Best Practice Network GS SOIL Promoting Access to European, Interoperable and INSPIRE Compliant Soil Information

Katharina Feiden, Fred Kruse, Tomáš Řezník, Petr Kubíček, Herbert Schentz, Einar Eberhardt, Rainer Baritz

Best Practice Network GS SOIL
promoting access to European, interoperable and
INSPIRE compliant soil information

Katharina Feiden[1], Fred Kruse[1], Tomas Reznik[2], Petr Kubicek[2], Herbert Schentz[3], Einar Eberhardt[4], Rainer Baritz[4]

[1]Coordination Center PortalU at the Lower Saxony Ministry of Environment and Climate Protection, Archivstr. 2, 30169 Hannover, Germany
[2]Masaryk University, Department of Geography, Laboratory on Geoinformatics and Cartography, Kotlářská 2, 61137 Brno, Czech Republic
[3]Umweltbundesamt GmbH, Spittelauer Lände 5, 1090 Vienna, Austria
[4]Federal Institute for Geosciences and Natural Resources, Information Systems Soil and Hydrogeology, Stilleweg 2, 30655 Hannover, Germany
gssoil@portalu.de

Abstract. INSPIRE provides the framework for the establishment of a European Spatial Data Infrastructure. The cross-border use and applicability of data requires that specific standards and rules are fulfilled by data providers. Such rules are currently being developed as data specifications. Soil as a theme in the INSPIRE annex III is included in this process, and was selected as the target theme for the EU best practice network GS SOIL „Assessment and strategic development of INSPIRE compliant Geodata-Services for European soil data". The project contributes to the harmonization and provision of interoperable soil geodata in Europe. The main deliverable of the project is the web portal http://gssoil-portal.eu/, which provides information, data management tools and links to data sources. Examples are the soil specific multilingual thesaurus, a metadata editor and catalogue service, provision of WMS and prototype WFS.

Keywords: INSPIRE, spatial data infrastructure, geoportal, soil data, metadata, data quality, thesaurus, controlled vocabulary, application schema, data specification.

Introduction

2   **Katharina Feiden1, Fred Kruse1, Tomas Reznik2, Petr Kubicek2,** Herbert Schentz3, Einar Eberhardt4, Rainer Baritz4

The network consortium comprises 34 partners from 18 EU member states. Project Coordinator is the Coordination Center PortalU at the Lower Saxony Ministry of Environment and Climate Protection (Germany). Overall 24 partners out of the consortium are soil data providers and will make the data available for the project. Hence, a complex and high quality data basis in a European context is assured. Beyond that, European institutions are also involved via the advisory board, as the European Environment Agency and the Joint Research Center of the EC.

The partners will establish and operate a network of services for spatial datasets and metadata. This network includes distributed services for data transformation, discovery, view and download. The final result of the project will be a central GS SOIL Portal, where European soil data from heterogeneous sources will be bundled. In order to ensure cross-border usability of the portal and related services, aspects of multilingualism and data interpretation will be considered thoroughly.

The project will extensively support the implementation of the INSPIRE requirements on basis of available experience in selected European countries and regions on different organizational levels. The planned results of the project are [3]:

− A consolidated soil-related theme catalogue and consolidated soil-related theme content-framework standards,
− An INSPIRE compatible metadata profile for spatial soil datasets, dataset series and services,
− Generic application schemes for soil information [12],
− A web portal (GS SOIL Portal) which provides access to all project soil data, including [8], [9], [10]
    − a view service which provides access to spatial soil data,
    − discovery and view of the INSPIRE conform metadata for the provided soil maps [5], [6],
    − interoperable spatial soil datasets (for exemplary soil products),
    − case studies on cross-boarder delivery of harmonised soil data access [7],
    − Best practise guidelines for
        − creating and maintaining metadata for soil database [12],
        − and for data harmonization.

The GS SOIL Portal

The GS SOIL community provides a centralized web access point for standardized, interoperable and INSPIRE compliant European soil information. In the GS SOIL Portal http://gssoil-portal.eu all soil related information from web pages, over databases to data catalogues will be made available and accessible [4]. Search results will be ranked and listed in shared result lists. Spatial soil data from OGC compatible Web Mapping Services (WMS) and Web Feature Services (WFS) will be visualized in a map viewer. For all tasks within the project the GS SOIL Portal will be used as a platform for an improved access to soil data. GS SOIL Portal has been built in an iterative cycle, adopting the relevant INSPIRE Implementing Rules (Network Services and related) and on the basis of the InGrid software designed for the German

Environment Information Portal (PortalU) [17]. Also general open tools and services will be provided for re-use by the project partners (data / service providers) and later technical integration of services and underlying geospatial data sets. Particular focus will be placed on mutual harvesting (CSW) with external systems. The current version of the GS SOIL Portal is already available in 13 project languages.
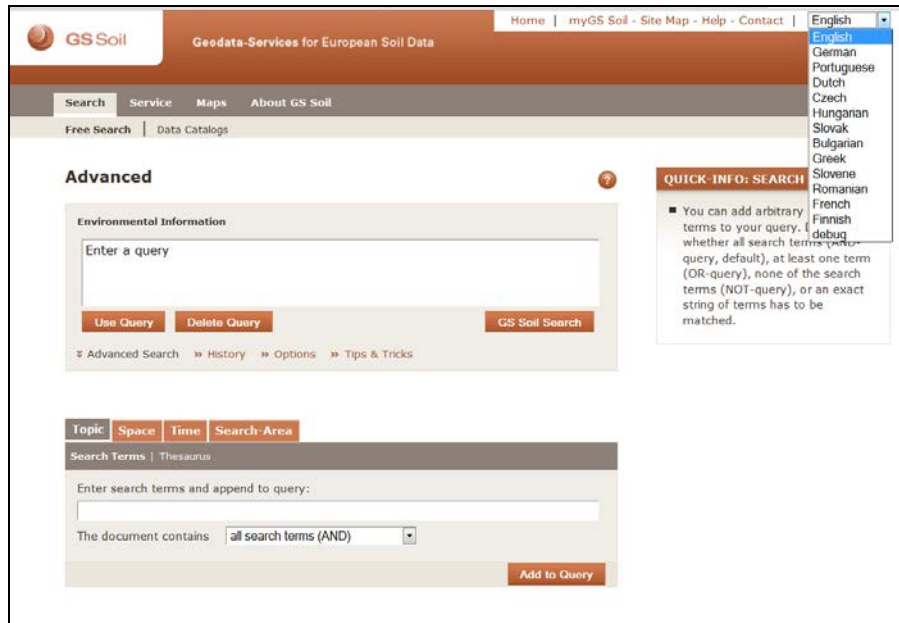

Fig. 1. The GS SOIL Portal (www.gssoil-portal.eu) and its advanced search function. The interface is available in 13 languages

The product harvesting, as well as the development of so-called content-framework standards (terminology, reference material and definitions to compare soil data), are based on a substantial search for existing soil data in European countries. The result has been implemented into a catalogue of soil data existing in Europe with a specific focus on products provided through web-services.

Data quality, data management and metadata

In GS SOIL two soil data relevant metadata profiles (the first one for datasets and dataset series while the second is intended for describing a service) compliant to INSPIRE metadata and ISO standards were developed in 2010. Subsequently, metadata proposed by the GS SOIL community was delivered as reference material to the Thematic Working Group for the sections dealing with metadata and data quality in the first version of INSPIRE Annex III Data Specification on soil. INSPIRE Data Specifications contain requirements (that are legally mandated) and recommendations

4   **Katharina Feiden1, Fred Kruse1, Tomas Reznik2, Petr Kubicek2,** Herbert Schentz3, Einar Eberhardt4, Rainer Baritz4

[13], [14]. On the other hand, INSPIRE Data Specifications may be hardly readable by data providers that are not aware of all underlying standards. Thus, it complicated to get an overall picture. GS SOIL Best Practice Guidelines for creating and maintaining metadata for soil databases are intended to overcome this insufficiency.

Proposed Best Practice Guidelines are covering both – metadata and data quality aspects. This approach consists of several steps, including

- a review of the GS SOIL metadata profile with respect to newly published INSPIRE documents (e.g. Commission Regulation 1089/2010 on interoperability of spatial data sets and services or new version 1.2 of Technical Guidelines on INSPIRE metadata),
- connections between dataset metadata and feature-level metadata through a feature catalogue,
- user-friendly guidelines for data providers who would like to describe their soil-related databases,
- the GS SOIL metadata profile applied on two real soil datasets as examples,
- guidelines on how to manage soil metadata in two or more languages,
- brief information about the development of the soil specific multilingual thesaurus (which is in detail written in the following section)
- and examples on usage of soil metadata within INSPIRE, i.e. spatial data infrastructure framework (e.g. how to get from dataset metadata to view/download service or dataset itself as well as description of related aspects).

Best Practice Guidelines are established as a recommended set of steps to obtain the right value that should be reported within INSPIRE compliant metadata. All steps are written in user-friendly way while an end user does not have to be aware of underlying standardization documents to successfully create INSPIRE valid metadata with GS SOIL profile. The consortium partners decide for a selected and open tool to integrate the metadata into the GS SOIL portal (e.g. GeoNetwork, InGrid Editor etc.). These open tools were adopted to the defined metadata structure and are INSPIRE compliant.

Another specific part of Best Practice Guidelines is related to data quality issues that are currently divided into two main streams. The first is focused on analyzing the existing quality specifications and measures of compliancy specified within the data themes of Annex I as well as data quality evaluation according to the ISO standards (like ISO 19157, 15158, 19113, 19114). The second activity is focused on the discussion paper "Data quality in INSPIRE: from requirements to metadata" (from October 2010). This document [15], although only on the level of discussion paper, could seriously change the demands and also the viewpoints for defining the data quality for all INSPIRE spatial data themes, including soils.

Metadata element name	Data Quality – Positional accuracy – Absolute or external accuracy
Definition	Closeness of reported coordinate values to values accepted as being true.
ISO 19115 number and name	18. dataQualityInfo
ISO/TS 19139 path	dataQualityInfo
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	117. DQ_AbsoluteExternalPositionalAccuracy
Domain	Lines 100-107 from ISO 19115
Implementing guidelines	This measure is for expressing of positional accuracy of point data (boreholes) only. Reported should be only one number expressing the positional accuracy according to the number of measured dimensions of soil dataset. Explicit formulas are defined in section 6.2. This result value is usually designated as Root Mean Square Error (RMSE) within GIS software. This is the only result to be reported within this metadata element. Following steps are recommended: 1. Recalculate resulting RMSE into metres, in case when resulting RMSE is in different units (for example GIS has calculated RMSE as 6 cm, i.e. recalculated RMSE is 0.06 m). 2. Write recalculated RMSE into the value metadata sub-element as a real number with a precision of 2 decimals (e.g. 1.85).Usage of decimal point is recommended. 3. Write "m" into the valueType metadata subelement
Example	value: **2.00** valueType: **m**

Example XML encoding	`<gmd:MD_Metadata ..`
	 `..`
	    `<gmd:dataQualityInfo>`
	    `..`
	       `<gmd:report>`
	        `<gmd:DQ_AbsoluteExternalPositionalAccuracy>`
	         `<gmd:result>`
	 `[…]`

Fig. 2. Example of (data quality) metadata element structure as a part of GS SOIL Best Practice Guidelines for creating and maintaining metadata for soil databases

Soil specific multilingual thesaurus

Sharing metadata and/or data generally leads to the need of common semantics, a common set of concepts and a common controlled vocabulary. This is a fact well-known long before the existence of IT and IT networks, and has lead to controlled lists of physical units, chemical elements, species lists or even to complete vocabularies like the medical or pharmaceutics vocabularies. Today thesauri and/or ontologies have become state of the art within communities and international projects even when the participants do not want to share data but "just" want to exchange their

6   **Katharina Feiden1, Fred Kruse1, Tomas Reznik2, Petr Kubicek2,** Herbert Schentz3, Einar Eberhardt4, Rainer Baritz4

knowledge. Public institutions who collect data/information/knowledge from various parties meet this well-known need establishing controlled vocabulary and making it accessible for their data providers. The European Environmental Agency e.g. uses and publishes GEMET, the GEneral Multilingual Environmental Thesaurus, Which is provided in 26 languages, as multilingualism is a need for any European collaboration, not only in the environmental sector. Referencing concepts of GEMET [12] is requested by the EU INSPIRE directive.

As GS SOIL aims to be INSPIRE compliant, GEMET has to be at least a part of the controlled vocabulary of that project. But GEMET represents a very general environmental vocabulary and needs more specific extensions for special domains, such as water, protected areas, air and – soil.

The GS SOIL Portal has a reference to a service for GEMET but for the reasons explained above, in addition to the preliminarily planned tasks, the GS SOIL community decided to establish a soil specific multilingual thesaurus in addition to the preliminarily planned tasks. This thesaurus shall be created as an extension of GEMET and shall contain the soil vocabulary of ISO11074 [16], the concepts of WRB [1] and concepts created and/or defined during the GS SOIL project.

Concerning the architectural aspects the thesaurus has to be linkable to the GS SOIL Portal and must be free and easily accessible. In order to make referencing GEMET as easy to handle as possible, it was decided to make the soil specific thesaurus not only semantically but also syntactically interoperable with this controlled vocabulary demanded by INSPIRE.

The architectural decisions and the willingness to stick to widely accepted standards lead to the decision to use SKOS/ RDF [20], a W3C standard, as formal language. Therefore an SKOS/RDF compatible thesaurus editor, which allows collaborative work, is chosen as tool for the establishment. Several tools for that standard are available, but so far the only freeware tool is iQvoc [19].

One benefit of the chosen formal language and tool is that content can also be presented as linked data, thus allowing interconnections with other controlled vocabulary, which will help to establish cross – domain semantic interoperability [18]. Moreover even links to concepts, which are defined for data and metadata transfer, could be established as soon as those models are transformed from UML to OWL.

The guidance and rules for the establishment of the multilingual content is set up in March and the continuous maintenance beyond the end of the project will be topic of the planning of GS SOIL.

Application schema, data harmonization and interoperability

Data can be exchanged in many different formats and structures. To ensure that the receiver does interpret the data in a proper way, a common or standardised application schema for soil data is needed. Because INSPIRE is determined to use already existing standards, GS SOIL adopted the ideas of the upcoming ISO 28258 Digital Exchange of Soil-related Data (i.e. the working draft available end of July 2010), developed it further and produced a practical implementation in two XML schema

files and a set of rules how data providers set up a further, extended XML schema definition and define properties and produce data files in an interoperable way.

ISO 28258 is basically an extension of ISO 19156 Observations and measurements for soil data. The core of the application schema is the feature catalogue. It lists all objects that can be described in the soil universe. The feature types and their relationships have been modeled using the Universal Modelling Language (UML). To ensure the interoperability of the data, this feature catalogue is complete and cannot be extended. The application schema is nonetheless generic in that it leaves the properties that can be described for each feature with the user of the standard (i.e. the data providers).
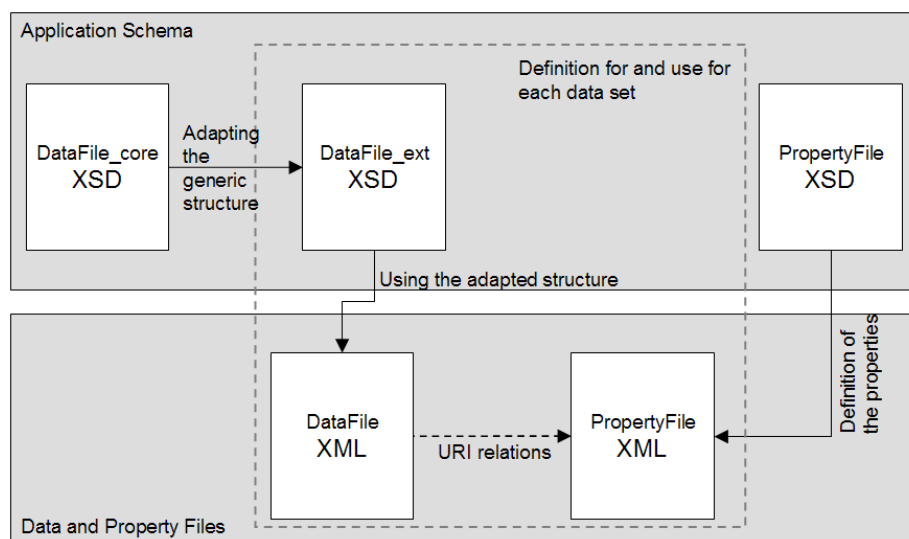

Fig. 3. Implementation of the application schema using XML and XML schema definitions

The application schema has been implemented in XML (Fig. 3). The feature types are defined in the DataFile_core.xsd. In order to have a mechanism to check XML data files against an xsd, the data file definition has to be extended for each use with the properties that shall be exchanged in the data file. In an additional PropertyFile.xml (that itself can be checked against the PropertyFile.xsd provided with the application schema) further property definitions are given that cannot be put into the DataFile_ext.xsd due to the impossibility to define more than the codes itself in xsd code list definitions. For interoperability, often further information is needed, e.g. the class boundaries of classified numerical values, or the link to the multilingual thesaurus.

The application schema and its implementation serve as reference material for the INSPIRE process of data specification development (e.g. proposal for a soil metadata profile). The GS SOIL amendments to the former ISO 28258 Working Draft were fed

back into the ISO process. The application schema is tested with data from the GS SOIL data providing partners using a hands-on testing procedure description. The GS SOIL SoilML is compared with the currently developed INSPIRE application schema for soil data.

Interoperability includes harmonization to varying degrees, depending on domain-specific agreements. A closer look to the descriptions and scope of the INSPIRE Annex Themes shows that soil data include a wide variety of data content. The historic development of soil data, especially high-resolution maps, has caused substantial regional differences in resolution, terminology, and taxonomy. The feasibility to harmonize the content of such products has not yet been systematically investigated. Independently of the agreements under INSPIRE, the array of possible harmonization methodologies is being investigated in detail in GS SOIL test cases. This includes high and low-resolution soil maps, as well as thematic maps, data from soil inventories such as long-term field experiments, and national soil monitoring. Thematic maps are applications of basic soil maps, for example susceptibility maps, very often combined with ancillary data such as climate and land cover. In the future, web processing services will allow such applications on the fly, using data and methodologies shared among distributed information systems. Given these frame conditions and outlook, semantic aspects of harmonization of soil information play an important role to the soil data user community.

Outlook

Until end of May 2012 the best practice network will finalize the self defined work plan. Technically, the GS SOIL Portal is already online and will be further improved. The soil thesaurus with the set of minimum terms will be realized. All data providers in the project and the extended network will make the interoperable data accessible and will also create metadata..

The Best Practice Guidelines for creating and maintaining metadata for soil databases were prepared to answer opened questions that data providers may have. At the same time, it serves as the "cookbook" where the process of metadata creation and management is described step by step to get to INSPIRE valid metadata with added relevant information on soil data. These Guidelines may be used in other application fields as well and in the full scope may be used for metadata according to the Commission Regulation No 1205/2008 (on INSPIRE metadata). Second benefit represents explicit guidelines on data quality including quantitative methodologies on specific measurements that are suitable for soil data. As a result, metadata and data quality information is well defined and contains values that are comparable.

In terms of harmonization GS SOIL first focus on the content framework. Concepts are being exchanged (for example, about map stratification, or about the definition of soil map units), and definitions agreed upon. This will allow focusing test cases on a common vocabulary. Recommendations will be provided as "Best practice guidelines for developing a content framework for interoperable soil data in Europe". Because the World Reference Base for Soil Resources plays an important

role in international data exchange, many GS SOIL partners will investigate the possibility of applying this reference system for identifying the taxonomic units of their products as the basic classification system. The next steps will be concentrated on comparing and adapting existing data sets across borders. Further harmonization steps will then look at the soil maps in detail. The geometric data content will be analysed against independent GIS-based criteria for comparing the structure and content of maps. The GS SOIL test cases will thus allow estimating the possibilities and effort to fully harmonize soil data. The results will be provided as "Data Harmonization Best Practice Guidelines".

References

1. Deckers, J.A., Driessen, P.M., Nachtergaele,. f.O., Spaargaren, O.C. World References Base for Soil Resources, Wageningen University and Research Center Publications. (2002)
2. European Environment Agency. EIONET. About GEMET. http://www.eionet.europa.eu/gemet/about
3. Feiden, K. et. al.: Progress of the transnational cooperation in building up a SDI for European soil data (eContentplus-project GS SOIL). INSPIRE conference 2010 – a framework for cooperation, 22-25. June 2010, Krakow (2010)
4. Feiden, K., Kruse, F., Epitropou, V., Karatzas, K.: The GS SOIL portal prototype and its integrated network. 24th International Conference on Informatics for Environmental Protection in cooperation with Intergeo 2010. 6.-8.10.2010 EnviroInfo Köln/Bonn 2010 / in: Greve, K. (Ed.): Integration of environmental information in Europe, pages 420-428. (2010)
6. GS SOIL consortium: D3.1 Metadata profile for soil geographic datasets and dataset series (2010)
7. GS SOIL consortium: D3.2 Metadata profile for soil geographic data services (2010)
8. GS SOIL consortium: D4.1 Theme specific test suite for developing data specifications (2010)
9. GS SOIL consortium: D5.1 Design specifications of the GS SOIL Portal and its network (2010)
10. GS SOIL consortium: D5.2 GS SOIL Portal (Prototype) (2010)
11. GS SOIL consortium: D5.3 First set of open tools and services (2010)
12. GS SOIL consortium: D3.4 Final best practice guidelines for Creating and Maintaining Metadata for Soil Database (2011)
13. GS SOIL consortium: D4.2 Generic application schemes for soil information (2011)
14. INSPIRE DT [Drafting Team] Data Specification. Definition of Annex Themes and Scope. Deliverable D2.3. Version 3.0. (2008a)

15. INSPIRE DT [Drafting Team] Data Specification. Methodology for the development of data specifications. Deliverable D2.6. Version 2.2. (2008b)
16. INSPIRE DT [Drafting Team] Data Specification. Data quality in INSPIRE: from requirements to metadata, discussion paper of the European Commission. (2010)
17. International Organization for Standardization. Soil Quality Vocabulary. ISO 11074:2005 http://www.iso.org/iso/catalogue_detail.htm?csnumber=38529
18. Kruse, F., Konstantinidis, S.:   InGrid® – eine Software zum Aufbau von Umweltinformationssystemen. – AGIT 2010, Salzburg, 07.-09.07.2010. (2010)
19. Maria Rüther, Joachim Fock, Thomas Bandholtz, Till Schulte-Coerne, Linked Environment Data. Workshop des Arbeitskreises „Umweltinformationssysteme" der Fachgruppe „Informatik im Umweltschutz",
20. Thomas Bandholtz, Till Schulte-Coerne, Robert Glaser, Joachim Fock, Tim Keller: iQvoc – Open Source SKOS(XL) Maintenance and Publishing Tool 6th Workshop on Scripting and Development for the Semantic Web SWSW2010, Heraklion, 31st May (2010)
21. W3C: SKOS Simple Knowledge Organization System. http://www.w3.org/2004/02/skos/