

Sharing European soil information - Best practice contribution to INSPIRE –

Katharina Feiden¹, Emil Fulajtár², Endre Dobos³, Herbert Schentz⁴, Einar Eberhardt⁵,
Rainer Baritz⁵, Carlos Figueiredo⁶, Sigbert Huber⁴, Hermann Klug⁷

Abstract

The GS SOIL best practice network comprises 34 project partners out of 18 European member states. It receives funding by the European Commission and the first two years of GS SOIL, starting in June 2009, were dedicated to the survey and the analysis of data availability and accessibility, the activities towards the GS SOIL metadata profile, the data model and the GS SOIL portal. The aim of the paper is to present the state of the art of the GS SOIL project and the ongoing activities to support the INSPIRE implementation in the soil domain, especially to the development of data specification for the annex III theme “soil (Feiden 2010A; see also GS SOIL Deliverables).

1. Introduction

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 European 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 outcome serves as reference material for the INSPIRE process of data specification development (e.g. proposal for a soil metadata profile). GS SOIL has also substantially contributed to the ISO project on soil data exchange (ISO 28258) by developing an operational set of basic data exchange files, and by implementing a consistent relation to OGC Observations and Measurements.

On the basis of a representative set of test cases throughout Europe, GS SOIL develops good practice guidance for data holders to provide INSPIRE-compliant soil data. This does not only focus on data exchange, but also on the various possibilities and methods to harmonize soil property information and legends. The main deliverable of the project is the web portal <http://gssoil-portal.eu/>, which provides information about soils, data management tools and links to data sources. Examples are the soil specific multilingual thesaurus, a metadata editor and catalogue service, provision of Web Mapping Services (WMS) and prototype Web Feature Services (WFS). The product harvesting, as well as the development of so-

¹ Coordination Center PortalU, Lower Saxony Ministry of Environment and Climate Protection, Archivstr. 2, 30169 Hannover, Germany

² Soil Science and Conservation Research Institute, Gagarinova 10, 82713 Bratislava, Slovakia

³ University of Miskolc, Miskolci Egyetem, Miskolc-Egyetemváros 3515, Hungary

⁴ Environment Agency, Spittelauer Lände 5, 1090 Vienna, Austria

⁵ Federal Institute for Geosciences and Natural Resources, Information Systems Soil and Hydrogeology, Stillweg 2, 30655 Hannover, Germany

⁶ EDISOFT s.a., Rua Quinta dos Medronheiros, Apt. 382, Lazarim, Caparica 2826-801, Portugal

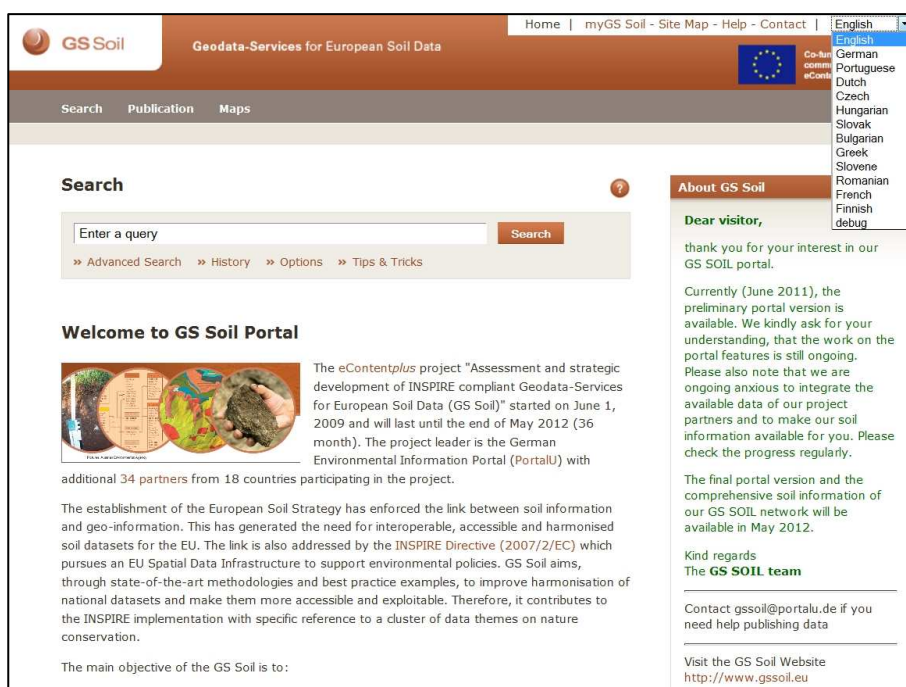
⁷ Paris Lodron University of Salzburg, Schillerstr. 30, Building 15, 5020 Salzburg, Austria

called content-framework standards (terminology, reference material and definitions to compare soil data), are based on a substantial collection of 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. (Feiden et al. 2011)

2. Sharing soil information via the GS SOIL Portal

The GS SOIL community provides a centralized web access point for standardized, interoperable and INSPIRE compliant European soil information, provided by the GS SOIL partners and the extended network. 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. 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). Technically, the portal is based on the software InGrid of the German Environmental Information Portal "PortalU". In addition, 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.



The GS SOIL portal entry page

Especially metadata and/or data generally leads to the need of common semantics, a common set of concepts and a common controlled vocabulary. Thesauri and/or Ontologies have become state of the art within communities and projects, which can be seen in the fact that the European Environmental Agency

uses and publishes GEMET, the GEneral Multilingual Environmental Thesaurus. Multilingualism is a need for any European collaboration, not only in the environmental sector.

GEMET represents a very general environmental vocabulary and needs more specific extensions for special domains, such as water, protected areas, air and – soil. Therefore, in addition to the preliminarily planned tasks, the GS Soil community decided to establish a soil specific multilingual thesaurus. This thesaurus shall be created as an extension of GEMET and shall contain the soil vocabulary of ISO11074, the concepts of WRB 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. It further has to be syntactically and semantically interoperable with GEMET since referencing to GEMET concepts is demanded within the INSPIRE implementation rules.

3. Content framework standards

One of the aims of the GS SOIL consortium is the documentation of the best practice content standards recommended to achieve harmonised large scale spatial soils and non-spatial property (semantic) data. This was split into two tasks: the Definition of Content-Framework Standards and the set up of a relevant Best Practice Guideline. The result is a review and analysis of the specific content requirements and characteristics that a harmonised distributed framework would possess as deriving from each soils information content provider. For spatial datasets this will entail consideration of the means by which the FAO World Reference Base (WRB) soil classification system can be adopted. Issues of scale and soil map unit geometric complexity should be assessed. For non-spatial property data, the key soil related parameters required by INSPIRE should be stated formally and compared to the offerings from the targeted data providers.

The works on mainly covers the following items:

- 1) Review of the content framework for existing soil data in European countries including common soil property data (analytical and field characteristics), soil profile data and site soil-related data as well as existing soil classifications (national classifications) used in European countries,
- 2) existing approaches of data gaps filling (data interpretation, spatial extrapolation, pedotransfer functions, etc.),
- 3) issues related to soil data harmonization (existing standards, guidelines for soil profile descriptions, soil survey and mapping guidelines) and
- 4) issues related to interoperability of soil data (semantic and technical).

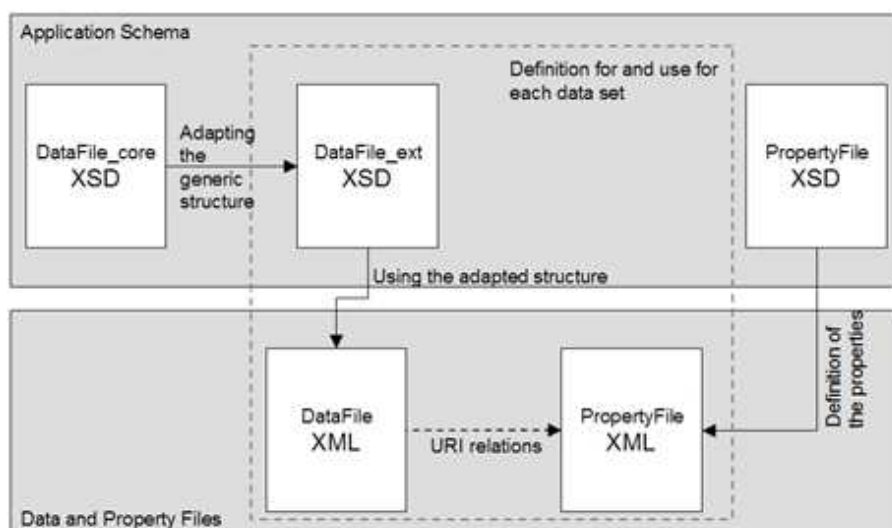
On the basis of these analysis the best practice guideline contains relevant examples for the identification of needed level of interoperability and required content framework at European scale, the preparation of best practice for data harmonisation and the issues related to soil cartography and adoption of WRB classification.

4. Interoperability and INSPIRE compliancy

The GS SOIL geoportal is a demonstration infrastructure that is compliant with INSPIRE principles and enhances the accessibility and exploitation of available digital soil data, as data sets / services and metadata, that was made available in the GS SOIL network. This technical solution of a geoportal only gains value if sufficient content is provided. The GS SOIL network aims with its activities to establish best practice in data specification development. These activities regarding an INSPIRE compliant GS SOIL data specification lead to the definition of GS SOIL specific metadata profiles for soil geographic datasets / dataset series and data services. They were developed following the INSPIRE IR for metadata, other international, and national standards (like the ISO 19115 and ISO 19119), and the needs of the data users. The focus was

also on data quality that is necessary for the soil spatial data theme. The results are directly defined in the soil metadata profiles. Additionally, recommendations were defined since this second part is not compliant to the existing ISO structure. Metadata structure developed in this project contains an example of the XML encoding soil specific resource.

Interoperability in GS SOIL is broadly defined as the ability of two or more autonomous entities to communicate and co-operate among themselves in a meaningful way despite differences in language, context or content (INSPIRE DT [Drafting Team] Data Specification). Technically, interoperability is the capability of a product to interact and function with other products without any access or implementation restrictions, and without immediate human intervention. This requires that specified data formats such as XML are maintained so that data sets can be exchanged and communicated (syntactic interoperability). Besides the simple exchange of information, it must be possible to interpret exchanged information meaningfully and accurately so that useful results are received by users (semantic interoperability). This requires that data sets conform to a common information exchange reference model, represented as a specific application schema and that the meaning of the contents is defined in common concepts and reference lists. The domain community has to agree on a common schema and on common concepts with which those schemata are filled. The data provider has to map his data to the concepts and fill in the schemata.



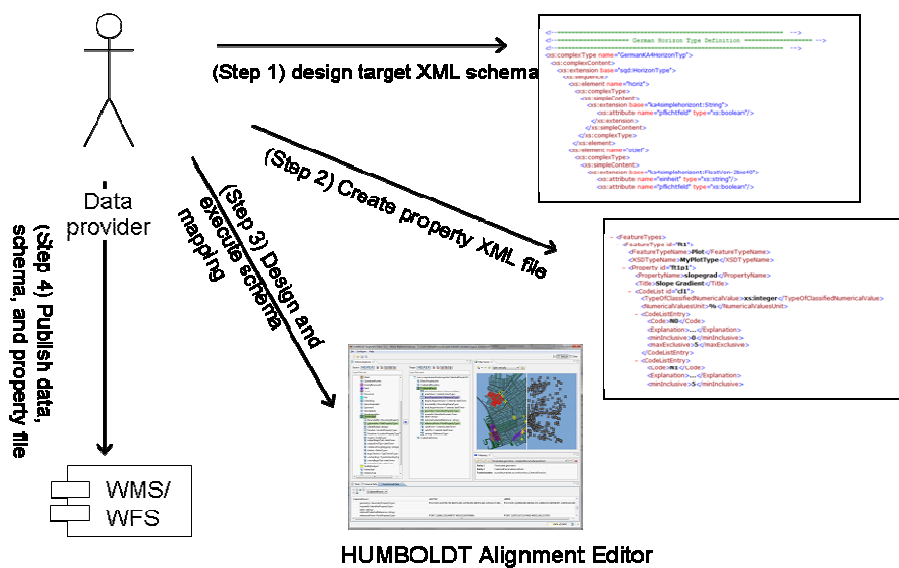
The GS SOIL application schema

The GS SOIL network developed an application schema for soil data, based on existing ideas of ISO 28258 draft for digital exchange of soil-related data. The schema allows data providers to define XML data structures for their soil data in a way that data receivers can immediately use and interpret the data in a meaningful and scientifically sound way. Soil data have been modelled using the Universal Modelling Language (UML). In order to make this schema usable, XML schema definitions (XSD) and rules how to apply them are presented. The application schema has been tested in a number of test cases in Europe, which are partly even cross-boundary. Besides the presentation of the test results, the application schema will be compared with the INSPIRE data specifications soil, version 2, which will be available for testing in June 2011. GS SOIL is registered as Spatial Data Interest Community (SDIC) and supports the INSPIRE development process in the theme soil. Significant effort to harmonise project activities with INSPIRE development was taken from the early start of the project and will be regarded in the “sustainability long term operational plan” of the project.

5. Data transformation

For the GS Soil consortium, the work with the harmonization approaches started in early 2011 with the identification of examples for source datasets to be transformed. The first step of the process is concerned with schema mapping. The GS soil common target schema for soil data will be based on the Feature Catalogue and Application Schema SoilML, defined in ISO CD 25258. In the initial setting, the source datasets will be transformed with the HUMBOLDT Alignment Editor (HALE), a tool for creating and executing schema mappings.

An initial view on the schema mapping and transformation procedure in GS Soil can be described by the following steps. Please note that the steps currently only include the schema mapping process and omit all other steps that might be required for soil data harmonization such as those identified in INSPIRE, e.g., coordinate referencing, portrayal or data quality.



GS SOIL uses the HUMBOLDT transformation alignment editor (Fitzner, 2010)

Step 1: This step consists of the design of the target schema. The individual FeatureTypes of the target schema are required to have a taxonomic subtype relationship to Feature Types of SoilML as defined in the relevant ISO. For example, a `GermanHorizonType` can be defined as a subtype of the generic `HorizonType` of SoilML, inheriting all of its attributes and adding additional ones.

Step 2: The attributes specific to the newly added Feature Types (such as `GermanHorizonType`) need to be defined in a property XML file. The file follows (i.e. needs to validate against) a XML schema file of SoilML as defined in ISO.

Step 3: Publishing datasets in the newly defined schema requires the manual definition of the correspondence between source and target types (i.e. a process called schema mapping) and the execution of the mapping on instance data. HALE allows domain experts to create mappings between two different schema and to execute them on instance data.

Step 4: After transforming the instance datasets, they need to be published in conjunction with the target schema and the property file. For example, the target schema might be published as response to a DescribeFeatureRequest on a WFS, while the instance data can be retrieved via GetFeature.

Based on the above steps and the experiences gained with the source datasets, a GS Soil internal best practice guideline on “soil data harmonization” will be compiled, including the topics:

- Where to start? What are the requirements on source datasets in order to perform the harmonization as proposed?
- What tools do we recommend (e.g., in addition to HALE) for performing the individual steps described above?
- How do we recommend / plan to deal with other data transformation and harmonization issues such as coordinate referencing, portrayal, data quality or consistency?

Acknowledgement



The project GS SOIL “Assessment and strategic development of INSPIRE compliant geodata services for European soil data” is co-funded by the European Community programme eContentplus. In the eContentplus programme, GS SOIL is defined as a Best Practice Network for Geographic Information.

Facts:

Project partners: 34 out of 18 EU member states

Duration: 06/2009 – 05/2012

Budget: ca. 5.1 Mio Euro

Funding: eContentplus 2008

The work presented in this paper is the joint effort of the GS SOIL consortium. The authors are representatives of the project coordinator and members of the work package on the technical establishment of the GS SOIL portal. They are solely responsible for the content. The paper does not represent the opinion of the European Community and the European Community is not responsible for any use that might be made of information contained therein.

The GS SOIL Portal www.gssoil-portal.eu

The project website www.gssoil.eu

Bibliography

- Feiden, K., Kruse, F., Reznik, T., Kubicek, P., Schentz, H., Eberhardt, E., Baritz, R.: Best Practice Network GS SOIL - Promoting Access to European, Interoperable and INSPIRE Compliant Soil Information. In: Hrebicek, Jiri; Schimak, Gerald; Denzer, Ralf (Eds.): Environmental Software Systems. Frameworks of eEnvironmentSeries: IFIP Advances in Information and Communication Technology, Vol. 359, p. 226-234
- Feiden, K. et. al. (2010a): 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, Poland.
- Feiden, K., Kruse, F., Epiropou, V., Karatzas, K. (2010b): 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.

INSPIRE DT [Drafting Team] Data Specification (2008a). Definition of Annex Themes and Scope. Deliverable D2.3. Version 3.0.

INSPIRE DT [Drafting Team] Data Specification (2008b). Methodology for the development of data specifications. Deliverable D2.6. Version 2.2.

Relevant project deliverables to be downloaded at www.gssoil.eu:

GS SOIL consortium (2010): D3.1 Metadata profile for soil geographic datasets and dataset series

GS SOIL consortium (2010): D3.2 Metadata profile for soil geographic data services

GS SOIL consortium (2010): D4.1 Theme specific test suite for developing data specifications

GS SOIL consortium (2010): D5.1 Design specifications of the GS SOIL Portal and its network

GS SOIL consortium (2010): D5.2 GS Soil Portal (Prototype)

GS SOIL consortium (2010): D5.3 First set of open tools and services

GS SOIL consortium (2011): D3.4 Final best practice guidelines for Creating and Maintaining Metadata for Soil Database

GS SOIL consortium (2011): D4.2 Generic application schemes for soil information.