OGC SOIL DATA IE SUMMARY

ABSTRACT

The OGC Soil Data engineering report summarised here will describe the results of the Soil Data Interoperability Experiment (the IE) conducted in 2015 under the auspices of the Open Geospatial Consortium (OGC) Agriculture Domain Working Group. Soil data exchange and analysis is compromised by the lack of a widely agreed international standard for the exchange of data describing soils and the sampling and analytical activities relating to them. Previous modelling activities in Europe and Australasia have not yielded models that satisfy many of the global data needs of soil scientists, data custodians and users. This IE evaluated existing models and proposed a common core model, including a GML/XML schema, which was tested through the deployment of OGC web services and demonstration clients. IE time constraints and limited participant resources precluded extensive modelling activities. However, the resulting model should form the core of a more comprehensive model to be developed by an OGC Soil Data Standards Working Group in collaboration with other parties.

OBJECTIVES

The objectives of the Soil IE were:

- Further develop and test a GML compatible information model for soil features, based on existing international initiatives.
- Seek participation by a number of soil agencies from across the world to establish data services using the defined model.
- Demonstrate delivery of standardised soil data services from multiple and disparate sources, and the use of these services within a number of user focussed applications.
- Prepare an OGC/IUSS engineering report with the intention to develop it into a data specification subsequent to the IE.

The Technical Approach for this Interoperability Experiment followed existing principles for development of information models as exemplified by GeoSciML, GWML2 and WaterML2 creation.

USE CASES

Four use-cases were developed for the IE to constrain the requirements of the information model and test implementations. The use cases addressed scenarios commonly encountered by providers and users of soil data: the delivery of data describing soil features and the use of algorithms or other processing methods to generate new data. A fifth use case addressing soil type mapping was proposed but not developed due to a lack of resources.

1. *Soil Data Integration*: Use the schema as a canonical structure into which heterogeneous soil data formats are transformed and published.
2. *Soil Sensor Data*: Identify the location and properties of sensors monitoring dynamic soil properties, and the provision of the measurements made at those sites.
3. *Soil Property Modelling and Predictions*: The provision of high resolution estimates of functional soil properties generated using digital soil mapping techniques.
4. *Pedo-transfer Functions*: Delivery of observed and interpreted soil properties (by soil type and/or by spatial distribution) in a standard format that allows the use of pedo-transfer functions.

REVIEW OF PREVIOUS WORK

EXISTING SOIL INFORMATION MODELS

Various national and international initiatives have worked on information models to support the exchange of soil data. This experiment intended to reconcile core concepts and features of these models into a single coherent, fully attributed, implementable, albeit provisional, standard. The five models that were reviewed in this IE were:

1. Australia and New Zealand Soil Mark-up Language (ANZSoilML);
2. e-SOTER Soil and Terrain Mark-up Language (SoTerML);
3. INSPIRE D2.8.III.3 Data Specification on Soil (INSPIRE Soil);
4. ISO 28258:2013 Soil quality – Digital exchange of soil-related data (ISO SoilML); and
5. IUSS/ISO 'Wageningen Proposal' (a variation of 4 with reference to 1, 2 and 3).

COMPARISON OF MODELS

The models were compared according to their scope (the breadth of information captured by each model), the modelling techniques and patterns used in their definition, whether they are readily available and implemented, and the context in which they have been used (e.g. production or prototype).

Scope refers to the breadth of information captured by each model. The FAO Guidelines for Soil Description, with additional guidance from the USDA Field Book for Describing and Sampling Soils, was adapted to provide a basis for comparison.

The criteria for comparison can be grouped into eight categories:

1. Site registration: identity, location, timing and other metadata about sampling sites;
2. Soil formation: environmental and human factors influencing the formation of the soil;
3. Soil description: physical, chemical and organic character of a soil;
4. Sampling: collection of physical samples;
5. Observation: field or laboratory measurements of soil properties;
6. Classification: categorisation of soil and horizons according to formal taxonomies;
7. Vocabularies: systems for managing terms and their definitions; and
8. Mapping: mapping the distribution of soils according to their type.

Modelling approach refers to how the model was defined. Models may be: *comprehensive,* attempting to cover as many dimensions of the soil as possible; *targeted*, hard-typing a selected set of essential properties while relying on soft-typing for a significant set of properties; or *framework,* a model that simply provides a framework of classes.

Accessibility refers to the availability of the model (UML models, XML schema, specification documents and other artefacts) in terms of access constraints or charges for access.

Implementation shows whether a standard has been implemented, either as a prototype in a production environment.

Table 1 is a comparison of the modelling approach and scope of each information model and Table 2 summarises matters of implementation. No existing model was accepted by the IE participants as the primary basis for a soil information model, due to a lack of agreement on the concepts in each of the models, or incompleteness due to a reliance on soft-typing.

The soil model developed as part of this IE (SoilIEML), although not part of the analysis of existing models, is included in the Table 1 for completeness. Green cells (F) denote aspects of a soil description covered by classes and formally defined properties; yellow cells (S) denote aspects of a soil description covered by classes and soft-typed properties; Grey cells (P) denote partial coverage; and white cells (X) denote aspects of a soil description that are not covered, or handled by a soft-typing mechanism.

Table 1: A comparison of concepts and properties available in existing soil information models.

Model / Soil Information	ANZSoilML	eSoTer	INSPIRE Soil	ISO SoilML	IUSS/ISO SoilML	OGC SoilIEML
Site Registration	F	S	F	F	F	F
Soil Formation	F	S	P	X	X	X
Soil Description	F	S	F S	P	P	F S
Soil Sampling	F	S	F	F	F	F
Observations	F	S	S	F	F	F
Soil Classification	F	F	F	F	F	F
Vocabularies	F (SKOS)	X	F (XSD/SKOS)	X	X	F (SKOS)
Soil Mapping	F	F	F	F	F	X

Table 2: Comparison of model accessibility and implementation.

Model	Accessibility	Implementation
ANZSoilML	Public; free of charge	Production services
eSoTer	Public; free of charge	Prototype (to be confirmed)
INSPIRE Soil	Public; free of charge	Prototype (GS Soil)
ISO SoilML	Private; charge for access	Prototype (GS Soil) (?)
IUSS/ISO SoilML	No model	No

SOIL DATA IE INFORMATION MODEL

As there was no suitable single starting model selected from existing candidates, a compromise model was developed (see Table 1). Wherever possible classes and patterns from the existing models were used. To constrain the model's scope, the selection of properties to be assigned to soil classes was based on the specifications of the GlobalSoilMap consortium with the intention to allow the model to expand in a modular fashion.

Three models were produced by the IE: a technology independent UML *conceptual model*; a UML *logical model* describing an ISO/OGC compliant application schema; and an XML *physical model* defining the encoding of data using Geography Mark-up Language.

To constrain modelling activity and help ensure that the model defined was not disruptive (conforming closely to existing OGC standards) or parochial (recognising that soil scientists must integrate data from many environmental domains), the group defined a set of guiding principles, including:

- Open publication of the results (UML, XSD) once approved through the OGC process.
- Open development of the model, subject to appropriate agreements with the OGC.
- Re-use of existing models wherever possible (if an existing model does not meet the project needs then work consult the relevant community to address issues before extending or branching the model).

The motivation for the principles was twofold:

- To ensure consistency of data types across domains - soil is a function of geology, climate, topography, and biology, and pedologists must aggregate these data during analysis.
- To ease process of deployment – a client developed for geology or climate observations encoded as O&M can be used for soil observations without modification (other than dealing with domain specific values).

IMPLEMENTATION

ARCHITECTURE

Three suites of components were deployed to implement the test services and clients. Details will be published in the Engineering Report

- Web services conforming to OGC specifications (WFS, WCS, WPS and SOS) for the delivery and processing of site registration, and soil formation, description, sampling, observation data.
- Web services conforming to World Wide Web Consortium (W3C) Linked Data standards for the resolution feature URIs and the delivery of soil classification and vocabulary definitions.
- Bespoke tools to present or transform data.

DEMONSTRATION

Demonstration services and client applications were successfully deployed. Readers are referred to the video of the Soil Data IE demonstration made at the Sydney meeting of the OGC Technical Committee in December 2015, and the client application developed by Federation University of Australia:

Soil Data IE Demonstration - Sydney, Australia – 2015-12-03 (YouTube):
https://www.youtube.com/watch?v=oR-c7Viu19k

CERDI Federation University Australia Soil Data IE Demonstrator:
http://data.cerdi.edu.au/soil_demo.php

CONCLUSION

This Interoperability Experiment has shown that the problem of soil data interoperability can be addressed and the definition of solutions has already been well advanced by existing initiatives. Overall, the information model defined by the Soil IE:

- Matches the conceptual needs of the GlobalSoilMap specification;
- Exceeds the coverage of the ISO SoilML, IUSS/IUSS and eSoTerML schema;
- Is broadly equivalent to the INSPIRE Soil model;
- Falls well short of the ANZSoilML model; and
- Requires reworking to better allow class and property extensions.

Current technology allows the deployment of web services to support soil data exchange. The main impediment to progress is the absence of a standards community that is well enough resourced and mentored to develop the necessary technical and modelling skills required to develop a model and systems that describe a complex and nuanced natural resource.

Future work will need to reconcile the need for a comprehensive and unambiguous conceptual model and a flexible, lightweight implementation model. At the time of writing the Soil Data IE engineering report, and its conclusions and recommendations, has not been approved for release by the OGC so recommendations and future work cannot be discussed in detail. They can be broadly summarised as:

1. recommendations for improvements or changes to the OGC standards and practices to better support soil data interoperability; and
2. the establishment of a Standards Working Group to undertake the modular development of a set standards for soil data in coordination with activities of the IUSS WGSIS and the FAO Global Soil Partnership (Pillars 5 and 4).