

SDI FOR NATURE CONSERVATION: THE NATURE-SDI*plus* DATA HARMONISATION APPROACH

Marketa, HANZLOVA¹, Pavel, VANIS², Peter, FABIAN³

¹GISIG, Geographical Information Systems International Group, Via Piacenza 56, 161 38, Genova, Italy
m.hanzlova@gisig.it

²VUGTK v.v.i., Research Institute of Geodesy, Topography and Cartography, Zdiby 98, CZ-250 66, Praha, Czech Republic
Pavel.Vanis@vugtk.cz

³Faculty of Management Science and Informatics, University of Zilina, Univerzitna 8215/1, 010 26, Zilina, Slovakia
peter.fabian@fri.uniza.sk

Abstract

The new EU policies approach for protected sites management has enforced the link between nature conservation and geo-information, generating the need for interoperable, accessible and harmonised datasets for the EU.

NATURE-SDI*plus* is an eContent*plus* Best Practice Network aiming at SDI for nature conservation. It supports the implementation of the INSPIRE Directive (2007/2/EC) and takes into consideration following data themes:

- Protected sites (Annex I),
- Biogeographical regions (Annex III),
- Habitats and biotopes (Annex III),
- Species distribution (Annex III).

NATURE-SDI*plus* Network pursues, through state-of-the-art methodologies and good practice examples, to improve harmonisation of national datasets and to make them more accessible and exploitable. The main objective of the NATURE-SDI*plus* Network is to: involve new stakeholders; share data and best practices; improve and stimulate exploitation and the re-use of information on nature conservation.

The results obtained from the user needs survey and the analysis of the data usability and accessibility within a wide European context (from the national data providers) were used as a fundamental base for the development of the NATURE-SDI*plus* metadata profile and data models to be used in nature conservation.

The heterogeneity among the data provided is very high, each data provider having already taken different data model approaches, and often using different metadata profiles or, in some cases, not using any metadata at all. During the analysis of the available datasets within the project scope, it was found that relatively large part of each theme's datasets - 37% of Protected Sites, 31% of Species, 40% of Habitats, 27% of Biogeography - had no metadata associated.

Therefore, the harmonisation of the data represents one of the most important steps during the project.

In order to decide about the suitable methods and workflows for data harmonisation, it is important to distinguish data harmonisation at different levels (conceptual schema, logical schema or physical schema). To facilitate the process for non-GI-expert users, so called Mediation approach was developed.

This approach introduces a virtual database as middleware that provides high level abstractions of the integrated data and supplies a unified, global schema to the users.

To perform the transformation, it is necessary to follow several steps which help data providers to find the similarities and differences between their data model and metadata files, and use them in the process of remodelling their information.

Several tools provided by the NATURE-SDI*plus* partners and some external applications have been analysed and are available for the data providers. Several tools designated for the use by the data providers are offered at no cost base to support the harmonisation efforts within NATURE-SDI*plus* project. There are also some open source tools available for the task.

Data quality control procedures/protocols and testing of GIS data transformed according to the NATURE-SDI*plus* specification have been also considered and identified across the task.

The final objective is to provide INSPIRE compliant datasets and metadata from the different partners available on the NATURE-SDI*plus* Geoportal.

The NATURE-SDI*plus* Geoportal has been set up to make the information, provided by the project partners, accessible by the NATURE-SDI*plus* Infrastructure. The infrastructure includes the NATURE-SDI*plus* Geoportal, the French national geoportal, other national geoportals (future task), the thesaurus framework web services, the metadata editing tool (MDWeb editor), and the published partners' web services; it creates a real sets of distributed interoperable services.

In the project, the Czech Republic is represented by the Research Institute of Geodesy, Topography and Cartography (VÚGTK, v.v.i.) who serves as the national coordinator. There is a cooperation with governmental Agency for Nature Conservation and Landscape Protection of the Czech Republic (AOPK CR) and some of its bodies (management bodies of protected areas).

The Slovak Republic is in the project represented by three institutions – University of Zilina, governmental Agency for the Environment (Slovenská agentúra pre životné prostredie – SAŽP – www.sazp.sk) and State Conservancy of the Slovak Republic (Štátna ochrana prírody – ŠOP – www.soprsr.sk). While the first one serves as the national coordinator of the efforts within the project and part of the training framework set up in the project in order to disseminate the accumulated knowledge and best practice, the role of SOP is the one of the data provider and the role of SAŽP is more complex, as it serves as data provider, Web services provider and the actual data harmonisation performer at the national level

Keywords: SDI, INSPIRE, nature conservation, metadata and data harmonization, geoportal, protected sites, biogeographical regions, habitats and biotopes, species distribution

THE PROJECT INTRODUCTION

Recently, there has been a strong need linking nature conservation with geoinformation in Europe. The INSPIRE Directive (2007/2/EC) has been an answer to the necessity of interoperable, accessible and harmonised datasets. It also addresses EU Spatial Data Infrastructure (SDI) to support environmental policies. NATURE-SDI*plus* is an eContent*plus* Best Practice Network aiming at SDI for nature conservation.

The approach of the NATURE-SDI*plus* network is to share data experiences and good practices, to improve exploitation of datasets, and to re-use information on nature conservation.

NATURE-SDI*plus* contributes to the implementation of the INSPIRE Directive with targeted reference to the cluster of data themes of the INSPIRE Annexes, i.e. Protected sites (Annex I), Biogeographical regions (Annex III), Habitats and biotopes (Annex III), Species distribution (Annex III).

The mission of the NATURE-SDI*plus* project is to establish a network on geographical information for nature conservation, to stimulate the community of nature conservation stakeholders improving the harmonisation, the exploitation and the accessibility of their datasets. In such a view, NATURE-SDI*plus* aims to support the implementation of the INSPIRE Directive at EU level through the evaluation of common metadata profiles and data models for the addressed data themes, compliant with the INSPIRE provisions and data specifications.

NATURE-SDI*plus* analyses the usability and accessibility of data. The results of this analysis were used to develop the NATURE-SDI*plus* European metadata profile and data model for datasets on nature conservation.

The project defines a common multilingual and multicultural approach for a simpler and standardised access to spatial data.

A demonstration infrastructure, compliant with the INSPIRE principles and supported by web services, provides the data accessibility with a dedicated geoportal: the main gateway to available datasets and services.

The final target of NATURE-SDI*plus* is to establish a long-term sustainable network of stakeholders dealing with geo-information for nature conservation.

NATURE-SDI*plus* joined testing of the INSPIRE Specification for Annex I Protected Sites and provided a proposal for the Annex III Data Themes addressed by NATURE-SDI*plus*. In this perspective a proactive collaboration has been established with the INSPIRE Drafting team, as demonstrated by participation of two project partners on "INSPIRE: Data Specifications Thematic Working Groups for Annex II and III".

The above activity allowed to consolidate a profitable interaction among the technical partners, the partners playing the role of National Co-ordinators, and the data providers in the different countries, and as well to start networking actions and stakeholders' involvement, especially at a national level.

THE NATURE-SDI_{plus} METADATA AND DATA MODELS

The metadata has been defined considering the requirements of the project's target users, as identified during the surveys conveyed on user needs and datasets and data policies analysis. The metadata specification is tied up with the work on the multilingual and multicultural aspects necessary for indexing interoperable data, with parallel work undertaken on data model specification, and with the specification of harmonization implementation leading to comprehensible information for evaluating their quality and suitability for any intended use.

The profile has been constructed using requested metadata elements, INSPIRE metadata implementing rules and the definition of complementary metadata relating to the four themes dealt with by this project.

As a respond to the needs of the target community, the ISO 19115 standard has been complemented by a certain number of metadata elements, collectively called extended metadata.

To facilitate the implementation of this specification and to produce the metadata as a function of the theme which a dataset is concerned, it was proposed to break up this specification by data theme. This means that for every data theme, there is a corresponding metadata profile containing the entities and metadata elements specific to that theme. To this, a generic metadata profile was added that allows geographic web services to be documented. Thus, a total of five metadata profiles were proposed for implementing this specification. The following figure (Fig. 1) presents an overview of the contents and rationale behind the construction of each metadata profile.

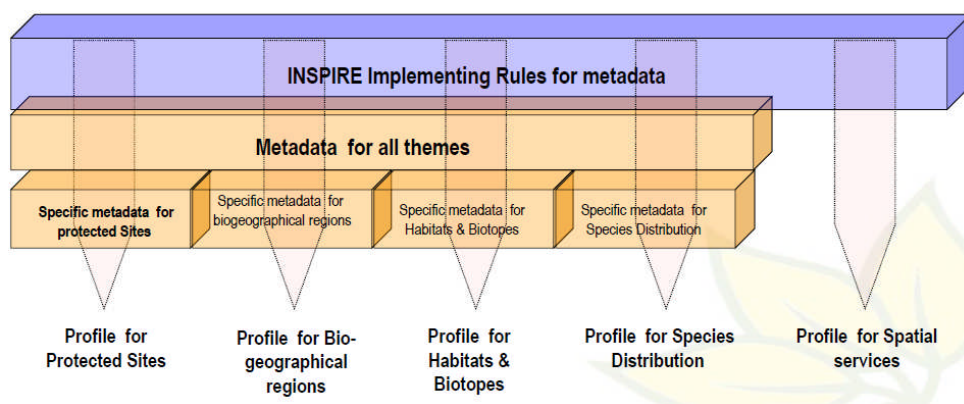


Fig. 1 Composition of metadata profiles for the NATURE-SDI_{plus} project

The elaboration of the data model followed as close as possible the methodology defined by the INSPIRE Drafting Team (DT) on Data Specifications and described in the following documents:

1. "D2.5: Generic Conceptual Model, Version 3.2" (DS-D2.5, 2009);
2. "D2.6: Methodology for the development of data specifications, Version 3.0" (DS-D2.6, 2008) and
3. "D2.7: Guidelines for the encoding of spatial data, Version 3.1" (DS-D2.7, 2009).

This methodology is based on a step-by-step approach in which user requirements are analysed and form the basis for the development of the use cases (in reality the selected use cases influence the development of the specs in the opposite direction as well), along with an analysis of the situation 'as-is' and of the gaps as related to those requirements.

The methodology was fine-tuned based on the experience of the development of the draft data specifications for theme 9 of annex I, i.e. protected sites.

One of the major outcomes of this process served as example for the work on the data model: “D2.8.1.9 – INSPIRE Data Specifications on Protected Sites – Guidelines” (D2.8.1.9, 2009).

Based on this experience it was decided to analyse and elaborate the potential use cases in more detail.

Information on the use cases was collected through dedicated interviews with users of nature conservation data and involved in related policy processes; but also through input from partners of the network and the analysis of documents.

The use cases were documented using a standard template used by the INSPIRE TWGs (Thematic Working Groups).

In a following step, the required spatial features and attributes were detected and described taking into account placeholders and candidate types, the voidable characteristics, possible code lists and enumerations, and stereotypes. UML was used to describe the application schema and data model (Enterprise Architect environment). The INSPIRE consolidated UML model (version October 2009) served as a starting point for the work. Freemind was used to prepare graphical overviews of the data model allowing non-expert users to read and understand the conceptual model (see Fig. 2).

Finally the models are delivered in the data exchange models corresponding to GML application schemas. An important aspect in this phase was the involvement and contribution of project partners for achieving broadly accepted data specifications. The data model is a key input to the INSPIRE data specification process and therefore was uploaded as reference material for the Thematic Working Group on Biogeographical Regions, Habitats and Biotopes and Species Distribution (TWG BR-HB-SD).

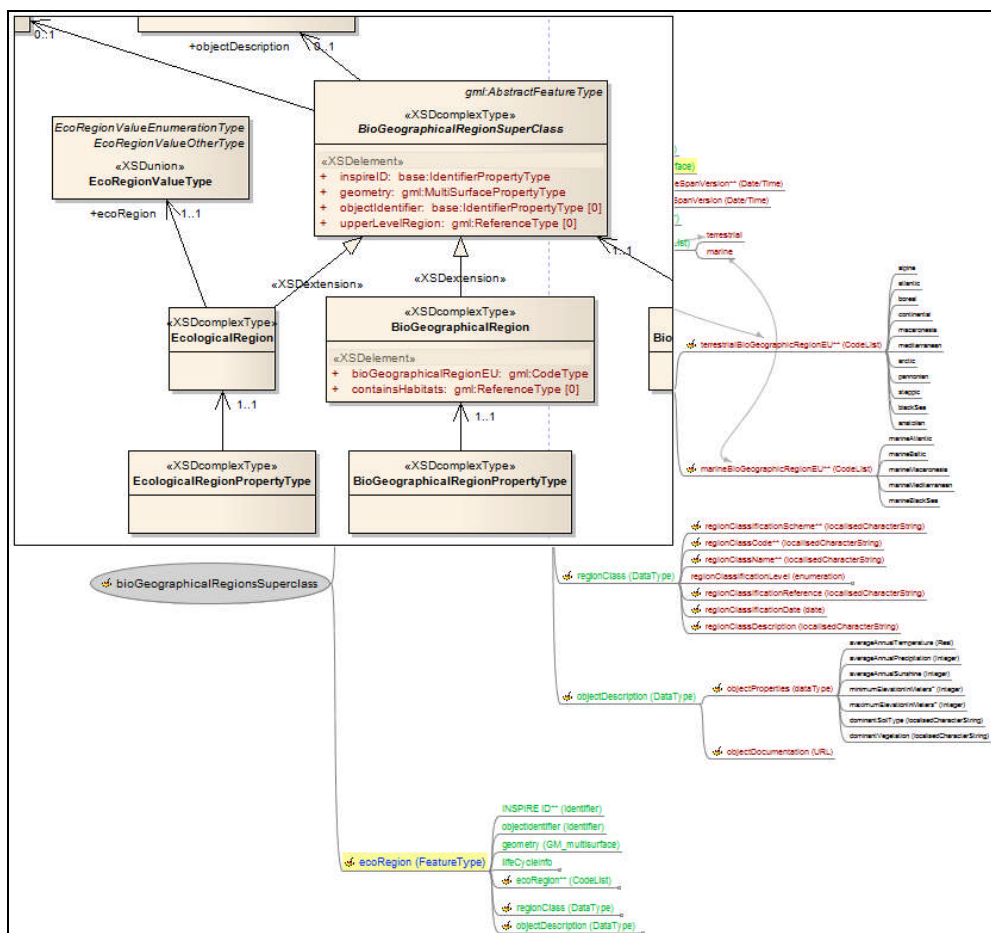


Fig. 2 Example of using UML (EA) schema format and Freemind format

METADATA AND DATA HARMONISATION PROCEDURES

In order to decide about the suitable methods and workflows for data harmonisation, it is important to distinguish data harmonisation on different levels (conceptual schema, logical schema or physical schema).

The Conceptual Model is a human oriented model of the world consisting of objects that a specific human considers relevant to a specific domain. It can be argued that there is only one reality, whereas there are almost as many conceptual models as there are people. The Logical Model is used to explore the domain concepts, and their relationships, and is often expressed as class models in UML. This stage is significant because it is the logical model which will be affected by the aggregation and degradation of data as it is harmonised. The Physical Model is used to design the internal schema of a database, depicting how the physical data is stored on a machine, i.e. the data tables, the data columns of those tables, and the relationships between the tables stored as flat files or databases.

To perform the transformation, it is necessary to follow several steps which help to find the similarities and differences between the data model and metadata files, and use them in the process of remodelling the information. The process can be summarised in four steps (Fig. 3):

1. to identify and to characterize the source and target information.
2. to compare source and target information using a matching table, documenting differences as essential input for the transformation and geoportals publication steps.
3. to choose the most suitable tools (ideally from a selection of tested and offered within the NATURE-SDIplus project) and to perform the transformation (remodelling).
4. to create the services and catalogue services for publishing the information under the NATURE-SDIplus Geoportals.

Several tools provided by the NATURE-SDIplus partners and some external applications have been analysed and are available for the data providers. Several tools designated for the use by the data providers are offered free of charge to support the harmonisation efforts within NATURE-SDIplus project. There are also some open source tools available for the task.

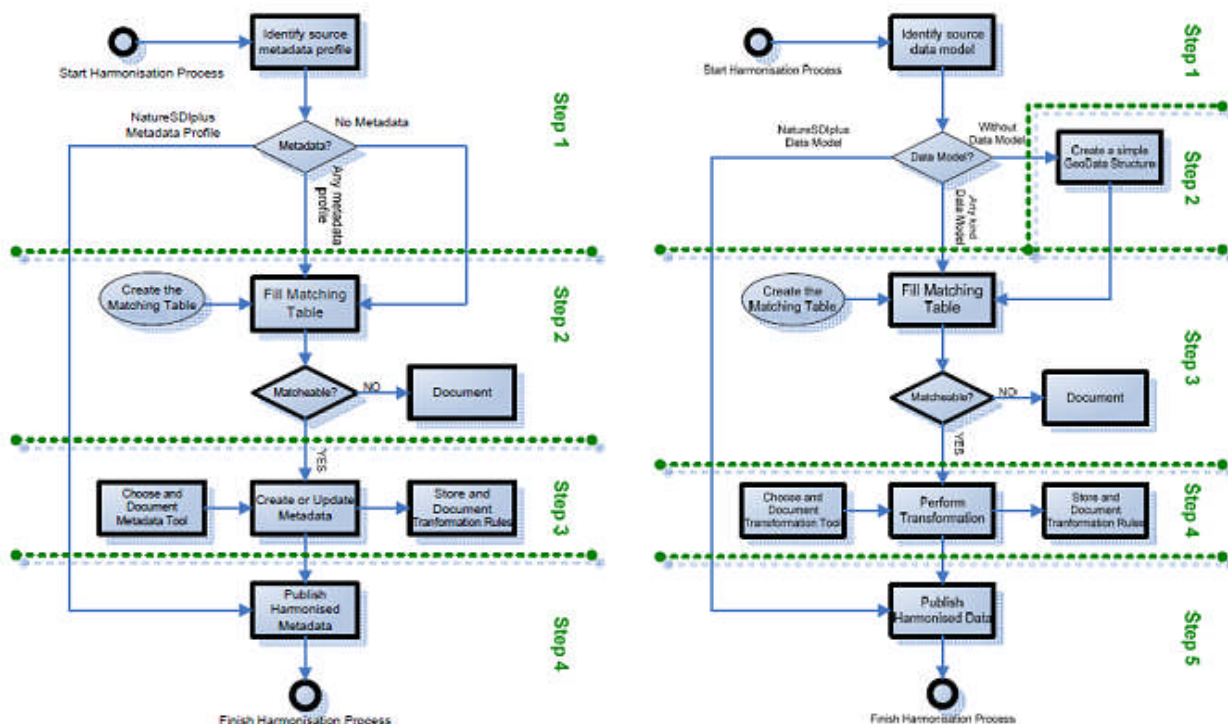


Fig. 3 Metadata (left) and data (right) Harmonisation Workflow

THE NATURE-SDIPLUS INFRASTRUCTURE

The NATURE-SDIplus infrastructure is a network of distributed entities that can be classified as follow:

- Data providers put, at disposal of the community, their own data related to Nature conservation harmonised according to the NATURE-SDIplus models and exposed in standard format and through standard services,
- Service providers are the technological partners which publish the national and central geoportals including the set of services and tools permitting to the users to discover, view and download the data distributed into the infrastructure.

From the functional point of view, the following picture (Fig. 4) shows the components of the NATURE-SDIplus infrastructure.

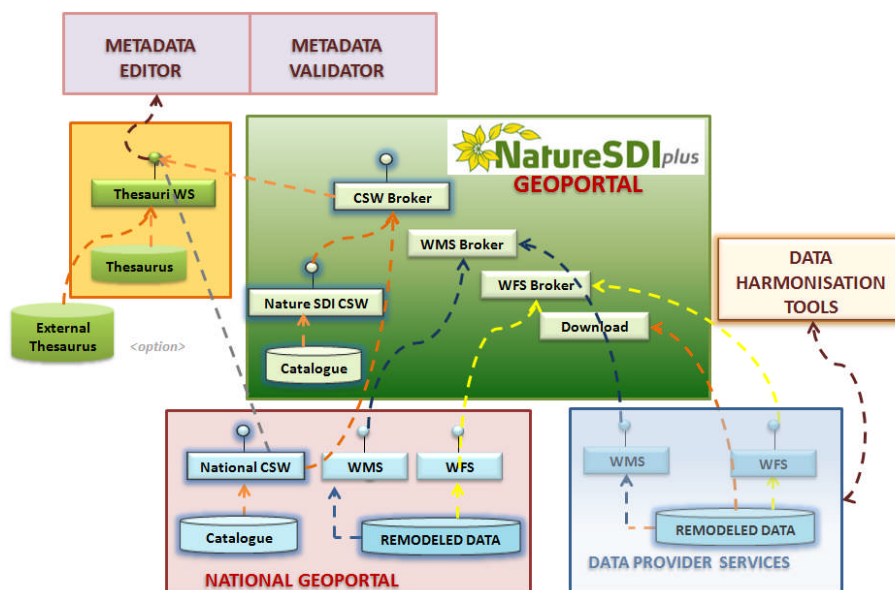


Fig. 4 NATURE-SDIplus Infrastructure Architecture

The *Metadata Editor Service* is a web application permitting the user to edit new metadata file according to the different profiles defined by the project. The metadata editor exploits the *Thesaurus Framework* facilities permitting to select the values of metadata keywords by the list of keywords collected and exposed by the thesauri framework services.

A *Metadata Validation Tool* has been deployed into the NATURE-SDIplus geoportal to check and validate the metadata produced by data provider before publishing it in a national or the central catalogue.

The geoportal administration tools include a *Catalogue Administration Tool* to manage and publish the metadata into the geoportal catalogue and to federate the national catalogue services into the geoportal catalogue client (desktop application).

The *Geoportal Web Client* is a browser interface including several facilities to search, display and download metadata and data exposed into the NATURE-SDIplus Infrastructure.

The *Geoportal Catalogue Client* permits to perform search to recover metadata related to the data published in the NATURE-SDIplus Infrastructure, looking into the central and federated catalogues of the entire infrastructure.

The following picture (Fig. 5) shows the main interface of the NATURE-SDIplus geo-portal.

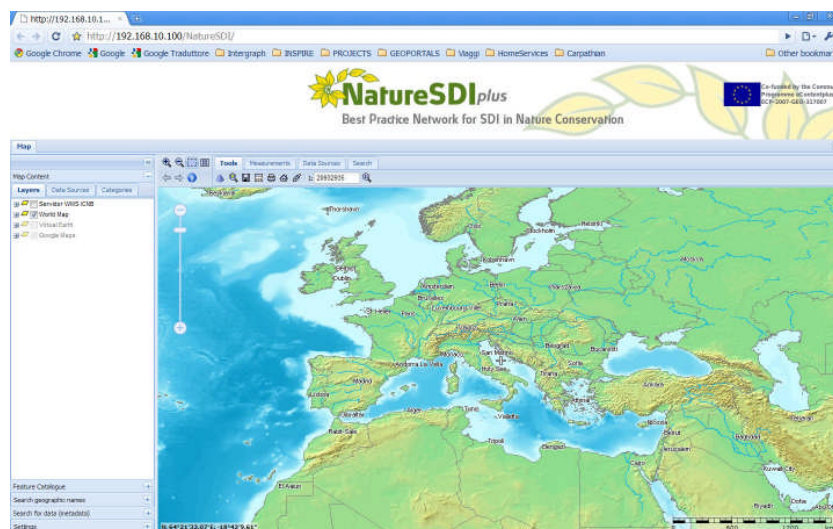


Fig. 5 NATURE-SDIplus geoportal interface

ADOPTED VALIDATION METHODOLOGY

The full NATURE-SDIplus validation methodology with respect to INSPIRE has been set-up, having identified what will be validated, against which requirement/standard and how the single items will be validated.

In terms of “What will be validated?”, it concerns harmonised datasets and metadata after the completion of the remodeling tasks.

After a detailed cross-check analysis between INSPIRE IRs and Technical Guidelines and EN/ISO and OGC standards (in order to answer “Against which requirement/standard?”), it has been decided to focus only on the applicable INSPIRE IRs and Technical Guidelines, as they also cover the other relevant standards.

Regarding the issue of “How the single items will be validated?”, the EN/ISO 19105 has been assumed as a reference context, but not as an additional requirement to conform to.

In this context, the first list of Test Suites has been drafted and adopted with the hierarchical structure, an approach of the EN/ISO 19105, based on “Test Modules” and “Test Cases”.

The “Test Case” components have been identified, i.e. Test case identifier, Test purpose, Test method, Reference to specific standards, Test type.

Regarding the “Test method”, templates for the Test Procedures and for the Test Reports have been drafted, in order to enable Data Providers to perform their self-validation.

Activity has started in the context of the data accessibility and usability testing, and the quality evaluation and datasets generalisation. The testing approach on data accessibility and usability was identified and considers the use cases accompanied user requirement survey, data model and metadata profile, geoportal requirements, respectively which will be further analysed and eventually adapted to the task needs. A comprehensive list of the quality issues, overlapping by INSPIRE documentation and NATURE-SDIplus metadata profile, was prepared.

CONCLUSIONS

The experience, that the NATURE-SDIplus partners gain along the project, has resulted in a wealth of knowledge and methods suitable for the harmonisation of GI in Europe and has underpinned the INSPIRE process in the field of nature conservation.

In this regard, the NATURE-SDIplus project establishes a Community of data and service providers and users in the addressed field, representative of the different European levels, from local to EU: the NATURE-

SDIplus Network, a first attempt towards the continuation and the sustainability of the initiative after the project conclusion (July 2011).

The European dimension of the project and its representativeness at European level are ensured by the growing numbers of Countries and Members represented in the Network (20 Countries, 44 members) and by the policy adopted by the project for an open participation to the network by relevant stakeholders in the field.

Moreover, a great effort has been achieved with INSPIRE TWG on Bio-geographical regions, Habitats and biotopes, and Species distribution (apart of the contribution to the Data Specification Testing for Protected Sites in the first year of the project). This collaboration with JRC is made by participation of two project partners (on behalf of the NATURE-SDIplus) in the activities of the INSPIRE TWG for the development of the Data Specifications.

Operationally, the next steps of the project will be devoted to the completion of the remodelling procedures resulting in the interoperable datasets, to the close collaboration with the INSPIRE TWG, and to the further actions in the validation phase.

NATIONAL EXPERIENCES AS GOOD PRACTICES

Czech Experience

In the Czech Republic the nature protection is in care of National Park Administrations, Agency for Nature Conservation and Landscape Protection of the Czech Republic (AOPK CR) (on the territory of large protected areas except National Parks) and Regional governments (out of the large protected areas). The datasets for nature protection are centralized at the AOPK CR.

For Nature-SDIplus project were studied data from all topics that are in the scope of the project. For protected areas were taken into account the databases of Small and Large Protected Areas that is stored at AOPK CR and also the protected areas layer from Fundamental Base of Geographic Data (ZABAGED®) managed by Survey office. The layer in ZABAGED® contains only fundamental information, i.e. the spatial extents of the protected areas, their name and type of protected areas. The original data from AOPK CR are more specific. There is also IUCN category, area, perimeter, some information touching the lineage of the record. From mandatory fields in the graphic file database the information about claiming the protected areas are missing. These information are in textual database that completes the graphic file.

The biogeographical regions are defined in the database of Biochoras. This database is more or less complete for the proposed schema. Only INSPIRE ID is missing. Coding of the bioregions and listing only the names of lowest level is a little problematic. The higher hierarchy levels must be found in a codebook.

The Biotopes database is in the stage of renovation. There are data for original data and updated data at AOPK CR. The data obtained for the project consist of a shapefile and textual database. The database for graphic part contains only fundamental information (ID, biotope code, date, formation group (rocks, forests, meadows etc.), and area) the more detailed information are in textual database file. There is different attitude to some items in the Czech data and the project proposal data structure. The quality of the biotope is evaluated from several points of view: The Czech database has different attributes for several issues, e.g. spatial and age structure, degradation, general evaluation, functionality evaluation etc. Also the impact of activities is evaluated in different way.

The species distribution is represented by Occurrence Database in the Czech Republic. It also consists of graphic file but without any database information and database textual file. From proposed mandatory fields the species conservatory status and monitoring assessment value are missing but further information identify the source of information, biotope and detailed description of localisation.

Slovak Experience

The Slovak Republic is in the project represented by three institutions – University of Zilina, governmental Agency for the Environment (Slovenská agentúra pre životné prostredie – SAŽP – www.sazp.sk) and State Conservancy of the Slovak Republic (Štátna ochrana prírody – ŠOP – www.sopsr.sk). While the first one serves as the national coordinator of the efforts within the project and part of the training framework set up in the project in order to disseminate the accumulated knowledge and best practice, the role of SOP is the one of the data provider and the role of SAŽP is more complex, as it serves as data provider, Web services provider and the actual data harmonisation performer at the national level.

ACKNOWLEDGMENT

The authors would like to thank the following Nature-SDIplus partners for their information contributions: Austrian Academy of Sciences (AT); TRACASA - Trabajos Catastrales (ES); Intergraph (IT, PL); VUGTK - Research Institute of Geodesy, Topography and Cartography (CZ); Epsilon International (IT, GR); Katholieke Universiteit Leuven (BE); University of Zilina and Slovakia Environment Agency (SK); University of Gavle (SE).

Nature-SDIplus is co-funded by the Community Programme eContentplus ECP-2007-GEO-317007.

REFERENCES

This paper was assembled using internal documents and reports conducting along the project life, i.e. the following project deliverables - “D3.2 Nature-SDIplus metadata profile (MetaData Specifications)”, “D3.4 Nature-SDIplus data model (final version)”, “D3.5 Specifications for data model implementation (Procedures for metadata profile and data model implementation)” - , the 4th progress report “D1.3.4 Periodic Report IV Semester”, and the second annual report “D1.4.2 Annual Report”.

D2.8.1.9 (2009), INSPIRE Thematic Working Group *Protected sites*: “D2.8.1.9 INSPIRE Data Specification on Protected sites – Guidelines”, 2009, 115p. [online document

http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_PS_v3.0.pdf]

DS-D2.5 (2009), INSPIRE Drafting Team "Data Specifications": DS-D2.5, Generic Conceptual Model, v3.2, 2009, 138p. [online document http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.5_v3.2.pdf]

DS-D2.6 (2008), INSPIRE Drafting Team "Data Specifications": DS-D2.6, Methodology for the development of data specifications, v3.0, 2008, 123p. [online document

http://inspire.jrc.ec.europa.eu/reports/ImplementingRules/DataSpecifications/D2.6_v3.0.pdf]

DS-D2.7 (2009), INSPIRE Drafting Team "Data Specifications: DS-D2.7, Guidelines for the encoding of spatial data, v3.1, 2009, 38p. [online document

http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/D2.7_v3.1.pdf]