eSoTer Project
www.esoter.org

SoTerML
Soil and Terrain Markup Language

Draft Release: Beta 5.0

As a deliverable for Working Package 6

29 Sep 2010

Amir Pourabdollah (editor) and Didier Leibovici
(Centre for Geospatial Science, the University of Nottingham, UK)

Vincent van Engelen and Piet Tempel
(ISRIC: World Soil Information, the Netherlands)

Stephen Hallett
(National Soil Resources Institute, Cranfield University, UK)

Rainer Baritz and Einar Eberhardt
(BGR: Federal Institute for Geosciences and Natural Resources, Germany)

Vit Penizek
(JRC: Joint Research Centre, the European Commission)

Introduction

Scope of the document

This document and the attached files describe the current situation in the development of SoTerML (Soil and Terrain Markup Language) a XML data exchange format to be used in the FP7 eSoTer project (www.esoter.org) and further use of eSoTer within its contribution to the GEOSS initiative at European and international level.

Version

This edition of SoTerML package is considered as Beta 5.0. The main change from version Beta 4.0 is removing the dependencies to GeoSciML. This issue has been discussed in eSOTER WP6 meeting (28/10/2010 in Godollo University) and as a proposition, it was decided to release a SOTERML package without such dependency, so the partners can evaluate the advantages and disadvantages of the new version.

In this version, no GeoSciML class has been used. For spatial descriptions, SoTerUnit and Profile (the only two objects that need spatial definition) has data elements of type gml:Polygon and gml:Point respectively. In addition, ParentMaterial is no longer a sub-class of EarthMaterial object of GeoSciML. Also, whenever CGI_TerValue or CGI_NumericValue were needed, classes CharacterString, Integer, Decimal or Number classes of GCO (ISO 19103, Geographic information - Conceptual schema language) have been used.

Finally, there is a change in Attribute database: SoilComponent can have text attributes of "WRBLegend" and "revisedLegend". This is different from the soil classification that may be assigned for each Profile.

Rationale

The eSoTer project, described in the next subsection, aims at developing and providing via web services Soil and Terrain data based on enhanced SOTER methodology. The future eSoTer data repository will use SoTerML as data exchange and serialization format.
The design of the XML format called SoTerML has two take into account two competing aspects. The first one is the various soil attribute "profiles" potentially used in SOTER databases such as the SOTER, WRB, the FAO classification schemas, and on the other hand the interoperability developments led by OGC about geographical dataset. On one side its goal is to be able to facilitate understanding and transfert from existing format and model and one the other side is to provide exchange of dataset in a harmonised and standardised fashion also compliant with existing standards. The context of development of this design especially with the ISO/TC190/SC 1 N140 "Recording and Exchange of Soil-Related Data" in progress, enriched the design to be able to be adapted to different soil profiles classification which is also the aim of the development from the ISO working group. Therefore SoTerML will be in principle compliant to the future ISO standard.

An OGC standard based on this specification will be submitted before the end of the eSoTer project.

The e-SoTer project

Soil and land information is needed for a wide range of applications but available data are often inaccessible, incomplete, or out of date. GEOSS plans a global Earth Observation System and, within this framework, the e-SoTer project addresses the felt need for a global soil and terrain database. As the European contribution to a Global Soil Observing System, it will deliver a web-based regional pilot platform with data, methodology, and applications, using remote sensing to validate, augment and extend existing data.

Technical barriers that have to be overcome include: quantitative mapping of landforms; soil parent material and soil attribute characterization and pattern recognition by remote sensing; standardization of methods and measures of soil attributes to convert legacy data. Two major research thrusts involve: 1) improvement of the current SOTER methodology using moderate-resolution optical remote sensing in combination existing parent material/geology and soil information; 2) within pilot areas, advanced remote sensing applications will be developed - geomorphic landscape analysis, geological re-classified remote sensing, and remote sensing of soil attributes.

Advances beyond the state of the art include: transformation of pre-existing data and addition of new information with remote sensing and DEM; interpretations of the e-SOTER database that address threats defined in the EU Soil Thematic Strategy and comparing the results with current assessments; and delivery through a web service of a data portal. e-

SOTER will deliver a Pilot Platform and a portal that provides open access to: 1) methodologies 2) enhanced SOTER databases and artifact-free 90m digital elevation model; 3) dedicated applications related to major threats to soil quality and performance.

Acronyms

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

GEOSS: The Global Earth Observation System of Systems (http://www.earthobservations.org)

XML: Extensible Markup Language (http://www.w3.org/XML)

UML: Unified Modelling Language (http://www.uml.org/)

DEM: Digital Elevation Model

EU: The European Union (http://europa.eu)

eSoTer: The FP7 European project (http://www.esoter.org)

SOTER: Soil and Terrain Database. A database system developed by ISRIC (the Netherlands) (http://www.isric.org/UK/About+ISRIC/Projects/Current+Projects/SOTER.htm)

SoTerML: Soil and Terrain Markup Language

GeoSciML: GeoScience Markup Language (http://www.geosciml.org)

GML: Geography Markup Language, also known as ISO 19136 (http://www.opengis.net/gml)

O & M: Observation and Measurement Encoding Standard (www.opengeospatial.org/standards/om)

FAO: Food and Agricultural Organization of the United Nations (http://www.fao.org)

GCO: ISO 19103, Geographic information - Conceptual schema language (http://www.isotc211.org/2005/gco)

1. SoTerML Packing Content

1- **SoTerML-beta5.0.pdf**
   This document, which briefly describes the whole package.
2- **SoTerML-Annexe A-beta5.0.pdf**
   The data model specification, including the details of the underlying UML design.
3- **SoTerML-Annexe B-beta5.0.pdf**
   The specification document for SoTerML attribute patterns. This document includes a table of different attributes applicable to different elements of SoTerML, and their permitted values.
4- **SoTerML-Annexe C-beta5.0.pdf**
   This annexe is a table for matching between the SOTER database attributes and their equivalent implementation in the SoTerML.
5- **SoTerML-beta5.0.eap**
   SoTerML's UML diagrams and specifications: This has been designed in Enterprise Architecture Professional Edition version 7.1 and includes two sub-packages of eSoTer package, called SoTerML and SoilClassification, in addition to "HollowWorld" package that includes ISO and OGC standard schemas to provide SOTERML package dependencies.
6- **SoTerML.xsd**
   The XML Schema of SoTerML. This is used to validate the core elements of a SoTerML file.
7- **SoilClassification.xsd**
   The XML Schema of the SoilClassification sub-package. This contains the validation rules for a part of SoTerML file that describes the soil classification (e.g. WRB)
8- **AttributeReference.xml**
   The XML containing the SoTerML's attribute reference information. This file must be used as included file within a SoTerML file. A sample is shown in SoTerMLSample.xml.
9- **SoTerMLsample.xml**
   This is new in this release and contains a sample file with real data about a SoTerUnit with a single Terrain Component and two Soil Components. This demonstrates how SoTerUnit elements (and particularly "attribute" element) can be used and put together.
10- **HTMLDocument Folder**
   This folder contains the documentation of the package in HTML format. To start, open index.html from within this folder.

2. Design Backgrounds

The design of the data model has been based on the development of different SOTER databases in the past, and soil classification documents. The main resources are:

1. **SOTER1995:** Global and National Soils and Terrain Digital Database (SOTER) Procedures Manual (revised edition), V.M.P. van Englen and T.T. Wen (editors), ISRIC, 1995.
2. **SOTER2002:** SOTER - Global and National Soils and Terrain Digital Database, Database Structure v3, ISRIC working paper No. 02/01, Piet Tempel, September 2002.
3. **SOTER2009 (January and April editions):** SOTER Manual Update for Discussion eSoTer partners (Draft): SOTER Attribute Data Coding + draft of Non-spatial attributes of a SoTer Unit.
4. **WRB2006/2007:** IUSS Working Group WRB. World Reference Base for Soil Resources 2006, first update 2007. World Soil Resources Reports No. 103. FAO, Rome, 2007.
5. **FAO1988**: Soil Map of the World Revised Legend with Corrections. FAO, Rome, 1988.

According to eSoTer project description, SoTerML is an extension of GeoSciML, thus GeoSciML and its own dependencies (including GML) are dependencies of SoTerML.

3. Class Hierarchy

The details and diagrams of the class hierarchy in SoTerML data model are described in annexe A (the attached file) as well as the HTML documentation (in a folder with the same name). A summary UML diagram of the design has been shown here in figure 1.

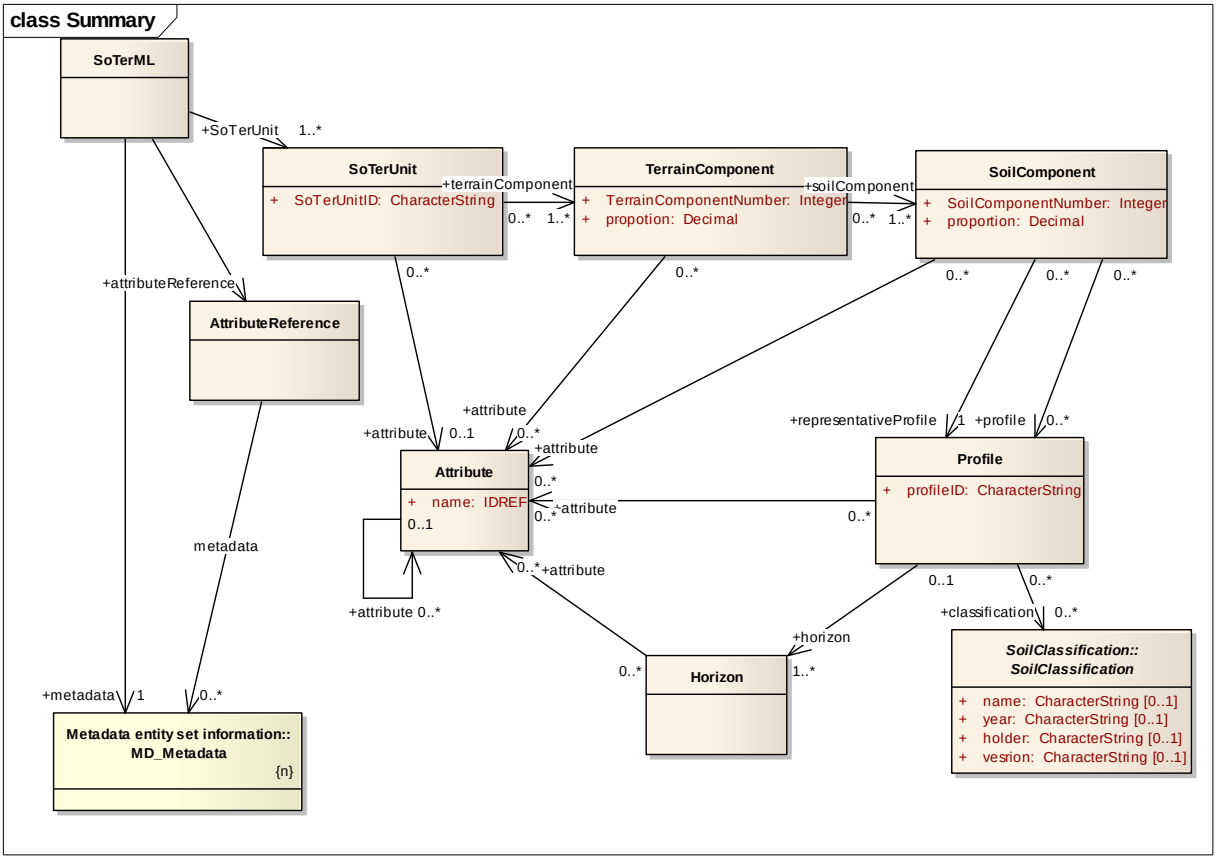

Figure 1: Top level description summarising of SoTerML objects

A brief review (as in Figure 2) shows that:

1- The root element is called **SoTerML**. This can include a number of **SoTerUnits**.
2- Each **SoTerUnit** may contain a number of **TerrainComponents**
3- Each **TerrainComponent** may contain a number of **SoilComponents**.
4- Also a number of Soil-**Profiles** can be associated to a **SoilComponent**.
5- Each **Profile** then itself includes a number of **Horizons**. A **Profile** is classified according to one of the standard **Soil-Classifications**. A soil classification (e.g. WRB2006) instance is

then associated to each profile. For the case of WRB 2006/2007, an RSG and to a number of **Qualifiers** (and optionally**, specifier**s) defined as prefix or suffix.

6- All of the above classes may have a number of SoTerML-specific or GeoSciML-specific **attributes**.

7- Each **attribute** above has a name and value. It may also be associated to a number of **AnalyticalMethods**.

8- The geometry of each SoterUnit, TerrainComponent, and SoilComponent is implemented by the link to a GML Polygon object. The geometry of a profile is also implemented by the link to a GML Point object.

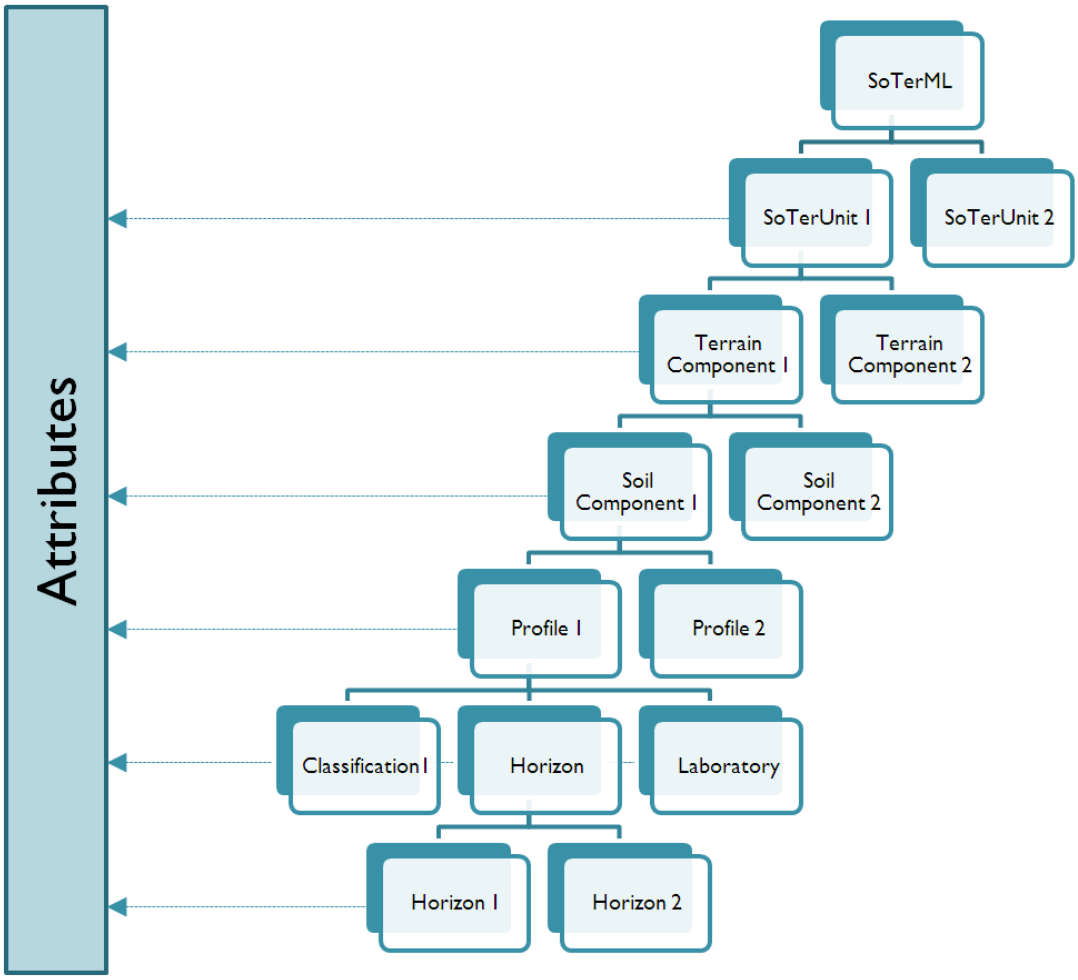

Figure 2: The class hierarchy of SoTerML

4. Attributes Design

The approach in designing SoTerML data attributes was to incorporate as much flexibility and reuse as possible. For this reason, the required data attributes in the past developments of the SOTER database structure have been categorized into the following four categories:

1- Some attributes are needed in SoTerML as derived directly from the SOTER database, like **SoTerUnitID**.

2- Some other attributes are required for a relational database design, but are not necessarily required for the SoTerML structure, partially because SoTerML uses a semi-structured XML design where redundant information is not always avoided, like **profile_ID**.

3- Many attributes can be automatically associated to or inherited from upper classes, or from Soil Classification classes, if the right classes are chosen for association/generalization in SoTerML data model.

4- A majority of attributes are designed to be implemented using the *Attribute mechanism* in SoTerML (detailed below). These are the attributes that are preferred to be fed from an *AttributeReference* bank, rather than being an explicit data member of SoTerML classes. Having these attributes as records of a data bank and not represented as columns can provide much more flexibility in the data model especially to the current and future amendments of the assignable attributes to each SoTerML class. These attributes are themselves divided to three groups:

   a. Numeric or literal: These attributes can accept a number or a literal string, like pH-KCl.

   b. Enumerated: The values of these attributes must be one of a set of predefined values, like vegetation having a set of permitted values as "I", "IA", "IA1", "IA2" etc.

   c. Sub-attributes: Some of the attributes are a container for some other sub-attributes of one of the two above groups, forming a recursive reference to other attributes, like dryColor attribute which has its own three attributes, being hue, value and chroma.

The attributes of the category 4 above with their properties and possible enumerations, has been called *Attribute Reference* in SoTerML. A descriptive list of Attribute Reference has been attached (as Annexe B) and an XML version of this list is also created (AttributeReference.xml). This XML file depends on the schema of the

AttributeReference element which is built in the SoTerML schema. As a result, a future SoTerML data file must have SoTerML.xsd as schema and AttributeReference.xml as an included XML file. For including a XML file within another XML –as shown in the included sample file- the XInclude mechanism is recommended as a W3G standard (http://www.w3.org/2001/XInclude). It is noticeable that AttributeReference may not be itself validated unless it is used within a validated SoTerML file. This abstraction (separating AttributeReference from the main schema and from a SoTerML document) allows that it can be maintained separately and provides a logical border from the language and associated values. In AttributeReference files, each attribute has been named together with its SourceElement, Description, Reference and a list of permitted enumerated values (where applicable). SourceElement is the name of SoTerML class (or the name of the parent attribute for the case of 4.c above) in which this attribute can be assigned. For each enumerated value, short and full terms and description are included.