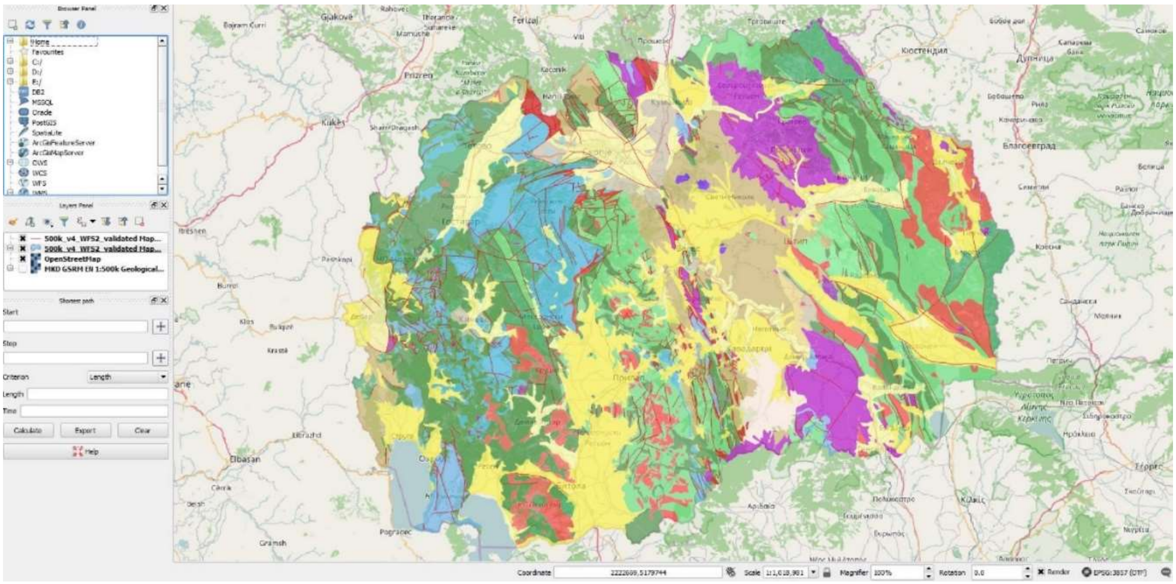


Figure 4-8 GML file import, validation and use in GIS client

Validation of the coordinate referential system ETRS89 is done by overlaying the resulting dataset with web services of the GSRM original basic geological dataset in scale 1:500,000 and the satellite basemap from Open Street Map in QGIS. Data consistency is checked by testing the topological correctness of the output data in comparison with the input data, regarding the number of objects, area they cover, as well their overlapping and existence of gaps. In this way, the envisioned result of the INSPIRE specified harmonisation process implementation was reached through a practical example on the basic geological map in scale 1:500,000.

4.2.1. Implementation of the INSPIRE data harmonisation locally

The presented implementation of the INSPIRE IR on Data Specification requirements sums the process characteristics and outputs giving the Western Balkans SDI stakeholder's view of the harmonisation process. This part depicts the INSPIRE requirements that ought to be met from the aspect of the GSRM as the stakeholder and serves as the basis for the approach to regional SDI stakeholders. The focus is put on the INSPIRE IR Data Specification implementation as the core process in data harmonisation and reaching interoperability of spatial data management. Framework for the process are the spatial standards which define data model structure, encoding, vocabularies and formats.

This practical example showed the importance of the source data structure and quality. Data structure in the technical sense determines the needs of the institution towards meeting the INSPIRE requirements, specifies mapping and transformation functions as well as the needed capacities for data preparation and processing.

Different data preparation and processing needs can be treated with the available software tools, making the harmonisation process smoother and closer to the institution's base data production and management process. Observing this from the business point of view, data harmonisation should be an integral part of the data management process in institution as much as possible, making the data processing and transformation feasible on the fly, modular or extended process to the routine work.

Results of the implemented harmonisation example showed the methodology, automation possibilities and limitations, validation process and capacities in performing the process. Thus, formalisation of the process for the specific regional needs is proven to be feasible and suitable for implementation. Its practical usage is dependent on the potential for integration in the existing production process.

5. Discussion

The paradigm of interoperable data exchange and Europe-wide usage is represented through the EU INSPIRE Directive. This top-down approach challenged the development of the data interoperability on the local level. As local specifics also differ, this thesis focuses on the Western Balkans region as the compact and significant data market regarding its common background and development status. Three main challenges that Western Balkans faces are data quality, legislation support and capacities for supplying and consuming the data. INSPIRE is seen as the framework for approaching this issue. Yet, literal transposition of the Directive is not regarded as a satisfactory approach.

This thesis examines the approach to the INSPIRE Directive implementation in the local context and contributes by revising the data harmonisation process through re-formalising the INSPIRE data model. In order to define the formalisation process, the study considers better comprehension of INSPIRE data harmonisation, analysis of the main regional challenges in Directive implementation and developing the use case of re-formalised INSPIRE data model implementation.

5.1. Comprehending the INSPIRE Directive

Overview of the INSPIRE Directive and its Implementing Rules shows scope of the Directive, domains of regulation and the progress achieved in the EU. INSPIRE decomposes interoperability implementation into segments covered by Implementing Rules on metadata, data, network services, data and service sharing, spatial data services and monitoring. These are technical rules on implementing the Directive in key domains, focusing on data environment and the data itself. Considering the motivation and the goals, the subject of this thesis covers the data harmonisation process as the core component in enabling interoperability and harmonisation.

INSPIRE Data Specification IR for the theme Geology presents the data model, while this study analyses the formalisation and extension approach in relation to stakeholder needs. The analysis shows that the theme scope exceeds the geology domain and the data model includes classes from several other themes. Further, it strives to cover all aspects of Geology theme subdomains (geomorphology, groundwaters, geophysics etc.), introducing complexity of expressing geological objects and relations. On the other hand, the data model offers a unified and standardised description of model elements, as well as the possibility for model expansion using GeoSciML as the data model language. This makes the data model structure firm but modular, thus securing interoperability and a standardised approach in defining elements like objects, relations, domain values and constraints. Detailed descriptions of the geological data model show model structure and the formal rules for its building and managing.

5.2. Formalisation approach

In the context of formalisation approach, most attention is given to the data model formalisation principles, namely application scheme modularity, voidability concept, as well as the extensions. There are different approaches which all are built upon the INSPIRE Generic Conceptual Model, putting it in light as the framework of spatial data interoperability all over Europe. Several examples of implementation of the formalisation concepts were presented, from the software for data harmonisation to examples of INSPIRE data models' extensions. Important findings in this sense are the definition of process flow of the data model extending by Geonovum (Grothe, Bulens & Reitz, 2016), possibility to incorporate user requirements within the INSPIRE GCM compliant data model (Hopfstock, 2016) and the developed approach of building up data model elements using INSPIRE elements as the template (Martirano, Morrone & Vinci, 2016).

Based on the analysis of the INSPIRE model formalisation rules, key phases in implementing INSPIRE compliant data consist of defining local data model and requirements, understanding relation to target – INSPIRE data model specification, schema and data transformation and finally validation of the resulting schema and dataset.

5.3. Formalisation concept and tools

The approach to formalising the INSPIRE data models is broaden with the analysis of the concepts behind the discussed examples. Thus, all phases of harmonising data are subject to standards which define methods, tools and services for managing spatial data. These are applied through (1) notation of data structure and schemas in standardised and interoperable way, (2) schema and data transformation, and (3) formal rules to which data and services need to comply, in order to be in conformity with referential standards. Even the geological INSPIRE data model itself is an example of a more generalised model extension, GeoSciML, which retains standardised structure and remains compliant with underlying standards after modifications.

Regarding the INSPIRE data model managing in a standardised manner, software supporting UML modelling tools are a great asset for revision, model extension and encoding. INSPIRE data model harmonisation process phases are supported and to various extent automated through the use of the software tools, which conform to the INSPIRE requirements. This thesis presents the analysis of the segments of the harmonisation process and the available software tools. As the most comprehensive software tool available as open source, Humboldt HALE is chosen and reviewed. The thesis describes how HALE reflects different harmonisation phases on a conceptual level and serves as the input for practical implementation within the study.

Validation and conformance testing is performed as the last phase of harmonisation. This topic is covered within the INSPIRE Generic Conceptual Model too, but it does not give the possibility for the fully automated process. Thus, this work reviews how Abstract Test Suite defines testing cases and in relation to that, which test can be automated using applicable software tools and which are subject to manual quality checks described by the INSPIRE requirements. The analysis shows that about one third of ATS is covered by the automated processes, while the rest is still in development, thus requiring expert level capacities and experience.

5.4. Possibility for common spatial data model

When the focus is put on the target region of interest, Western Balkans, there are several key factors of interest for the development of the common spatial data model. The SWOT analysis regarding the development of the spatial data infrastructures in WB region reviewed the historical development of spatial data management, legal background and practical experience. Results imply a good normative and strategic position for the INSPIRE Directive implementation, but also absence of standardisation and institutional inconsistence with regard to INSPIRE Directive implementation. In addressing these issues, several examples were presented before formalising data harmonisation for the geology theme in the WB region.

An example from Sweden, within the Nature-SDIplus project and ESDIN project, emphasised the impossibility of a fully automated harmonisation process and the importance of the source model in order to take the most of the transformation process automation. The approach used the Extract – Transform – Load model to simplify and enable the highest efficiency of the process. Portuguese national geological authority used another approach. They restructured the source model in order to enable higher automation level of the data transformation towards the target model. Different experiences show that orientation towards the more automated process caused more significant changes in the source model, while prioritising the retention of the source model structure was more time consuming and required additional processing in comparison with the described structure. Thus, the approach to be taken is directly dependent on the end user needs and the data providers' priorities. This segment of the thesis introduces Study 1, which examines regional critical factors for the INSPIRE harmonisation process.

5.5. Western Balkans constellation of the spatial data modelling

As seen by the geological data management institutions' representatives, harmonisation is not just an obligation, but a necessity in the current work on the geological data. It is seen as the key factor in improving the services for the private sector. Besides general knowledge of the INSPIRE Directive and the harmonisation process among the data producers, the lack of a coordinated approach and performance in this direction is a cause of slow progress. Aside from the low capacities, due to the current state of the data structures, harmonisation is viewed as a highly complex process and a goal that is difficult to reach.

Relevance of these findings are confirmed through the current implementation status of the INSPIRE Directive. It shows high level of comprehension, higher priority milestones reached and the awareness for the obstacles in further work. Data structure is by far the highest rated cause of the slow progress of the INSPIRE implementation. Aside from the institutional IT capacities and level of digitisation, legal issues also influence the slow down. This situation requires a multidisciplinary approach and expertise in the area where the capacities are on quite a low level. Such a constellation reflects also on the data model formalisation achievements. There are few examples of data model development and implementation. These are mostly for internal or research use and have not yet had wider applicability.

The results of Study 1 are a valuable input and representation of the needs for formalising an INSPIRE conformant geology data model, especially because some progress has already been made and the environment for implementation is well known. The key factors and questions stated within Study 1 are also the focus of Study 2, which shows the formalised data model implementation process with a strong focus on the regional needs.

5.6. Data model formalisation and implementation

Implementation of the INSPIRE data harmonisation process as the theme of Study 2 firstly deals with the source model structure and the data processing, followed by the formalisation of the target model according to the stakeholders' needs in the region. These are inputs for the transformation process that results in a harmonised geological dataset.

In comparison to the analysed experiences Europe-wide, the approach implemented in Study 2 reflects the Portuguese geological authority concept of structuring the source model to meet both INSPIRE and national requirements. Such orientation is taken bearing in mind the key findings of Study 1. Namely, capacities, current data quality, legal issues and experience in the field implied an approach towards satisfying the main goals – creating a market-oriented, interoperable and accessible dataset, meeting national legal requirements regarding the geological data management and increasing efficiency.

Formalisation of a simple and effective target data model, which complies with the INSPIRE requirements is the result of the stakeholders' needs analysis and the source data model structure. The target data model was defined in relation to obligatory INSPIRE Geology theme application schema elements, bearing in mind the efficient transformation process from the source schema. In this aspect the formalisation included taking elements from the core INSPIRE data specification application schema and composing its modular elements in reflection to the source model and the user needs.

Structuring the source data, in relation to the INSPIRE requirements regarding the geology schema, enabled an automated harmonisation process towards the target data model. HALE software used for the whole process showed independence of the processed data and the underlying technology. It is important to note that available open source solutions can be utilised with guarantee of the high-quality outputs, user-friendly environment and available documentation.

The described approach to formalising the INSPIRE data model in a local context indicates the concept for developing other themes target data models for various level spatial data infrastructures.

6. Conclusions

During the evolution process of the INSPIRE Directive implementation, the concept of interoperability diverged significantly on the EU level of implementation from the local (national, regional, sub-regional) levels. The INSPIRE recommendations and approach are accepted much wider, so today the situation is that a number of regional and local SDIs exist and the Directive is formally accepted even outside of the EU. The described situation caused a need for seamless data exchange guided by more specific and regional dependent characteristics. This thesis focused on the Western Balkans region and the formalisation of the INSPIRE Geology theme spatial data harmonisation as the possible implementation approach. Further, the thesis addressed bringing a better understanding and contribution to the qualitative approach to spatial data infrastructure development in the Western Balkans. Analysing the implementation process background, technical framework, examples from mature SDIs and regional needs in the field of spatial data exchange resulted in the approach to formalising the INSPIRE data harmonisation process. The main conclusions of this approach are summed up in this chapter.

The first objective, studying the INSPIRE data harmonisation process showed how the INSPIRE Directive excels in the conceptual approach to data and services interoperability. The Directive provides a well-structured environment and rules for developing data models, as well as the means for accessing and exchanging data in an efficient and standardised manner. All relevant examples of data models' development covered in this thesis are based on INSPIRE Generic Conceptual Model. Although they differ in their purpose, structure and functionalities, they all comply with the recommended formalisation rules. Finally, the presented analysis of the INSPIRE Directive, IR and the experiences in the field is a valuable source for capacity building and approaching the interoperability process in a systematic and widely adopted manner.

With the aim to propose a methodology of data harmonisation through the formalised INSPIRE data model according to local needs, the thesis focused on the data specification, transformation and conformity with the INSPIRE requirements. The reviewed literature along with the practical examples in this domain provided insight into valuable software tools for data harmonisation. In this context, the description and underlying concepts of INSPIRE data model formalisation provide the means for a practical implementation of INSPIRE technical documents and specify the needed functionalities of the software tools.

The second objective referred to the analysis of the critical factors of the harmonisation process. When put in the local context of the Western Balkans region, the concepts, practice and development of the interoperable data usage showed a dynamic but inconsistent situation. Although there has been certain progress in adopting and implementing the INSPIRE Directive, there is no systematic or synchronised approach towards solving common problems. Study 1 revealed the needs in the regional context which are reached further in this thesis. It can be said that the interest and the potential for data harmonisation and utilisation is high. An important fact is that an integrative and strong effort needs to be dedicated in order to develop such potential. This thesis offers an approach to data harmonisation in the technical domain. Yet, a lot has to be done in fields of (1) regulating domains of INSPIRE recommendations (data management, data accessibility restrictions, data and service sharing licencing models etc.); (2) capacity building on publisher, data management and user levels; and (3) the integral approach to common problems solving, through better SDI coordination in the region.

As the third objective, the test implementation of the INSPIRE harmonisation process, the focus of the work was on the implementation of the INSPIRE harmonised data model presented in Study 2. It is shown that the recommended concepts and current experiences in the field can form the approach, but the key for an effective harmonisation process is to satisfy the user requirements. In the case of the geology data model used at the national geological authority in FYR Macedonia, this included formalising the source data model and the processes which can automate data transformation to meet both national and INSPIRE requirements. Further, user needs were the key factor for re-formalising the target INSPIRE data model. Structuring the application scheme bearing in mind these concepts led to better utilisation, acceptance and further development of the spatial datasets.

The general aim was reached through defining, design and implementing the INSPIRE data harmonisation. The resulting harmonised geological dataset is one possible alternative in the implementation of the INSPIRE harmonisation. As the simplified version of the geological dataset was processed, extensions to object models or their relations were not included. The intension was not to produce a complete and final geological data model, but to define the process, with key factors, steps and methods as the contribution to the INSPIRE data harmonisation on national and regional levels. Through the presented approach and solution, this thesis introduced the basic methods and means which should be further developed as needs evolve.

As a further advance in the area of INSPIRE implementation, tighter cooperation and more integral work is needed on the local SDI levels. In the presented example of the Western Balkans, this would mean including data producers in the SDI technical teams for the purpose of knowledge exchange and building applicable common data models, but also other INSPIRE compliance components. Following this approach, new opportunities, like the need for data model extensions, geoprocessing web services, regulatory documents development etc. could be grasped and expanded on all INSPIRE topics. The final outcome is seen in more efficient data management and a data exchange boost for the benefit of the entire market.

References

Aleksic, I. R. (2013). INSPIRATION – Spatial Data Infrastructure in the INSPIRATION : National Report SERBIA.

Alhir, S. S. (2002). Guide to Applying the UML.

Cetl, V., Tóth, K., Abrami, A. & Smits, P. (2013). Report on the Status of INSPIRE in the Balkan Countries.

Cetl, V., Tóth, K. & Smits, P. (2014). Development of NSDIs in Western Balkan Countries in Accordance with INSPIRE, *Survey Review*, vol. 46, no. 338, pp.316–321.

Commission of the European Communities. (2008). Western Balkans: Enhancing the European Perspective.

Crompvoets, J., Vandenbroucke, D., Nedović-Budić, Z. & Todorovski, D. (2010). Detailed Survey Concerning INSPIRE - Coordination, Funding and Sharing Measures in South-East Europe, in *International Conference SDI 2010*, 2010.

Drafting Teams 'Data Specification' 'Network Services' 'Metadata'. (2007). INSPIRE Technical Architecture - Overview, vol. 2007, no. march, pp.1–12.

eENVplus Consortium. (2014). eENVplus Geological Map Harmonization in Italy and Slovenia, *Online*, Available Online: http://showcase.eenvplus.eu/client/ep09.htm [Accessed 7 July 2017].

European Commission. (2015). INSPIRE Implementation Roadmap, Available Online: http://inspire.ec.europa.eu/inspire-roadmap/61 [Accessed 22 January 2017].

European Commission. (2017a). INSPIRE Principles, Available Online: http://inspire.ec.europa.eu/inspire-principles/9 [Accessed 22 January 2017].

European Commission. (2017b). INSPIRE Consolidated UML Model, Available Online: http://inspire.ec.europa.eu/data-model/approved/r4618-ir/html/ [Accessed 25 March 2017].

European Commission & European Environmental Agency. (2014). Mid-Term Evaluation Report on INSPIRE Implementation.

Geologicharka. (2012). Sheets of the Basic Geologic Map of SFR Yugoslavia - Cyrillic Version, Available Online: https://commons.wikimedia.org/wiki/File%3AOGK-SFRJ-Listovi-cyr.png [Accessed 25 March 2017].

GEORZ Lab & Research department. (2011). INSPIRE Prototypes (Phase 2) - Dutch Kadaster.

GeoZS. (2017). Geological Maps of Slovenia, *Online*, Available Online: http://www.geo-zs.si/index.php/en/products/publications2/geological-maps [Accessed 25 March 2017].

GISIG. (2013). eENVplus Project, *Online*, Available Online: http://www.eenvplus.eu/ [Accessed 7 July 2017].

Grothe, M., Bulens, J. & Reitz, T. (2016). Geonovum Study INSPIRE Extensions, *WeTransform Study*, Available Online: http://www.geonovum.nl/sites/default/files/20161121-INSPIRE-Extensions.pdf [Accessed 25 March 2017].

Hedefalk, F. & Östman, A. (2011). Making Swedish Environmental Geodata INSPIRE Compliant: A Harmonization Case Study, *Mapping and Image Science*, no. 3, pp.30–37.

Hemmatnia, E., van den Broucke, J. & van Raamsdonk, K. (2010). Harmonising Dutch National Geodata Conformant to INSPIRE Using Combined Transformation, *International Journal of Spatial Data Infrastructures Research*, vol. 5, pp.365–381.

Hopfstock, A. (2016). ELF Data Specification.

Humboldt Consortium. (2011). Survey on the Situation and Perspectives of the GI Market in Europe, Available Online: http://www.esdi-humboldt.eu/survey.html [Accessed 3 April 2017].

INSPIRE. (2003). Report on the Feedback of the Internet Consultation on a Forthcoming EU Initiative Establishing a Framework for the Creation of an Infrastructure for Spatial Information in Europe.

INSPIRE Drafting Team 'Data Specifications'. (2013). INSPIRE Generic Conceptual Model, *INSPIRE Data*

Specification, p.157.

INSPIRE Registry. (2017). INSPIRE Registry - Mapping Frame.

INSPIRE Thematic Working Group Geology. (2011). D2.8.II.4 INSPIRE Data Specification on Geology – Technical Guidelines.

IUGS Commission for the Management and Application of Geoscience Information. (2013). GeoSciML v3.2 Application Schema, Available Online: http://www.geosciml.org/geosciml/3.2/documentation/html/index.htm?goto=1 [Accessed 25 March 2017].

Jovičić, L. (2016). Реализација На Геопросторни Мрежни Услуги Во Рамките На Геоинформациониот Систем На Геолошкиот Завод На Република Македонија, in *Third Congress of Geologists of Republic of Macedonia*, 2016.

JRC. (2017). INSPIRE Validator, Available Online: http://inspire-sandbox.jrc.ec.europa.eu/etf-webapp/ [Accessed 15 December 2017].

Laxton, J. L., Serrano, J.-J. & Tellez-Arenas, A. (2010). Geological Applications Using Geospatial Standards - an Example from OneGeology-Europe and GeoSciML, *International Journal of Digital Earth*, vol. 3, pp.31–49.

Martirano, G., Morrone, S. & Vinci, F. (2016). Using and Extending INSPIRE Data Models, in *Workshop The GeoSmartCity Hub: A Data Platform for Supporting the Operativeness of Smart Cities*, 2016, Barcelona: INSPIRE Conference 2016, p.23.

Mihaljević, I. (2011). Schema Transformation between Different Modeling La Nguages in the INSPIRE Context Using GML.

Ministerstvo za Ekonomija. (2007). Pravilnik Za Izrabotka Na Osnovna Geoloska Karta - 2 Na Republika Makedonija 1:50,000.

Ministry of Mining and Energy of the Republic of Serbia. (2015). BEWARE Project Description & Goals, *Online*, Available Online: http://geoliss.mre.gov.rs/beware/?page_id=8 [Accessed 7 July 2017].

Neuschmid, J., Beyer, C., Eizinger, C., Schrenk, M. & Wasserburger, W. (2010). Plan4all – State of the Art in the Harmonisation of Spatial Planning Data, in *Real Corp 2010 Proceedings/Tagungsband*, 2010.

Ogrizovic, V. (2013). 6 TH REGIONAL STUDY ON CADASTRE AND NSDI Vukan Ogrizović Table of Contents.

Östman, A. & Abugessaisa, I. (2014). Data Harmonisation Concepts.

Pavlov, D., Jovičić, L., Postolova, S. & Mitev, I. (2015). Logical Model of Mapping and Labeling of Lithological Units on the Basic Geological Map of the Republic of Macedonia, in *Proceedings – The First Congress of Geologists of Bosnia and Herzegovina*, 2015, Tuzla, pp.71–75.

Pereira, A. M., Luís, A. G. & Cabral, P. B. (2013). INSPIRE Geology Data Model Implementation in Digital Geological Map Production in Portugal: A Preliminary Approach, *International Journal of Spatial Data Infrastructures Research*, vol. 8, pp.128–143.

Reitz, T. (2016). INSPIRE Data Specifications Extensions in GeoSmartCity, *WeTransform News*, Available Online: https://www.wetransform.to/news/2016/04/07/inspire-data-specifications-geosmartcity/ [Accessed 25 March 2017].

Richard, S. M. & CGI, I. W. G. (2007). GeoSciML -- A GML Application for Geoscience Information Interchange, in *Digital Mapping Techniques ’07--Workshop Proceedings: U.S. Geological Survey*, Vol. 1285, 2007, pp.47–59.

Rinne, I. (2014). Opening the INSPIRE Conformance Testing, Available Online: http://www.spatineo.com/2014/12/opening-inspire-conformance-testing/ [Accessed 25 March 2017].

Rudarsko-geološki fakultet. (2011). Geološki Informacioni Sistem Srbije - Geološka Terminologija I Nomenklatura, Available Online: http://geoliss.mre.gov.rs/recnik/OProjektu.aspx.

Tamash, N. (2012). INSPIRE Compliant Datasets INSPIRE Compliant Datasets – Transformation & Conformance Testing.

The European Parliament and the Council of the European Union. (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), 50 Official Journal of the European Union 1–14.

The Geological Surveys of Europe. (2014). EGDI SCOPE - Integrating 3D Model Data into European Data Infrastructure, Available Online: http://www.egdi-scope.eu/.

The HUMBOLDT project. (2010). The HUMBOLDT Framework, Available Online: http://www.dhpanel.eu/humboldt-framework/overview.html [Accessed 25 March 2017].

Tomas, R., Harrison, M., Barredo, J. I., Thomas, F., Llorente Isidro, M., Pfeiffer, M. & Čerba, O. (2015). Towards a Cross-Domain Interoperable Framework for Natural Hazards and Disaster Risk Reduction Information, *Natural Hazards*, vol. 78, no. 3, pp.1545–1563.

Tóth, K., Portele, C., Illert, A., Lutz, M. & Lima, M. N. De. (2012). Interoperability Specifications in Spatial Data Infrastructures.

Villa, P., Molina, R., Gomarasca, M. A. & Roccatagliata, E. (2012). Harmonisation Requirements and Capabilities towards a European Spatial Data Infrastructure (ESDI): The HUMBOLDT Protected Areas Scenario, *International Journal of Digital Earth*, vol. 5, no. 5, pp.417–438.

wetransform. (2016). Survey on Extending INSPIRE Data Models, Available Online: http://inspire-extensions.wetransform.to/survey.html.

Appendices

Appendix A - Main UML stereotypes in INSPIRE Geology schema

Table A-1 Main UML stereotypes in INSPIRE Geology schema adapted from (INSPIRE Thematic Working Group Geology, 2011)

Stereotype	Model element	Description
<<applicationSchema>>	Package	An INSPIRE application schema according to ISO 19109 and the Generic Conceptual Model.
<<leaf>>	Package	A package that is not an application schema and contains no packages.
<<featureType>>	Class	A spatial object type.
<<type>>	Class	A type that is not directly instantiable, but is used as an abstract collection of operation, attribute and relation signatures. This stereotype should usually not be used in INSPIRE application schemas as these are on a different conceptual level than classifiers with this stereotype.
<<dataType>>	Class	A structured data type without identity.
<<union>>	Class	A structured data type without identity
<<enumeration>>	Class	An enumeration, or a list of values attribute can take, stored in the data model.
<<codeList>>	Class	A code list. List of attribute values stored in external list (not in data model). INSPIRE requires code lists to be available through registers.
<<import>>	Dependency	The model elements of the supplier package are imported.
<<voidable>>	Attribute, association role	A voidable attribute or association role. Used to describe object characteristic when the value is not present in dataset but can exist in real world. Code list: Unpopulated, Unknown, Withheld.
<<lifeCycleInfo>>	Attribute, association role	If in an application schema a property is considered to be part of the life-cycle information of a spatial object type, the property shall receive this stereotype.
<<version>>	Association role	If in an application schema an association role ends at a spatial object type, this stereotype denotes that the value of the property is meant to be a specific version of the spatial object, not the spatial object in general.

Appendix B - Theme Geology Conceptual data model

UML diagram showing overview of the classes and relations for the geological conceptual data model is presented in the Figure B-1. Only the top level of the data model is depicted along with the classes and relations without data types, code lists and enumerations.

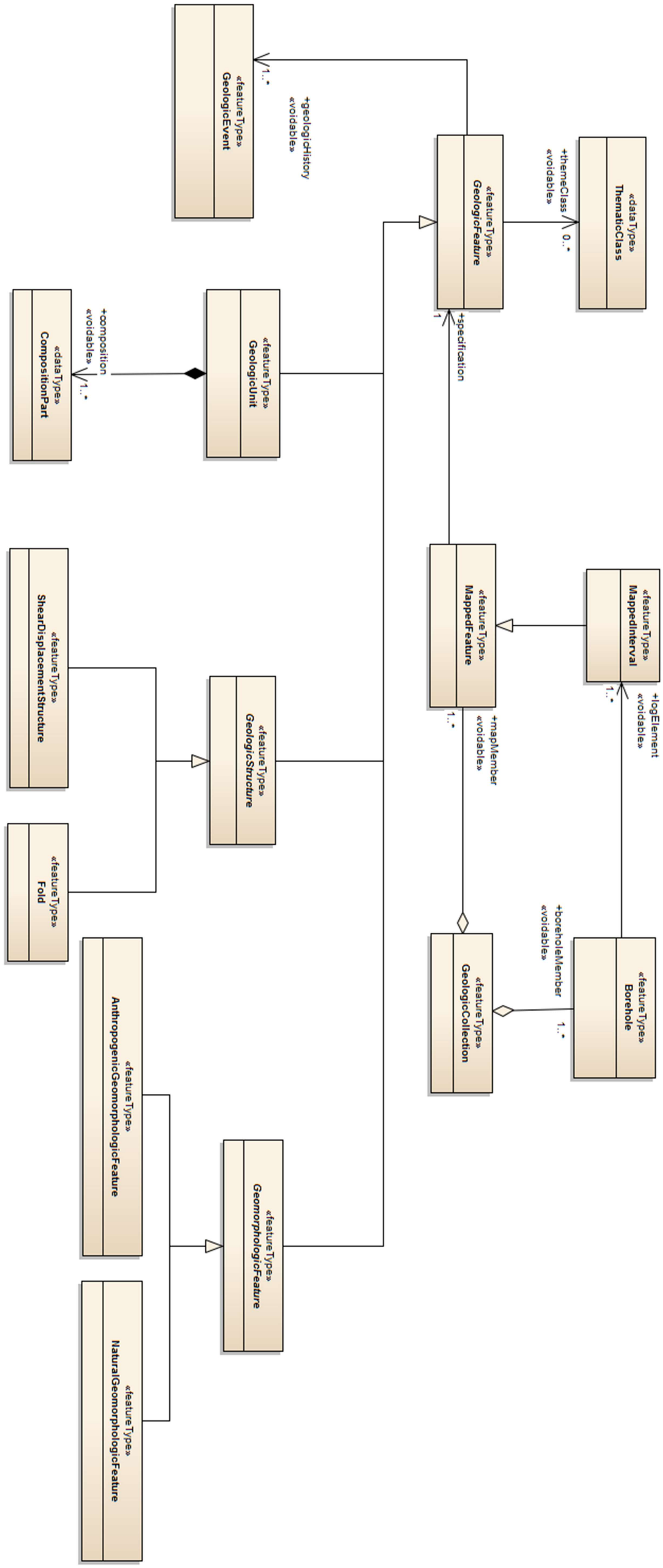


Figure B-1 INSPIRE Consolidated UML Model (European Commission, 2017b)

Representative classes of one segment of the model are shown in detail in order to depict how they are formalised and managed. For further information about classes as well as referred external data models the source literature is INSPIRE Data Specification on Geology, version 3.0. (INSPIRE Thematic Working Group Geology, 2011) The central part of the model is formalised through the description of its classes and related objects (Figure B-2).

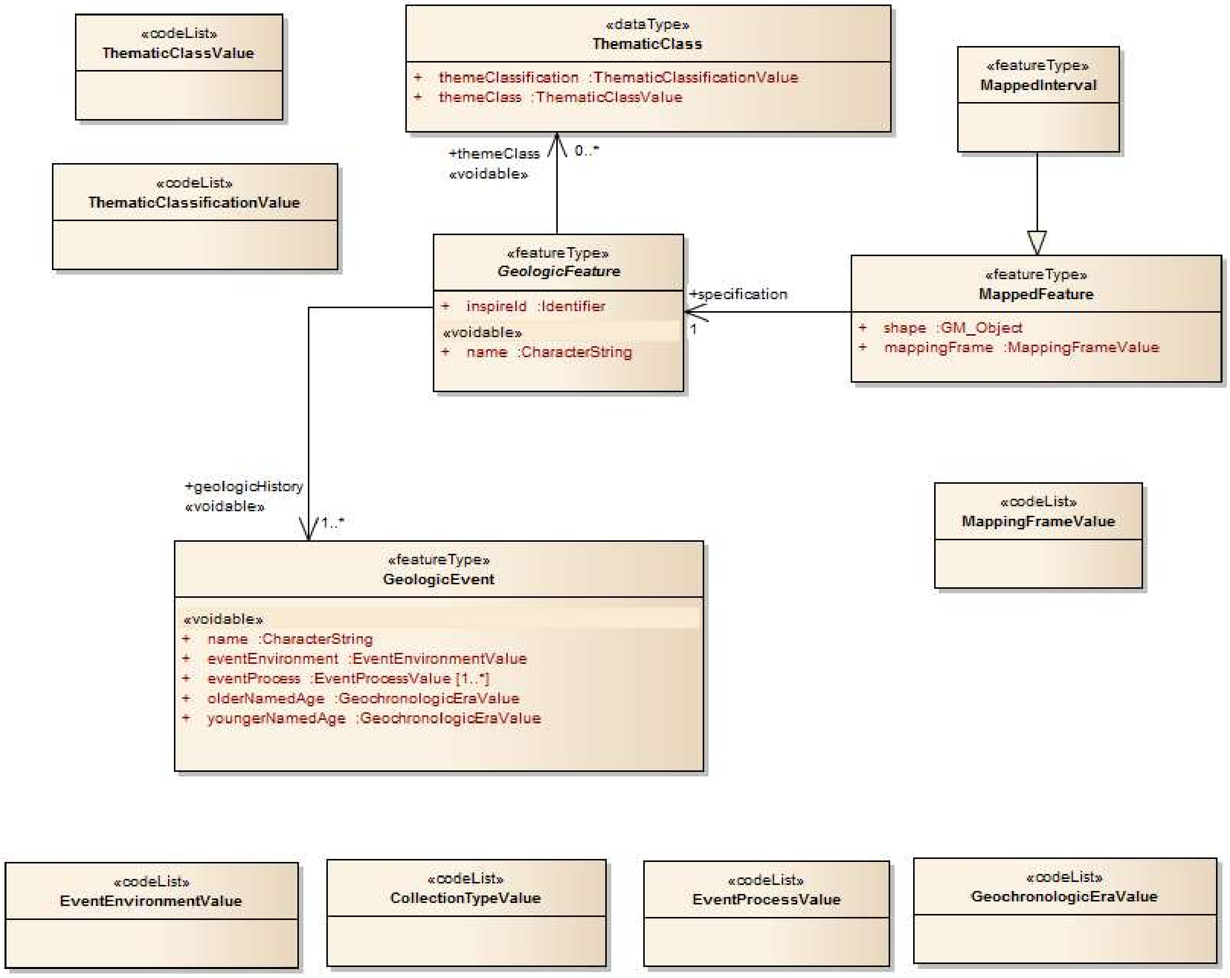


Figure B-2 Schema of the INSPIRE data model segment, theme Geology. Source: (INSPIRE Thematic Working Group Geology, 2011)

Model segment shows classes and their relations as well as reference code lists. Description of each feature class, attributes, code list with relevant information is contained within the Feature Catalogue, as the separate part of the Data specification. An example is presented in Figure B-3.

5.3.2.1.6. GeologicFeature

GeologicFeature (abstract)

Definition:	A conceptual geological feature that is hypothesized to exist coherently in the world.
Description:	This corresponds with a "legend item" from a traditional geologic map. While the bounding coordinates of a Geologic Feature may be described, its shape is not. The implemented Geologic Feature instance acts as the "description package"
Stereotypes:	«featureType»

Attribute: inspireId

Value type:	Identifier
Definition:	External object identifier of the spatial object.
Multiplicity:	1

Attribute: name

Value type:	CharacterString
Definition:	The name of the geologic feature.
Description:	EXAMPLE: a lithostratigraphic unit, mineral occurrence, or major fault. Not all GeologicFeatures will have names, for example minor faults.
Multiplicity:	1
Stereotypes:	«voidable»

Association role: themeClass

Value type:	ThematicClass
Definition:	A thematic classification of the geologic feature.
Description:	A GeologicFeature may be classified according to one or more thematic schema, for example ground stability or mineral resource potential.
Multiplicity:	0..*
Stereotypes:	«voidable»

Association role: geologicHistory

Value type:	GeologicEvent
Definition:	An association that relates one or more geologic events to a geologic feature to describe their age or geologic history.
Multiplicity:	1..*
Stereotypes:	«voidable»

Figure B-3 Feature catalogue elements – example of GeologicFeature attributes and associations description. Source: (INSPIRE Thematic Working Group Geology, 2011)

Appendix C – Questionnaire

INSPIRE Harmonization Models in Western Balkans

This is the questionnaire for the Master thesis research that addresses regional SDI stakeholders with the aim to identify needs and specifics regarding the implementation of the INSPIRE harmonisation model. Domain of research is narrowed to INSPIRE theme geology. The questionnaire treats data structuring approach and expectations, level of INSPIRE Directive implementation, experience with data harmonisation in order to form the basis for regional model development in accordance to INSPIRE requirements.

* Required

1. **Email address** *

INSPIRE harmonization needs

Organisations' view on the needs for spatial data harmonisation.

2. **1/6 How do you estimate the necessity within the Western Balkans region to harmonise geological data in relation to its content, quality and actuality?** *

Mark only one oval.

	1	2	3	4	5	
Very low	◯	◯	◯	◯	◯	Very high

3. **2/6 In which application areas is the interoperability of spatial data of major importance for your organisation?**

4. **3/6 Your organisation sees the ensuring of the data interoperability as:** *

Mark only one oval.

- ◯ periodical campaign work
- ◯ continuous work of a dedicated department/team
- ◯ work demanding resources which increase over time
- ◯ work that considers engaging outsource expertise knowledge
- ◯ need for external resources in terms of spatial data and services management
- ◯ Other:

5. **4/6 The implementation of INSPIRE is well co-ordinated between my country and its neighbouring countries** *

Mark only one oval.

	1	2	3	4	5	
Agree strongly	◯	◯	◯	◯	◯	Disagree strongly

6. **5/6 Which are the most valuable benefits for your organisation expected from the data harmonisation by INSPIRE?** *
Mark only one oval per row.

	Agree strongly	Agree	No opinion	Disagree	Disagree strongly
Reducing duplication of data collection costs (by accessing a single, harmonised data set)	⬭	⬭	⬭	⬭	⬭
Development of standardised fundamental core spatial databases, from which new products and services can be developed more cheaply and more quickly.	⬭	⬭	⬭	⬭	⬭
Enabling easier discovery of datasets via use of standardised metadata and publication of such metadata via electronic means (web catalogue services).	⬭	⬭	⬭	⬭	⬭
Better, cross-departmental co-ordination of spatial data collection and publishing regimes, due to the availability of harmonised datasets.	⬭	⬭	⬭	⬭	⬭
Faster access to spatial data (especially with Web-based delivery)	⬭	⬭	⬭	⬭	⬭
Efficiency gains from wider access to better quality data both internally to an organisation and across organisations and disciplines.	⬭	⬭	⬭	⬭	⬭
Benefits to society, e.g. better policy making, implementation, and monitoring.	⬭	⬭	⬭	⬭	⬭
Brings added value for private business.	⬭	⬭	⬭	⬭	⬭

7. **6/6 Which are the biggest drawbacks for your organisation expected from the data harmonisation by INSPIRE? ***

Mark only one oval per row.

	Agree strongly	Agree	No opinion	Disagree	Disagree strongly
Need of additional resources in terms of hardware, software or expert knowledge	⬭	⬭	⬭	⬭	⬭
Extensive and time-consuming data processing and restructuring	⬭	⬭	⬭	⬭	⬭
Current data/data specifications quality	⬭	⬭	⬭	⬭	⬭
Lack of institutional coordination and support on the national level regarding INSPIRE implementation	⬭	⬭	⬭	⬭	⬭
Creation of the redundant data (INSPIRE compliant data and data in the present format)	⬭	⬭	⬭	⬭	⬭
Gap between the data harmonisation requirements and the current data management procedures	⬭	⬭	⬭	⬭	⬭
Data accessibility restrictions and/or data and service sharing licencing models issues	⬭	⬭	⬭	⬭	⬭
Clarity of INSPIRE technical documentation and complexity of INSPIRE models and requirements	⬭	⬭	⬭	⬭	⬭

INSPIRE harmonised data status

Overview of the level of harmonisation of the spatial data within the organisation.

8. **1/5 My organisation has been involved in projects implementing INSPIRE measures (If yes please provide names/links)** *

9. **2/5 The spatial datasets and services of my organisation follow INSPIRE specifications for:** *

Mark only one oval per row.

	Yes	No	Partially
Documenting (have metadata)	⬭	⬭	⬭
Structuring, follow INSPIRE data specifications	⬭	⬭	⬭
Discovery, through web-based services	⬭	⬭	⬭
Viewing, through web-based services	⬭	⬭	⬭
Downloading, through web-based services	⬭	⬭	⬭

10. **3/5 How do you access third party spatial data?** *

11. **4/5 Which documents in your organisation refer to structure, creation and usage of the basic geological maps and when were they adopted?** *

12. **5/5 In which level the current legislation treats or refers in any way INSPIRE, harmonisation, interoperability, seamless accessibility and usage of the geological data?** *

Mark only one oval.

- ⬭ It doesn't
- ⬭ Very low
- ⬭ Moderate
- ⬭ Low
- ⬭ High
- ⬭ Very high

INSPIRE harmonisation implementation

Organisations' approach to process of spatial data harmonisation according to INSPIRE.

13. 1/8 What is your organisation's approach in implementing the harmonisation of spatial data? *

Mark only one oval.

- Implementation not yet decided
- Own implementation
- External service provider

14. 2/8 Do you already use applications or tools that are offering harmonisation functionality for spatial data? If yes, which? *

____

15. 3/8 What kind of problems or difficulties related to spatial data did you already experience? (Multiple answers possible) *

Check all that apply.

- [ ] Integration/combination of spatial data from different sources
- [ ] Different ontologies (semantics)
- [ ] Complex and time-consuming conversion of data formats
- [ ] Different or inconsistent data models
- [ ] Datasets in different languages
- [ ] Missing, inconsistent or obsolete metadata
- [ ] Inconsistencies of datasets
- [ ] Other: ______________________________

16. 4/8 Spatial data inconsistencies can differ very much. Please prioritise sources causing problems in data harmonisation processes: *

Mark only one oval per row.

	Very low	Low	Moderate	High	Very high
Data formats					
Scales / resolutions / level-of-detail					
Spatial consistency issues (edge-matching etc.)					
Geographic coordinate reference systems					
Ontologies (semantics)					
Natural languages					
Visualisation					
Conceptual schemas (data models)					
Classification schemes					
Metadata profiles					

17. **5/8 What has to be done in your opinion to maximize the speed of the INSPIRE harmonisation process?** *

18. **6/8 Which capacities you have or plan in direction to data harmonisation?**

19. **7/8 How knowledgeable are you about INSPIRE standards for geospatial data and services?** *

Mark only one oval.

- Expert
- Professional user
- Occasional user
- No experience

20. **8/8 Have you worked with any INSPIRE Data Specification or a GML Application Schema?** *

Mark only one oval.

- Yes, to create a specific data model aligned with INSPIRE Data Specifications
- Yes, to integrate different data sets
- Yes, to create a map or download service
- Yes, for other purposes
- No.

Extending INSPIRE harmonisation models

This part of the questionnaire relates only to the organisations that already have INSPIRE harmonisation model developed!

21. **Did your organisation create at least one data model that references or extends INSPIRE data specifications, e.g. by including properties that are references to INSPIRE objects?** *

Mark only one oval.

- Yes, for research
- Yes, to improve internal processes in my organisation
- Yes, to improve external processes (such as reporting) with other organisations
- Yes, to fulfil a legal obligation
- Not yet, but we are planning to Skip to question 32.
- No Skip to question 32.

Extending INSPIRE harmonisation models

This part of the questionnaire relates only to the organisations that already have INSPIRE harmonisation model developed!

22. 1/11 Compared to the INSPIRE Data Specifications, what have you done to create your data model? *

Check all that apply.

- [ ] Created new data types (Feature types, Classes)
- [ ] Added new properties to existing data types
- [ ] Profiled an INSPIRE model/data types by taking away elements
- [ ] Added or extended Codelists
- [ ] Added formal constraints (e.g. Schematron, OCL)
- [ ] Other: ______________________

23. 2/11 What was the process flow of creating and maintaining the data model? *

____

____

____

____

____

24. 3/11 What software package was the model created in? *

____

25. 4/11 What are your input data and which geographical region it covers? (e.g. lithology, stratigraphy, physical properties on local, national, cross-border level) *

____

26. 5/11 What is your current/planned method of harmonised data delivery? (If available, please provide links) *

____

____

____

____

____

27. **6/11 Is the data model already used?** *

Mark only one oval.

- ( ) Yes, in academic research
- ( ) Yes, in pilots or testbed environments
- ( ) Yes, in production environments
- ( ) No, it is in development
- ( ) No, it is deprecated
- ( ) Other: ______________________

28. **7/11 What is the license of your data model?** *

____

29. **8/11 Is the data model documented? If available, please provide links on published documents in checkbox "Other".** *

Check all that apply.

- [ ] Yes, in non-public documents
- [ ] Yes, in public documents/websites
- [ ] Yes, in scientific or professional publications
- [ ] No.
- [ ] Other: ______________________

30. **9/11 What type of challenges did you encounter when designing the data model?** *

Mark only one oval per row.

	Blocking	Serious	Minor	No Issue
Knowledge about base standards insufficient	( )	( )	( )	( )
Unsure about extension methodology	( )	( )	( )	( )
Insufficient knowledge of UML or other modelling languages	( )	( )	( )	( )
Insufficient resources (budget, staff)	( )	( )	( )	( )
Stakeholder involvement and political support	( )	( )	( )	( )

31. **10/11 Which challenges did you encounter when implementing the data model?** *

Check all that apply.

- [ ] Model transformation (e.g. from UML to XML Schema)
- [ ] Data Transformation
- [ ] Systems Integration
- [ ] Data Publishing
- [ ] Data Usage
- [ ] Other: ____________________

32. **11/11 Which best practices have you learned from creating the model? Please include links to any examples or documentation** *

____

____

____

Thank you for your time!

If you are interested in the results of this study, please provide the following optional information. Of course, any personal information you supplied will be duly kept confidential until its deletion. This information will not be distributed to third parties.

Participant information

33. **Name**

____

34. **Email address**

____

35. **Name of your organisation/institution and department** *

____

36. **Application area(s) you are active in:**

____

37. **Your role in the business/organisation:**

Mark only one oval.

- ( ) Managerial function (Executive Board / Managing Director)
- ( ) Department Head / Area Manager
- ( ) Project management
- ( ) Technical expert
- ( ) Other: ____________________

38. **What kind of tasks and/or responsibilities do you have in relation to spatial data? (Multiple answers possible)**

Check all that apply.

- [ ] Provision of GIS or GI-based tools and services
- [ ] Development of GIS or GI-based tools and services
- [ ] Customisation of GI processing tools and services
- [ ] Spatial data acquisition
- [ ] Spatial data analysis
- [ ] Processing of spatial data / geographic information from different sources
- [ ] Data harmonisation / integration
- [ ] Other: ____________________

39. **Please provide any other comment of preference**

____

____

____

- [ ] **Send me a copy of my responses.**

Powered by Google Forms

Appendix D – Institutions contacted for participation in the questionnaire

Table D-1 Institutions contacted for participation in the questionnaire

Country	Stakeholder	Respondent position	Respondent's fields of expertise
Albania	Polytechnic University of Tirana	Professor at the Department of Applied Geology, Environment and Geoinformatics	Provision of GIS or GI-based tools and services; Spatial data acquisition; Spatial data analysis; Processing of spatial data / geographic information from different sources
Bosnia and Herzegovina	The Geological Survey of the Federation of Bosnia and Herzegovina – FZZG	Technical expert	Provision of GIS or GI-based tools and services;
	The Republic Survey for Geological Researches of the Republic of Srpska	Senior expert geologist in landslides and GIS	Development of GIS or GI-based tools and services; Spatial data analysis; Processing of spatial data / geographic information from different sources
Croatia	Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb	Project management in engineering geology and geohazards	Spatial data acquisition; Spatial data analysis; Processing of spatial data / geographic information from different sources
Kosovo*	Independent Commission for Mines and Minerals	Department Head / Area Manager	Provision of GIS or GI-based tools and services; Spatial data analysis; Data harmonisation / integration
FYR Macedonia	The Geological Survey of Republic of Macedonia	Project management in GIS development, spatial data harmonisation, web services implementation	Provision of GIS or GI-based tools and services; Development of GIS or GI-based tools and services; Spatial data analysis; Processing of spatial data / geographic information from different sources; Data harmonisation / integration
Montenegro	University of Montenegro, Faculty of Civil Engineering, Geology Department	Lecturer at the Department of Hydrogeology	Spatial data analysis; Processing of spatial data / geographic information from different sources
Serbia	University of Belgrade, Faculty of Mining and Geology	Project management in Landslides	Spatial data acquisition; Spatial data analysis; Data harmonisation / integration
Slovenia	The Geological Survey of Slovenia	Department Head / Area Manager in GIS and IT	Provision of GIS or GI-based tools and services; Development of GIS or GI-based tools and services; Customisation of GI processing tools and services; Spatial data acquisition; Spatial data analysis; Processing of spatial data / geographic information from different sources; Data harmonisation / integration

* This designation is without prejudice to positions on status, and is in line with UNSCR 1244/1999 and the ICJ Opinion on the Kosovo declaration of independence.

Appendix E – Conceptual data model of the basic geological map

Following lines represent the analysis and the depiction of the Guiding technical rules for making the basic geological map. The analysis also shows how these rules refer to geological data structure, representation and comprehension.

Lithological units completely cover area treated in the geological map, thus topologically being equivalent to the polygons with no gaps or overlapping between them. Surface representation of the lithological units are characterised by the name, symbol, colour, hatch and age. Depending on the type, different combination of attributes and values can be present. Exception from this rule are bedrock alterations representing process occurred in a specific area disregarding the type or border of the lithological units beneath. These are represented on the map as the polygons overlaying lithological units independent of the lithological units themselves. They are described by the area, name and the hatch with defined possible values for the name and hatches. Structure of the lithological units is shown on the Figure E-1.

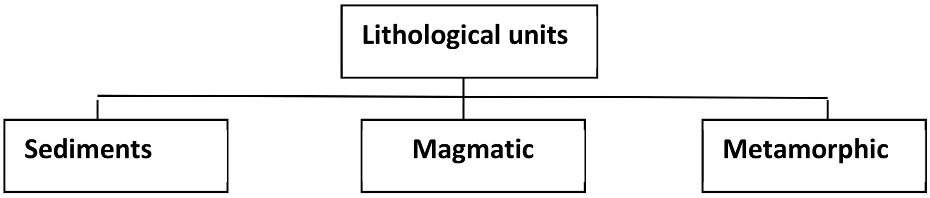


Figure E-1 Basic geological map, lithological units' structure

Following figures are the result of the guideline analysis shown in (Pavlov et al., 2015) and represent the structure of the geological map polygon elements, where each is characterised by the: source of data represented on the map, type of the data value, entry type (free text, domain value, automatically assigned), mandatory constraint, data values domain definition, display on the map (true or false). Additionally, hatch element has the attribute colour of the hatch.

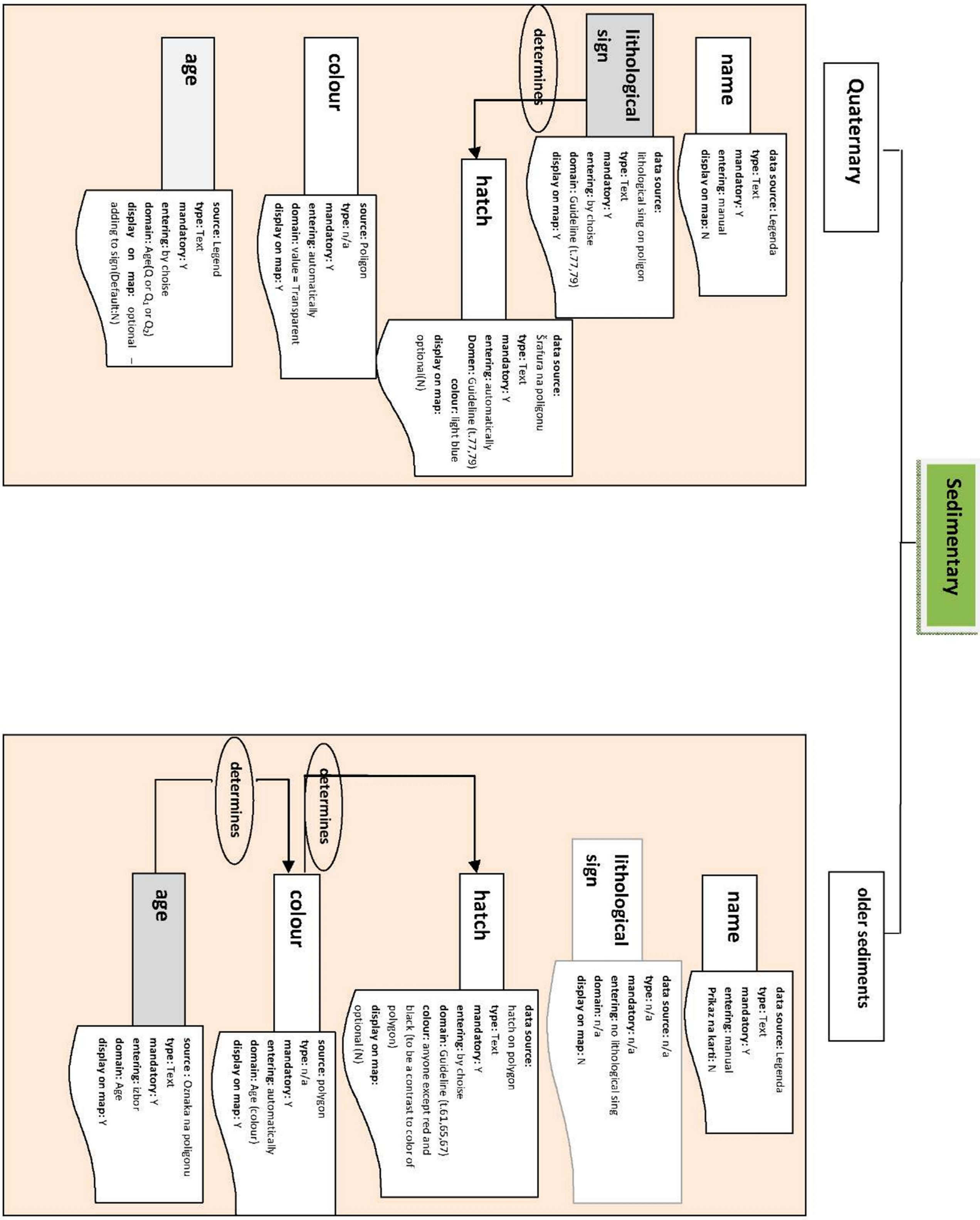


Figure E-2 Sediments and their structure according to mapping principles of the basic geological map (Pavlov et al., 2015)

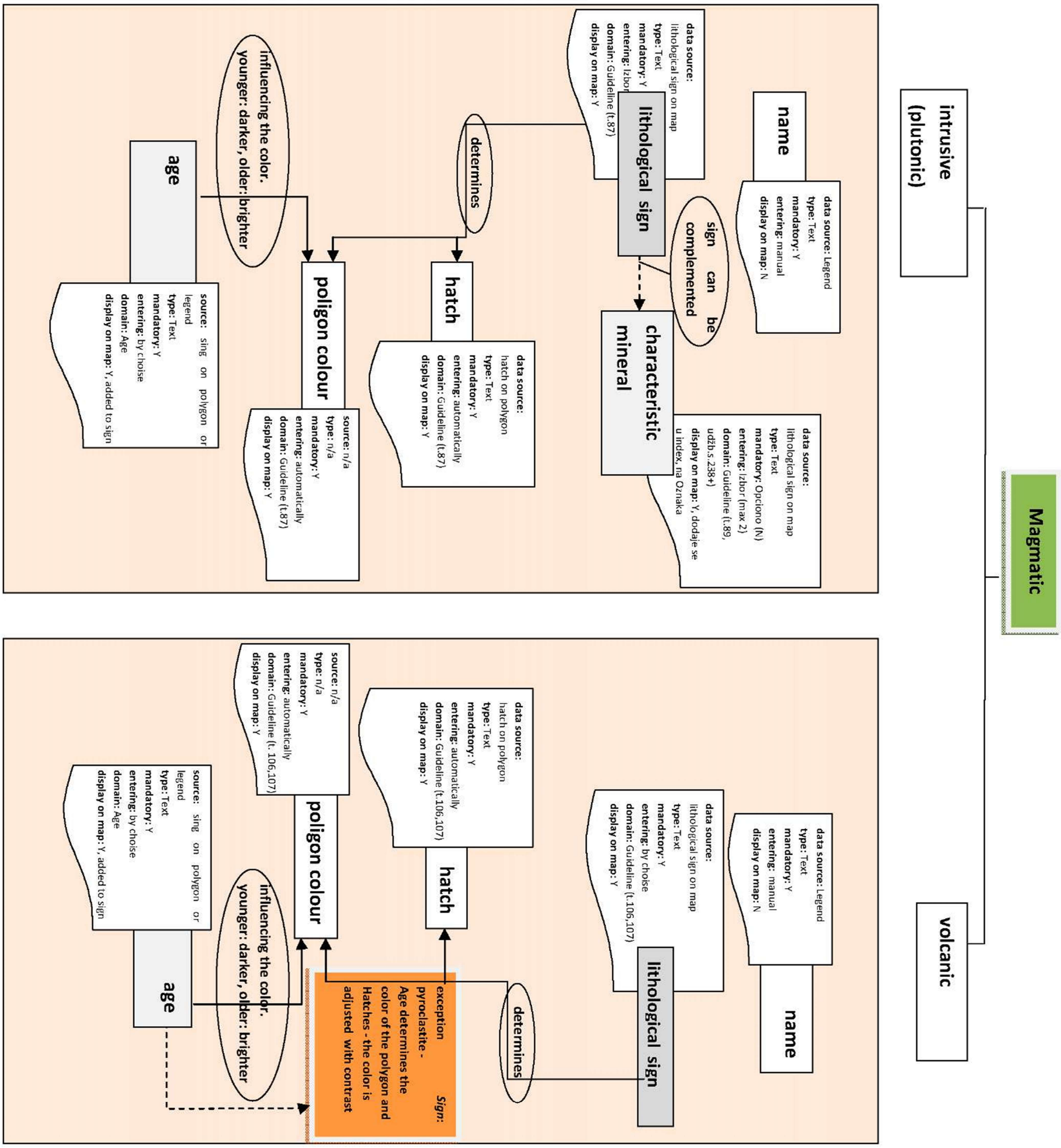


Figure E-3 Magmatic rocks and their structure according to mapping principles of the basic geological map (Pavlov et al., 2015)

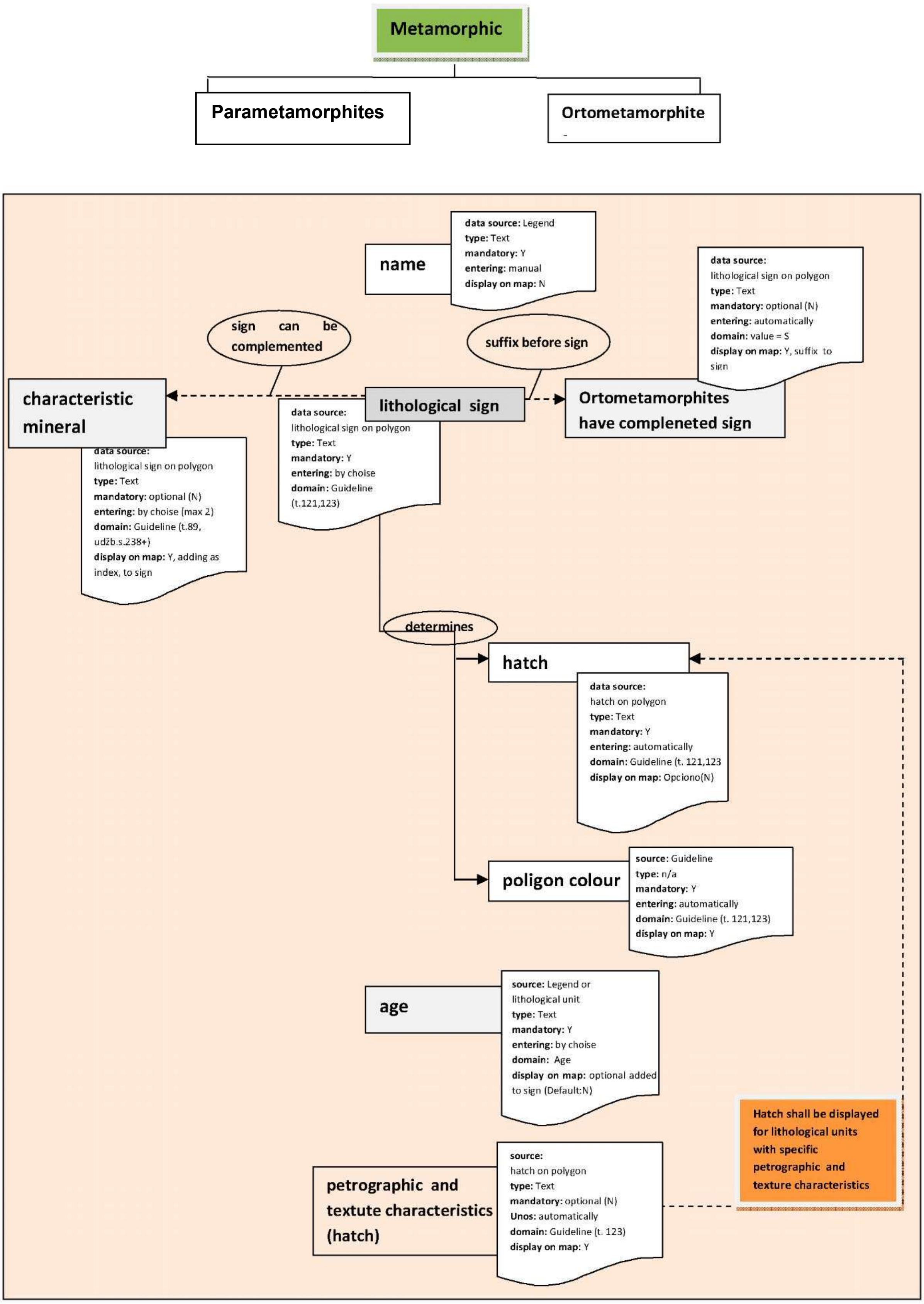


Figure E-4 Metamorphic rocks and their structure according to mapping principles of the basic geological map (Pavlov et al., 2015)

Linear features on the map differ by the type, name and symbology on the map. Symbols are formalised with the technical drawings in the guideline. Figure E-5 and Figure E-6 represent types of linear and point features shown on the map.

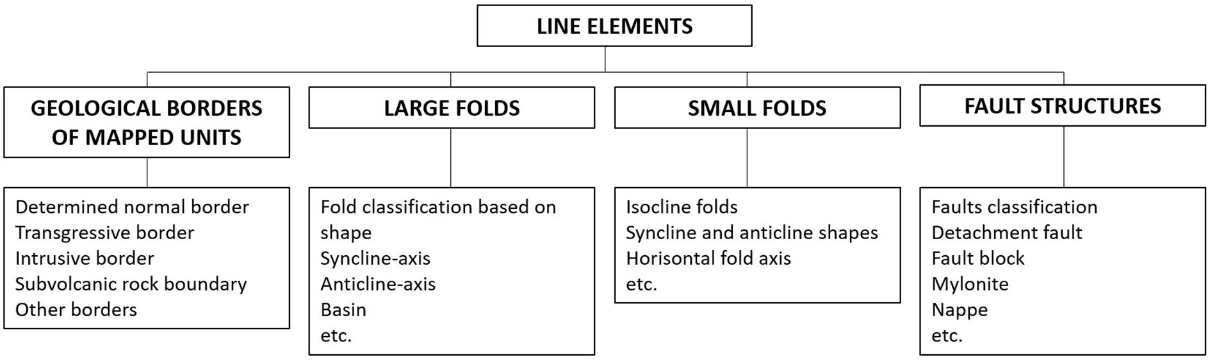


Figure E-5 Types of linear objects of the basic geological map (Pavlov et al., 2015)

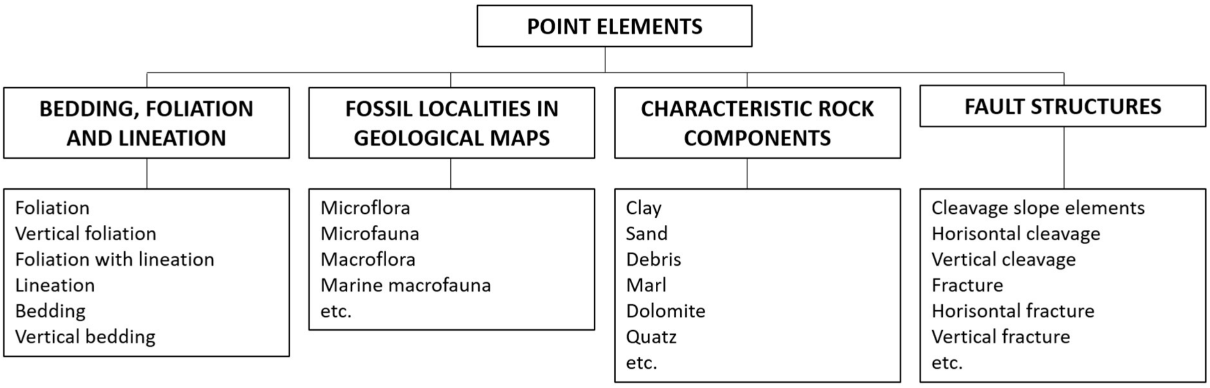


Figure E-6 Types of point objects of the basic geological map (Pavlov et al., 2015)

Appendix F – Formalising the logical data model for the Geological map of FYR Macedonia in scale 1:500,000

On the basis of the Guiding technical rules of making the basic geological map and the consultations with the filed experts from the GSRM, logical data model was formulated for the example dataset – Geological map of FYR Macedonia in scale 1: 500,000.

Two feature classes were created and represented by the three layers shown in the Figure 4-3, namely *Mapped unit* (polygons representing lithological units), *Structural unit* (line features representing tectonic objects) and *Border* (line features representing lithological units' borders). The dataset is visualised using symbology created in ArcMap according to the paper map objects representation. Table F-1 and Table F-2 describes the data model classes.

Table F-1 Data model class description: Mapped unit

Mapped unit	**Data type**	**Description**
ObjectHronoID	short integer	Object ID sorted by the element chronostratigraphic order
NameLegendMkCyr	string	Name of the element as shown in the legend, using Macedonian Cyrillic naming
NameLegendMkLat	string	Name of the element as shown in the legend, using Macedonian Unicode naming
NameLegendEn	string	Name of the element as shown in the legend, in English
AgeMk	string	Name of the age (Macedonian) which characterise map element
AgeEn	string	Name of the age (English) which characterise map element
TectonicZoneMk	string	Name of the tectonic zone characterising map element, in Macedonian
TectonicZoneEn	string	Name of the tectonic zone characterising map element, in English

Mapped units' names in the legend often contain age distinction too. Furthermore, age is specified with different generalisation level, depending on the characteristics of the mapped units, coverage area and level of generalisation of mapped units. In consultations with GSRM attribute table was filled, yet values for the mapped units' age still vary in accuracy level and in some cases, are not even determined.

Tectonic zones are determined for the small number of mapped objects and for the majority could not be determined.

Tectonic units describe borders, faults and thrusts and are defined only by their type as shown in the Table F-2.

Table F-2 Data model class description: Tectonic unit

Tectonic unit	**Data type**	**Description**
NameMk	string	Type of the tectonic element in Macedonian
NameEn	string	Type of the tectonic element in English

Structure of the described features is simple yet contains in a few fields elements that can be further interpreted in order to satisfy basic geological map guideline requirements and consequently to serve as the source data for the harmonisation process. Example of this situation is depicted in the feature description in Table F-3 and Table F-4.

Table F-3 Mapped unit feature description

Mapped unit attribute	Value
ObjectHronoID	18
NameLegendMkCyr	Еоцен – конгломерати
NameLegendMkLat	Eocen – konglomerati
NameLegendEn	Eocene – conglomerates
AgeMk	Палеоген
AgeEn	Paleogene
TectonicZoneMk	Неодредено
TectonicZoneEn	Unknown

Table F-4 Tectonic unit feature description

Tectonic unit	Value
НазивМК	Utvrden rased
НазивЕН	Determined fault

As seen, *NameLegendEn* can be further interpreted and decomposed, since it contains lithological unit type, age and visual representation as distinction which have definition in map guideline document. The same principle leads definition and visualisation of the line elements, while point elements are not represented in this map scale. Thus, rules for interpretation and transformation of the dataset elements are present. This work does not tend to produce definite data model but gives possible interpretation of the basic geological data model as the source model for the harmonisation process.

Appendix G – Source to target model mapping and transformation using HALE

Basic geological map in scale 1:500,000 digital structure as the source model consists of polygon and line object classes described with presented attribute tables. Transformation to target model required creating elements of GML simple object features, in this case polygon and line features.

The first source model feature class *Mapped unit* contained multipart polygons and needed to be transformed into simple features – single part polygons. This task was performed in ArcGIS Desktop as a GIS client, since HALE does not support such transformation. As it was mentioned during analysis of the source model within the section 4.1.1, attribute describing mapped unit age needed to be generalised, while for the object for which no values existed, *Unknown* as a value was set as it is specified in INSPIRE specification for the void value reason. Attributes that did not exist in the source model were added to it as long as they do not violate quality of the information and are requirement of the INSPIRE specification.

This pre-processing of the data resulted in restructured source class of mapped units, called *litologija* (Table G-1) obtained from the base geological map information as described in its logical data model with the class *Mapped unit*.

Table G-1 Source model feature class litologija describing mapped units

Source data model feature class *litologija*		
Attribute name	**Data type**	**Description**
LegendaID	Short integer	Refers to mapped unit attribute ObjectHronoID
OpisMK_UC	String	Refers to mapped unit attribute NameLegendMkLat
OpisEN	String	Refers to mapped unit attribute NameLegendEn
Era_eng	String	Refers to mapped unit attribute AgeEn
Period_eng	String	Refers to mapped unit attribute AgeEn
Epoch_eng	String	Refers to mapped unit attribute AgeEn
TektonskaZonaMK	String	Refers to mapped unit attribute TectonicZoneMk
TektonskaZonaEN	String	Refers to mapped unit attribute TectonicZoneEn
MappFrame	String	'the surface on which the *MappedFeature* is projected' (INSPIRE Registry, 2017), is part of the *MappedFeature* class of the INSPIRE schema presented in Appendix B - Theme Geology Conceptual data model and takes values from the INSPIRE defined code list
PolygonID	String	object ID in the source model and acts as the local identifier, having following structure *PolygonID_*,* where * stands for the source object ID
GUtypeID	String	unique object identifier in general, thus recommended form of the value structure can look like (namespace, schema class ID and finally polygon ID) MK.INSPIRE.GLSRM.GE.GU.100.
GUtypeDesc	String	Describes geological unit type, taking value from the INSPIRE code list GeologicalTypeValue
Namespace	String	Designation of all source datasets, in format like (country code, specification code, institution code, dataset theme). In this case it was decided to be MK.INSPIRE.GLSRM.GE

Similar actions were taken for the line class *StructuralUnit*, where identifying attributes were added as well as attribute describing fault type whose values come from the domain taken from the INSPIRE code list *FaultTypeValue*. Finally, source data line class has the following structure:

Table G-2 Source model feature class tektonika describing strucutral units

Source data model feature class *tektonika*		
Attribute name	**Data type**	**Description**
Tektonika	String	Refers to structural unit attribute *НазивМК*
Tectonics	String	Refers to structural unit attribute *НазивЕН*
MappFrame	String	part of the *MappedFeature* class of the INSPIRE schema presented in section Appendix B - Theme Geology Conceptual data model and takes values from the INSPIRE defined code list
LineID	String	object ID in the source model and acts as the local identifier, having following structure *LineID_*,* where * stands for the source object ID
Namespace	String	Designation of all source datasets, in format like (country code, specification code, institution code, dataset theme). In this case it was decided to be MK.INSPIRE.GLSRM.GE
RasedTip	String	Describes tectonic element type, taking value from the INSPIRE code list FaultTypeValue

Data preparation gave source model more flexibility and potential for the automated process of transformation, without devaluing source data or existing data management process. In this way, mapping of the source data model towards target data model was enabled through the HALE software.

Within the HALE software, process firstly required source data model to be imported. In this case shapefiles representing mapped units and structural units were loaded with the presented prepared data structure, followed by importing data itself. After, target data model – INSPIRE theme Geology core data schema was imported through the URL of the schema from the INSPIRE geoportal (Figure G-1).

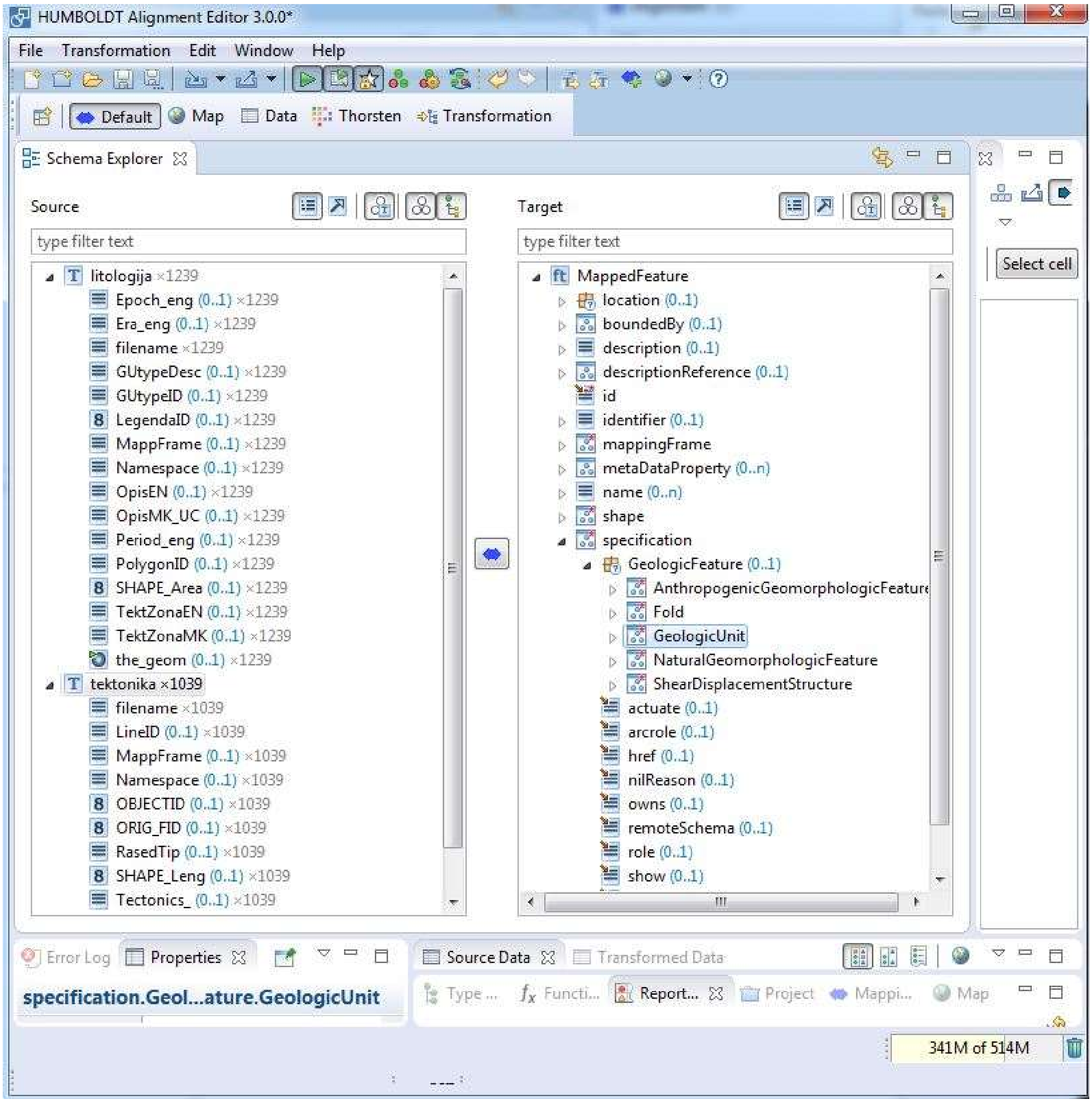


Figure G-1 Imported source and target data model schemas

Bearing in mind that target schema was loaded in its full extents, it was necessary to choose relevant elements for the user requirements. Figure G-2 represents Geology core overview schema described in Appendix B - Theme Geology Conceptual data model, highlighting classes represented by the described features on the base geological map in scale 1: 500,000.

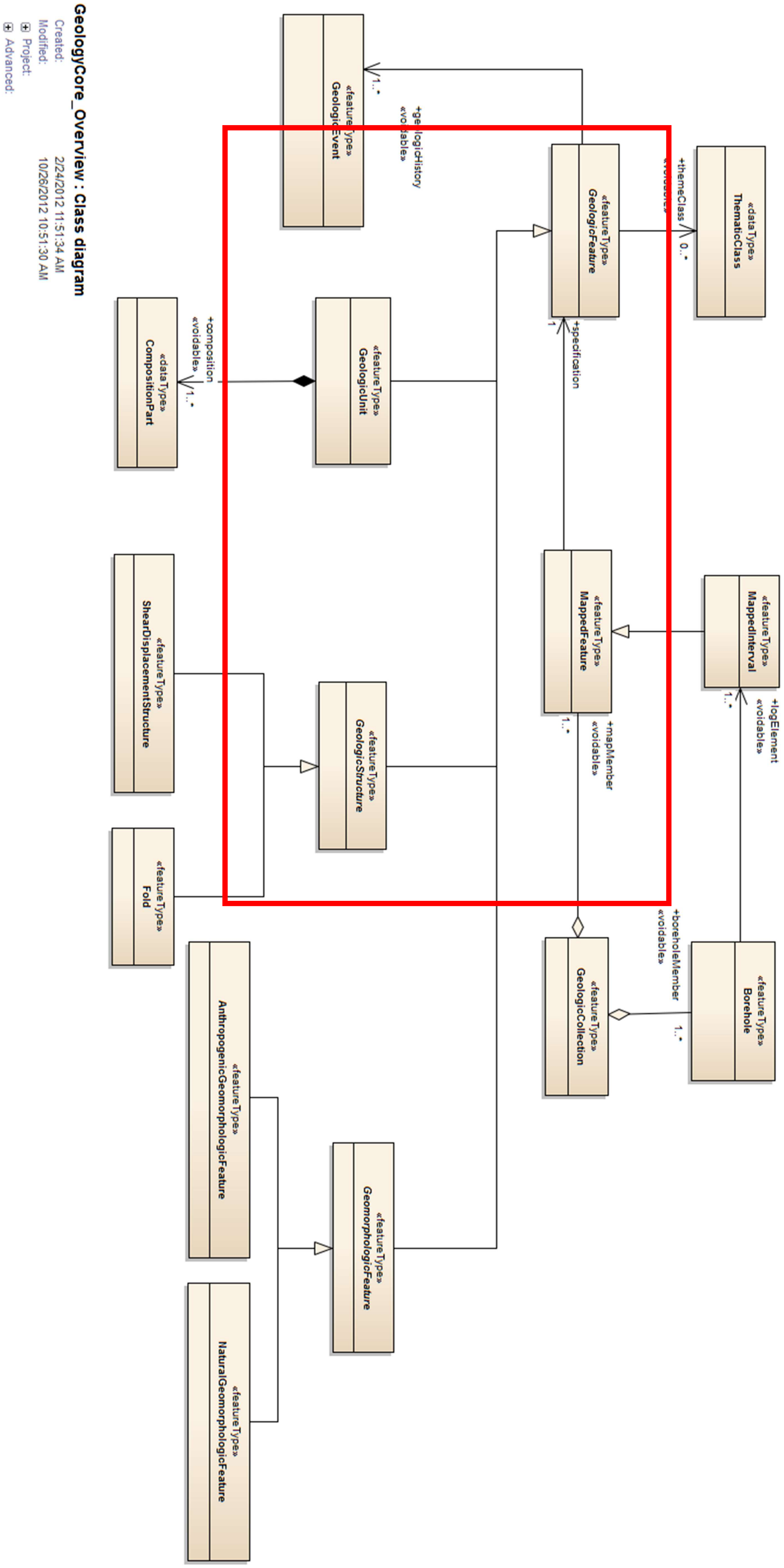


Figure G-2 INSPIRE Consolidated UML Model, Geology core overview schema with highlighted segment of interest

Target schema classes required for representation of the chosen geological dataset and the transformation were *MappedFeature*, *GeologicalFeature* and associated *GeologicalUnit, GeologicalStructure* and *ShearDisplacementStructure* (Figure G-3). Referent classes for the source model were selected using HALE functionality *Edit mapping relevant target types*.

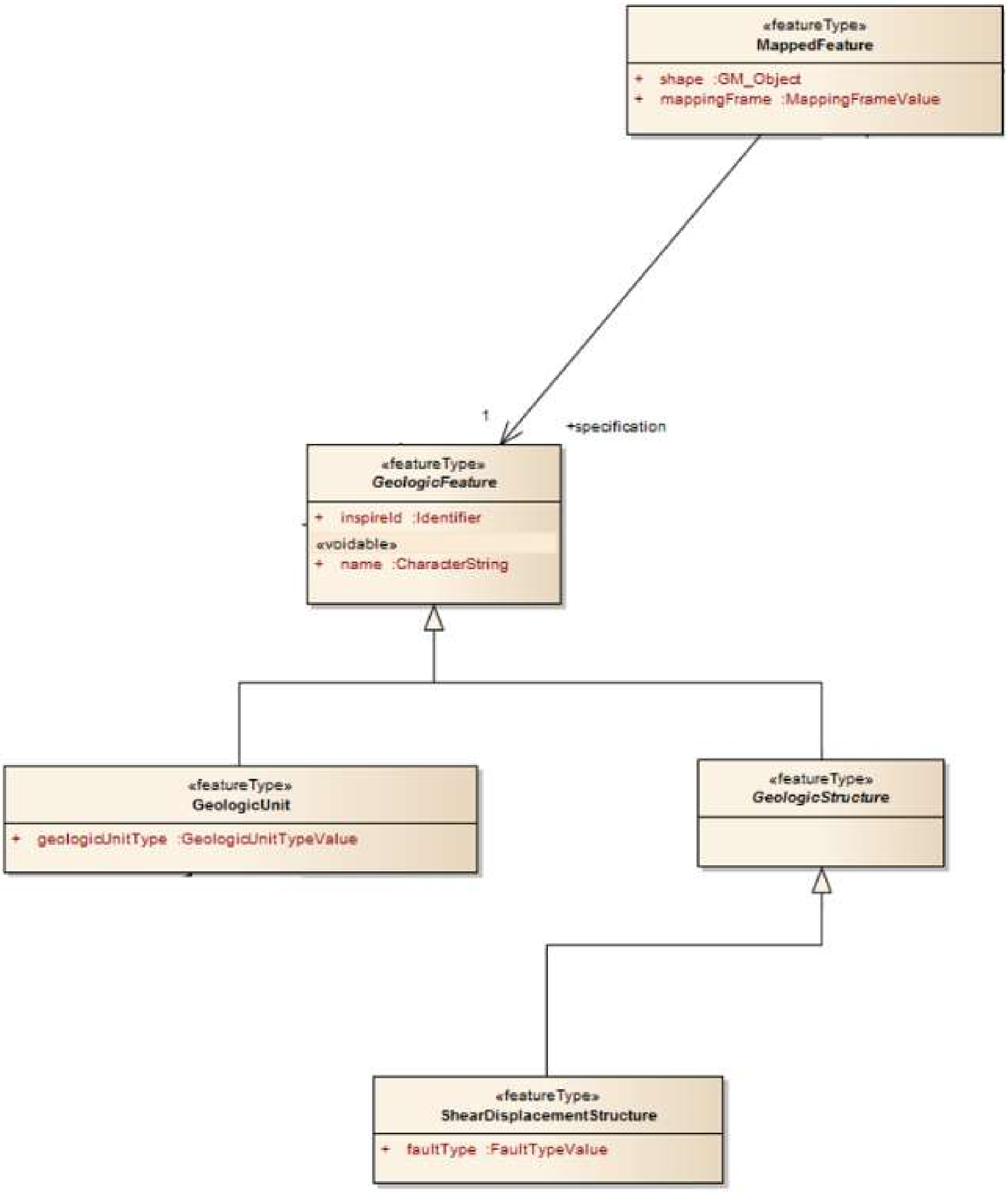


Figure G-3 INSPIRE Consolidated UML Model, Geology core class diagram with selected classes – target schema

Discussed on the logical level, mapping process is considered as the definition of rules for different types of managing data characteristics (geometric, spatial, attribute etc.) in order to meet the target schema element requirements. Mapping elements of the class representing mapped units (named *litologija.shp*) started with marking classes from both models that should be related. In this case *litologija* is connected to *MappedFeature* using HALE function *Retype,* meaning that firstly changing type of the class was needed to meet the target model requirements. The principle is to map target class elements using transformation functions and referent source data model elements. Figure G-4 represents how mapping was done and which functions were used in transforming mapped units' class *litologija* to *MappedFeature* target class.

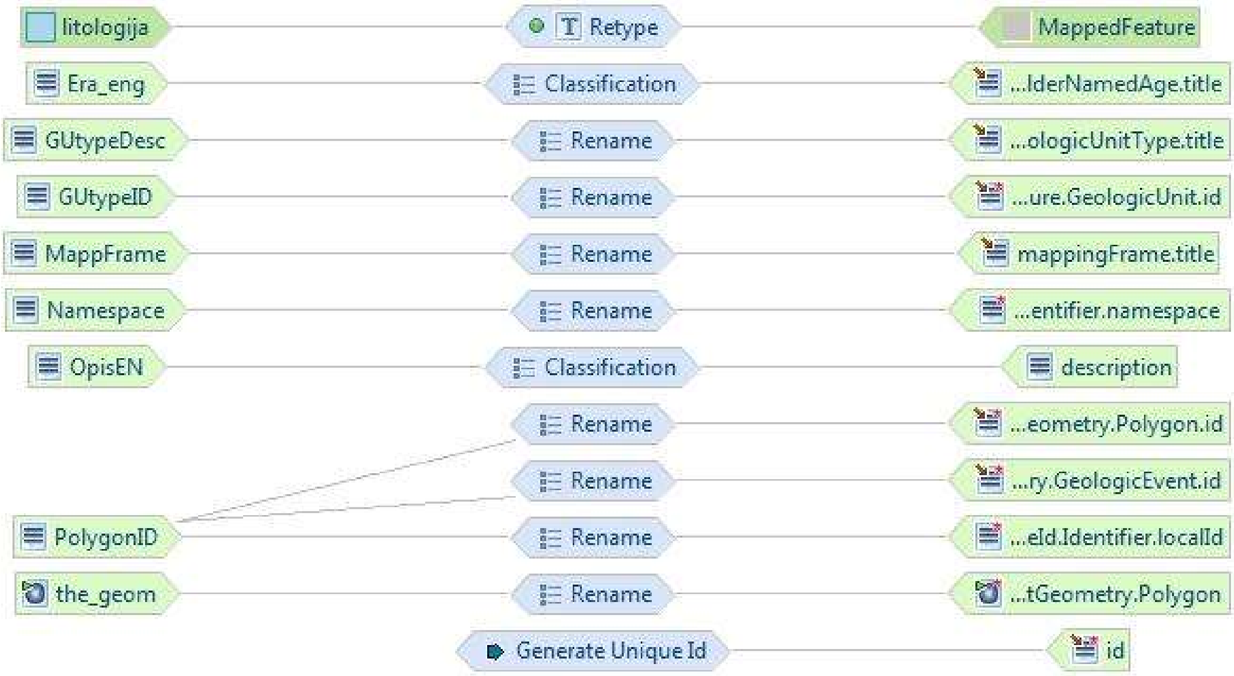


Figure G-4 Mapping functions overview within HALE

Target data model schema includes associated classes too. Thus, mapping shown on the diagram above includes associated class *GeologicUnit*. Detailed mapping rules provided in Figure G-4 are:

Table G-3 Mapping functions in relation to source and target schema objects for MappedUnit class

Source schema element – class *litologija*	Mapping functions	Target schema element – *MappedUnit*
-	*Generate Unique Id*	*id*
MappFrame	*Rename*	*mappingFrame.title*
OpisEN	*Classification*	*description*
the_geom	*Rename*	*shape.AbstractGeometry.Polygon*
PolygonID	*Rename*	*shape.AbstractGeometry.Polygon.id*
Source schema element – class *litologija*	**Mapping functions**	**Target schema element – *MappedUnit specification.GeologicFeature.GeologicUnit.***
PolygonID	*Rename*	*geologicHistory.GeologicEvent.id*
Era_eng	*Classification*	*geologicHistory.GeologicEvent.olderNamedAge.title*
GUtypeDesc	*Rename*	*geologicUnitType.title*
PolygonID	*Rename*	*inspireID.Identifier.localId*
Namespace	*Rename*	*inspireID.Identifier.namespace*
GUtypeID	*Rename*	*id*

Using same concepts, mapping *StructuralUnit* class *tektonika* had following flow:

Table G-4 Mapping functions in relation to source and target schema objects for StructuralUnits class

Source schema element – class *tektonika*	Mapping functions	Target schema element – *StructuralUnit*
-	*Generate Unique Id*	*id*
MappFrame	*Rename*	*mappingFrame.title*
RasedTip	*Rename*	*description*
the_geom	*Rename*	*shape.AbstractGeometry.LineString*
LineID	*Rename*	*shape.AbstractGeometry. LineString.id*
LineID	*Rename*	*specification.GeologicFeature.ShearDisplacementStructure.id*
LineID	*Rename*	*specification.GeologicFeature.ShearDisplacementStructure.inspireId.Identifier.localId*
Namespace	*Rename*	*specification.GeologicFeature.ShearDisplacementStructure.inspireId.Identifier.namespace*
RasedTip	*Rename*	*specification.GeologicFeature.ShearDisplacementStructure.faultType*

Presented example shows usage of several mapping functions – *Rename, Classify, Generate Unique Id,* which are part of the available HALE transformation functions. Other functions which could be utilised from HALE are *Coordinate System Transformation* or *Data Extraction*. All functions are well documented and this work will not discuss them in detail. From the mapping diagram example above, it is also represented how element from the source model can be decomposed into several elements containing same or modified value. HALE also provides a reversed process, generalisation.

Series from Lund University

Department of Physical Geography and Ecosystem Science

Master Thesis in Geographical Information Science

1. *Anthony Lawther:* The application of GIS-based binary logistic regression for slope failure susceptibility mapping in the Western Grampian Mountains, Scotland (2008).
2. *Rickard Hansen:* Daily mobility in Grenoble Metropolitan Region, France. Applied GIS methods in time geographical research (2008).
3. *Emil Bayramov:* Environmental monitoring of bio-restoration activities using GIS and Remote Sensing (2009).
4. *Rafael Villarreal Pacheco:* Applications of Geographic Information Systems as an analytical and visualization tool for mass real estate valuation: a case study of Fontibon District, Bogota, Columbia (2009).
5. *Siri Oestreich Waage:* a case study of route solving for oversized transport: The use of GIS functionalities in transport of transformers, as part of maintaining a reliable power infrastructure (2010).
6. *Edgar Pimiento:* Shallow landslide susceptibility – Modelling and validation (2010).
7. *Martina Schäfer:* Near real-time mapping of floodwater mosquito breeding sites using aerial photographs (2010).
8. *August Pieter van Waarden-Nagel:* Land use evaluation to assess the outcome of the programme of rehabilitation measures for the river Rhine in the Netherlands (2010).
9. *Samira Muhammad:* Development and implementation of air quality data mart for Ontario, Canada: A case study of air quality in Ontario using OLAP tool. (2010).
10. *Fredros Oketch Okumu*: Using remotely sensed data to explore spatial and temporal relationships between photosynthetic productivity of vegetation and malaria transmission intensities in selected parts of Africa (2011).
11. *Svajunas Plunge:* Advanced decision support methods for solving diffuse water pollution problems (2011).
12. *Jonathan Higgins:* Monitoring urban growth in greater Lagos: A case study using GIS to monitor the urban growth of Lagos 1990 - 2008 and produce future growth prospects for the city (2011).
13. *Mårten Karlberg:* Mobile Map Client API: Design and Implementation for Android (2011).
14. *Jeanette McBride:* Mapping Chicago area urban tree canopy using color infrared imagery (2011).
15. *Andrew Farina:* Exploring the relationship between land surface temperature and vegetation abundance for urban heat island mitigation in Seville, Spain (2011).
16. *David Kanyari*: Nairobi City Journey Planner: An online and a Mobile Application (2011).

17. *Laura V. Drews:* Multi-criteria GIS analysis for siting of small wind power plants - A case study from Berlin (2012).
18. *Qaisar Nadeem:* Best living neighborhood in the city - A GIS based multi criteria evaluation of ArRiyadh City (2012).
19. *Ahmed Mohamed El Saeid Mustafa:* Development of a photo voltaic building rooftop integration analysis tool for GIS for Dokki District, Cairo, Egypt (2012).
20. *Daniel Patrick Taylor*: Eastern Oyster Aquaculture: Estuarine Remediation via Site Suitability and Spatially Explicit Carrying Capacity Modeling in Virginia's Chesapeake Bay (2013).
21. *Angeleta Oveta Wilson:* A Participatory GIS approach to *unearthing* Manchester's Cultural Heritage '*gold mine'* (2013).
22. *Ola Svensson:* Visibility and Tholos Tombs in the Messenian Landscape: A Comparative Case Study of the Pylian Hinterlands and the Soulima Valley (2013).
23. *Monika Ogden:* Land use impact on water quality in two river systems in South Africa (2013).
24. *Stefan Rova:* A GIS based approach assessing phosphorus load impact on Lake Flaten in Salem, Sweden (2013).
25. *Yann Buhot:* Analysis of the history of landscape changes over a period of 200 years. How can we predict past landscape pattern scenario and the impact on habitat diversity? (2013).
26. *Christina Fotiou:* Evaluating habitat suitability and spectral heterogeneity models to predict weed species presence (2014).
27. *Inese Linuza:* Accuracy Assessment in Glacier Change Analysis (2014).
28. *Agnieszka Griffin:* Domestic energy consumption and social living standards: a GIS analysis within the Greater London Authority area (2014).
29. *Brynja Guðmundsdóttir:* Detection of potential arable land with remote sensing and GIS - A Case Study for Kjósarhreppur (2014).
30. *Oleksandr Nekrasov:* Processing of MODIS Vegetation Indices for analysis of agricultural droughts in the southern Ukraine between the years 2000-2012 (2014).
31. *Sarah Tressel:* Recommendations for a polar Earth science portal in the context of Arctic Spatial Data Infrastructure (2014).
32. *Caroline Gevaert:* Combining Hyperspectral UAV and Multispectral Formosat-2 Imagery for Precision Agriculture Applications (2014).
33. *Salem Jamal-Uddeen:* Using GeoTools to implement the multi-criteria evaluation analysis - weighted linear combination model (2014).
34. *Samanah Seyedi-Shandiz:* Schematic representation of geographical railway network at the Swedish Transport Administration (2014).
35. *Kazi Masel Ullah:* Urban Land-use planning using Geographical Information System and analytical hierarchy process: case study Dhaka City (2014).
36. *Alexia Chang-Wailing Spitteler:* Development of a web application based on MCDA and GIS for the decision support of river and floodplain rehabilitation projects (2014).

37. *Alessandro De Martino:* Geographic accessibility analysis and evaluation of potential changes to the public transportation system in the City of Milan (2014).
38. *Alireza Mollasalehi:* GIS Based Modelling for Fuel Reduction Using Controlled Burn in Australia. Case Study: Logan City, QLD (2015).
39. *Negin A. Sanati:* Chronic Kidney Disease Mortality in Costa Rica; Geographical Distribution, Spatial Analysis and Non-traditional Risk Factors (2015).
40. *Karen McIntyre:* Benthic mapping of the Bluefields Bay fish sanctuary, Jamaica (2015).
41. *Kees van Duijvendijk:* Feasibility of a low-cost weather sensor network for agricultural purposes: A preliminary assessment (2015).
42. *Sebastian Andersson Hylander:* Evaluation of cultural ecosystem services using GIS (2015).
43. *Deborah Bowyer:* Measuring Urban Growth, Urban Form and Accessibility as Indicators of Urban Sprawl in Hamilton, New Zealand (2015).
44. *Stefan Arvidsson:* Relationship between tree species composition and phenology extracted from satellite data in Swedish forests (2015).
45. *Damián Giménez Cruz*: GIS-based optimal localisation of beekeeping in rural Kenya (2016).
46. *Alejandra Narváez Vallejo:* Can the introduction of the topographic indices in LPJ-GUESS improve the spatial representation of environmental variables? (2016).
47. *Anna Lundgren:* Development of a method for mapping the highest coastline in Sweden using breaklines extracted from high resolution digital elevation models (2016).
48. *Oluwatomi Esther Adejoro:* Does location also matter? A spatial analysis of social achievements of young South Australians (2016).
49. *Hristo Dobrev Tomov:* Automated temporal NDVI analysis over the Middle East for the period 1982 - 2010 (2016).
50. *Vincent Muller:* Impact of Security Context on Mobile Clinic Activities
    A GIS Multi Criteria Evaluation based on an MSF Humanitarian Mission in Cameroon (2016).
51. *Gezahagn Negash Seboka:* Spatial Assessment of NDVI as an Indicator of Desertification in Ethiopia using Remote Sensing and GIS (2016).
52. *Holly Buhler:* Evaluation of Interfacility Medical Transport Journey Times in Southeastern British Columbia. (2016).
53. *Lars Ole Grottenberg*: Assessing the ability to share spatial data between emergency management organisations in the High North (2016).
54. *Sean Grant:* The Right Tree in the Right Place: Using GIS to Maximize the Net Benefits from Urban Forests (2016).
55. *Irshad Jamal:* Multi-Criteria GIS Analysis for School Site Selection in Gorno-Badakhshan Autonomous Oblast, Tajikistan (2016).
56. *Fulgencio Sanmartín:* Wisdom-volkano: A novel tool based on open GIS and time-series visualization to analyse and share volcanic data (2016).

57. *Nezha Acil:* Remote sensing-based monitoring of snow cover dynamics and its influence on vegetation growth in the Middle Atlas Mountains (2016).
58. *Julia Hjalmarsson:* A Weighty Issue: Estimation of Fire Size with Geographically Weighted Logistic Regression (2016).
59. *Mathewos Tamiru Amato:* Using multi-criteria evaluation and GIS for chronic food and nutrition insecurity indicators analysis in Ethiopia (2016).
60. *Karim Alaa El Din Mohamed Soliman El Attar:* Bicycling Suitability in Downtown, Cairo, Egypt (2016).
61. *Gilbert Akol Echelai:* Asset Management: Integrating GIS as a Decision Support Tool in Meter Management in National Water and Sewerage Corporation (2016).
62. *Terje Slinning:* Analytic comparison of multibeam echo soundings (2016).
63. *Gréta Hlín Sveinsdóttir:* GIS-based MCDA for decision support: A framework for wind farm siting in Iceland (2017).
64. *Jonas Sjögren:* Consequences of a flood in Kristianstad, Sweden: A GIS-based analysis of impacts on important societal functions (2017).
65. *Nadine Raska:* 3D geologic subsurface modelling within the Mackenzie Plain, Northwest Territories, Canada (2017).
66. *Panagiotis Symeonidis*: Study of spatial and temporal variation of atmospheric optical parameters and their relation with PM 2.5 concentration over Europe using GIS technologies (2017).
67. *Michaela Bobeck:* A GIS-based Multi-Criteria Decision Analysis of Wind Farm Site Suitability in New South Wales, Australia, from a Sustainable Development Perspective (2017).
68. *Raghdaa Eissa*: Developing a GIS Model for the Assessment of Outdoor Recreational Facilities in New Cities Case Study: Tenth of Ramadan City, Egypt (2017).
69. *Zahra Khais Shahid*: Biofuel plantations and isoprene emissions in Svea and Götaland (2017).
70. *Mirza Amir Liaquat Baig*: Using geographical information systems in epidemiology: Mapping and analyzing occurrence of diarrhea in urban - residential area of Islamabad, Pakistan (2017).
71. *Joakim Jörwall*: Quantitative model of Present and Future well-being in the EU-28: A spatial Multi-Criteria Evaluation of socioeconomic and climatic comfort factors (2017).
72. *Elin Haettner*: Energy Poverty in the Dublin Region: Modelling Geographies of Risk (2017).
73. *Harry Eriksson*: Geochemistry of stream plants and its statistical relations to soil- and bedrock geology, slope directions and till geochemistry. A GIS-analysis of small catchments in northern Sweden. (2017).
74. *Daniel Gardevärn:* PPGIS and Public meetings – An evaluation of public participation methods for urban planning. (2017).
75. *Kim Friberg:* Sensitivity Analysis and Calibration of Multi Energy Balance Land Surface Model Parameters. (2017).
76. *Viktor Svanerud:* Taking the bus to the park? A study of accessibility to green areas in Gothenburg through different modes of transport. (2017).

77. *Lisa-Gaye Greene*: Deadly Designs: The Impact of Road Design on Road Crash Patterns along Jamaica's North Coast Highway. (2017).
78. *Katarina Jemec Parker*: Spatial and temporal analysis of fecal indicator bacteria concentrations in beach water in San Diego, California. (2017).
79. *Angela Kabiru*: An Exploratory Study of Middle Stone Age and Later Stone Age Site Locations in Kenya's Central Rift Valley Using Landscape Analysis: A GIS Approach. (2017).
80. *Kristean Björkmann*: Subjective Well-Being and Environment: A GIS-Based Analysis. (2018).
81. *Williams Erhunmonmen Ojo*: Measuring spatial accessibility to healthcare for people living with HIV-AIDS in southern Nigeria. (2018).
82. *Daniel Assefa*: Developing Data Extraction and Dynamic Data Visualization (Styling) Modules for Web GIS Risk Assessment System (WGRAS). (2018).
83. *Adela Nistora*: Inundation scenarios in a changing climate: assessing potential impacts of sea-level rise on the coast of South-East England (2018).
84. *Marc Seliger*: Thirsty landscapes - Investigating growing irrigation water consumption and potential conservation measures within Utah's largest master-planned community: Daybreak. (2018).
85. *Luka Jovičić:* Spatial Data Harmonisation in Regional Context in Accordance with INSPIRE Implementing Rules. (2018).