Web enabling soil data for the Farm Dairy Effluent Storage Calculator

Envirolink 1067-MLDC72

Landcare Research
Manaaki Whenua

Web enabling soil data for the Farm Dairy Effluent Storage Calculator (MLDC72)

Alistair Ritchie, Sam Carrick and Linda Lilburne

Landcare Research

Prepared for:

Marlborough District Council

15 Seymour Street
PO Box 443
Blenheim 7240
New Zealand

July 2012

Disclaimer

This report has been prepared by Landcare Research for Marlborough District Council. If used by other parties, no warranty or representation is given as to its accuracy and no liability is accepted for loss or damage arising directly or indirectly from reliance on the information in it.

Contents

Summary

Web service delivery of S-map Online

- This report proposes the delivery of soil data from Landcare Research's S-map Online database using the spatially enabled OGC Web Feature Service. The data will be structured using a flexible data model, defined and published as an OGC Geography Mark-up Language application schema. For each occurrence of a S-map soil polygon a GML document attributing the polygon with soil type and dairy effluent risk categories will be provided (with time, the data model should grow to support the provision of data for other applications). Qualitative or quantitative estimates of uncertainty will be included for each individual property value where appropriate. Client applications will be able to invoke all spatial filter and data operations specified by the WFS protocol when accessing data. The data itself will reflect the most up to date data published by Landcare Research via S-map Online.

Project background

- Studies show that poor management of farm effluent can result in contamination of surface and groundwater. New Zealand regional councils are therefore reviewing their rules around farm effluent management. DairyNZ is running a significant programme to help councils and farmers improve effluent management systems, with the Farm Dairy Effluent Storage Calculator (DESC) as a key tool. Once implemented, the improved management of farm effluent will help improve water quality in farmed catchments, particularly those with intensive dairy farming.

Soil data services

- Environmental modelling tools like the DESC require up-to-date and accurate data to ensure that their calculations are robust and efficiently executed. As a means of achieving this, much greater use is being made of technology and standards of the World Wide Web.

- Landcare Research is developing and applying information standards for the provision of environmental data in this context, aligning them with national and international open-data initiatives. The intent is to provide core scientific data, and derived products, in a way that allows their use for a number of purposes and promotes the integration of data from multiple domains and agencies.

- The use of common data formats based on the standards of the Open Geospatial Consortium (OGC) ensures data can be accessed by multiple tools with reduced effort in data manipulation and software redevelopment. In turn this reduces the time required, and therefore the costs incurred, to provide data and develop applications that use the data.

Recommendation

- At the time of writing the proposal has not been tested beyond a conceptual rule and manual creation of data. A test deployment is required to ensure that the proposal meets the needs of the DESC. If this is successful then a project to deploy the service in an appropriately governed environment is required. A fundamental aspect of this work will be ensuring the services are well maintained and consistently available with a corresponding allocation of appropriate resources and funding.

1 Introduction

Studies have shown that poor management of farm effluent can result in contamination of surface and groundwater (Houlbrooke et al. 2004). New Zealand regional councils are currently reviewing or modifying their rules around farm effluent management, and DairyNZ is running a significant programme to help farmers improve their effluent management systems, with the farm Dairy Effluent Storage Calculator (DESC) as a key tool. Once implemented, the improved management of farm effluent will, over a number of years, help improve water quality in farmed catchments, particularly those with intensive dairy farming.

Environmental modelling tools like the DESC must be easily accessible and require up-to-date and accurate data to ensure that their calculations are robust and efficiently executed. As a means of achieving this, much greater use is being made of technology and standards of the World Wide Web. DairyNZ are in the process of redeveloping the DESC, moving it from a stand-alone desktop application to an interactive website.

Landcare Research is complementing this work by developing and applying information standards for the provision and exchange of environmental data, and deploying web data delivery services that conform to these standards. This work is aligned with national and international initiatives, including the New Zealand Spatial Data Infrastructure (NZGO 2011) and Global Soil Map[1]. The intent is to provide core scientific data, and derived products, in a way that allows their use for a number of purposes and promotes the integration of data from multiple domains and agencies.

The main benefits of standardised web-delivery of data are greater accuracy and efficiency of decision making; it ensures tools have access to the most up to date information for modelling purposes (real-time access to databases); and the use of common data formats across agencies allows data to be accessed by multiple tools with reduced effort in data manipulation and software redevelopment.

This report does not evaluate the DESC or document the tool in detail. Rather, it uses the DESC as a case study to determine the feasibility of Landcare Research's web services as authoritative data sources for environmental modelling tools. It does so by proposing a data model and web service architecture that will allow S-map data to be delivered to the DESC. The report briefly considers other environmental modelling systems to illustrate the benefits of supporting applications with requirements other than those of the DESC: e.g. reducing the number of systems maintained by Landcare Research (and operational costs) and ensuring the provision of consistent and comparable data to related modelling applications.

This report is part of a body of work describing soil information services from Landcare Research (see also the related National Soils Database scoping report – Hewitt et al. 2012). The solution defined here will be reconciled with NSD data delivery needs – defining an integrated approach to environmental data delivery. It is expected that further design work and testing will be required before a final, production-ready solution can be deployed. Consideration must also be given to how such services will be financially sustained.

---

[1] GlobalSoilMap.net: http://www.globalsoilmap.net/

2 Background

2.1 Farm dairy effluent risk

The DESC was first developed by Massey University, with support from Horizons Regional Council (Horne et al. 2011), and has subsequently been adopted by almost all regional councils and the wider dairy industry (Horizons Regional Council 2008, Environment Southland 2009, Bay of Plenty Regional Council 2010, Waikato Regional Council 2010, Wellington Regional Council 2011, Marlborough Regional Council 2011, West Coast Regional Council 2012). Underpinning the DESC is the farm dairy effluent (FDE) risk framework, developed by AgResearch, which classifies soils into one of five categories based on soil–landscape features (Table 1). For each category there is a suite of minimum management practices, based on each category's inherent risk for direct loss of FDE during land application. In the DESC the five categories of the risk framework are classified as either high risk (categories A, B, C) or low risk (categories D, E). The low-risk soils allow irrigation of small depths of FDE once the soil has reached field capacity, whereas the high-risk soils require full deferred irrigation principles where application depth must be less than soil water deficit, Thus greater FDE storage is required for high-risk soils during periods of insufficient soil water deficit, compared to low-risk soils where storage is limited to during and immediately after rain events.

The use of both the FDE risk framework and the pond storage calculator is complementary. Figure 1 illustrates how DESC uses the high- or low-risk class determined by the FDE risk framework as an input into the calculations to determine the required storage volume.
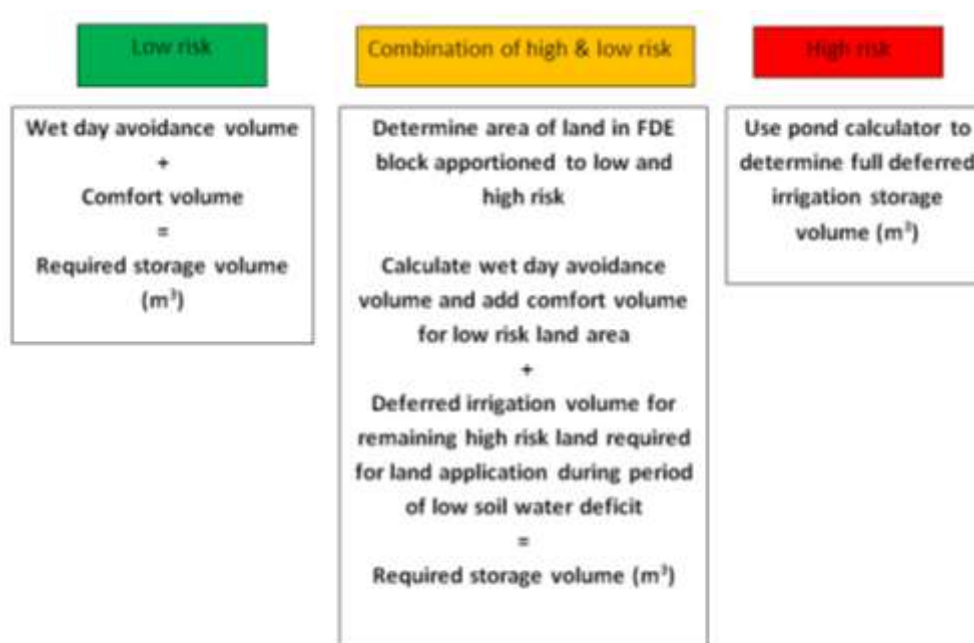

Figure 1 Conceptual diagram of how DESC uses the high or low risk class determined by the FDE risk framework to calculate the required storage volume (Houlbrooke et al. 2011).

Table 1 The five categories of the FDE risk framework, with the risk class used in DESC, and the suite of minimum management practices (Houlbrooke et al. 2011)

Category	A	B	C	D	E
Soil and landscape feature	Artificial drainage or coarse soil structure	Impeded drainage or low infiltration rate	Sloping land (>7°) or land with hump & hollow drainage	Well-drained flat land (<7°)	Well-drained very light flat land (<7°)
Risk Class	High			Low	
Application depth	< soil water deficit			< 50% of water holding capacity in top 300 mm of soil	≤ 10 mm & < 50% of water-holding capacity in top 300 mm of soil
Instantaneous application rate			< soil infiltration rate		
Average application rate	< soil infiltration rate			< soil infiltration rate	
Storage requirement	Apply only when soil water deficit exists			Wait for 24 hours after soil is wet	
Maximum N load	150 kg N/ha/yr			150 kg N/ha/yr	
Max depth: High rate tool	10 mm			25 mm (10 mm at field capacity)	10 mm
Max depth: Low rate tool	25 mm	25 mm	10 mm	25 mm	10 mm

2.2   Case study: Marlborough District Council

The Marlborough District Council is currently implementing the use of the pond storage calculator for farm effluent management, with a full report on application of the FDE risk framework to Marlborough soils provided by Houlbrooke et al. (2011). Knowledge of the soil type on a property is a key parameter for the calculator, and the council is incrementally working on providing this information to the S-map database.

At present Marlborough District Council are collecting new soils data, adding it to S-map database and making it available through S-map Online. Soil mapping is also underway in other regions, e.g. Canterbury, Horizons, Hawke's Bay, and West Coast. However, the problem is that each time the soil map for a new area becomes available, the DESC will also need to be updated with the new soil information. There is the real risk that S-map Online and the DESC will become out of synch, resulting in confusion and poor decision-making.

This situation will be avoided by real-time access to data for the soil-type component of the pond storage calculator over the Internet, but development of this is beyond the skills of the council.

3 Farm Dairy Effluent Storage Calculator

3.1 Description

The farm Dairy Effluent Storage Calculator (DESC) calculates farm-specific storage requirements. It takes input data relating to climate (30 years of local meteorological data on a daily time-step); soil risk (see Section 3.2); the pond catchment area (including existing effluent ponds); wash water generated during milking; and irrigation hardware and management (daily pumping volume). The climate data are provided by NIWA and Plant & Food Research, and the soil risk categories and data are from AgResearch and Landcare Research.

Once processed the tool provides estimates of the maximum and minimum required pond volume; the probability that a given pond storage volume is sufficient; a plot of cumulative effluent volumes that would have been required historically; and cumulative irrigation days required for a given period. All of these outputs are summarised in a single, printable report.

The tool is used by a range of agencies and individuals, all of which have varying levels of comprehension of the scientific basis for the application and the ultimate application of the model outputs. Users include:

- Regional councils
- Dairy companies (Fonterra)
- Effluent system designers and implementers
- Dairy farmers
- Researchers (e.g. catchment modelling)
- Educators
- Agricultural consultants

In its current form it is deployed as a stand-alone combination of a Microsoft .NET application and Microsoft Access database holding all the regional input data not provided by the end-user. There is no mechanism for updating the data in the database; instead the application is valid for a period of time before it expires. At this point, a new version of the application, with updated data, is provided.

In 2012, DairyNZ engaged contractors to migrate the application from the PC desktop to a web-based application (with potential for off-line access). The functionality will remain the same in terms of inputs and outputs, but it will now make use of data provided in near real-time from authoritative data sources (e.g. Landcare Research web data services), rather than cached data sources within the application. The application will also be able to display, manipulate and process spatial information.

3.2   Use of Landcare Research soil data

The DESC requires estimates of low- and high-soil risk (see Section 2.1) as an input parameter (Figure 2a). In the current version these are entered manually by the end-user after they have used the Landcare Research S-map Online application[2] to establish what soils are in the area of interest and the relevant percentage of the area covered by each soil. The risk type for each soil is then established from a look-up table provided with the application (Figure 2b) and the area of soil risk is manually estimated (the sum of the percentage of the area of interest covered by each soil risk type).
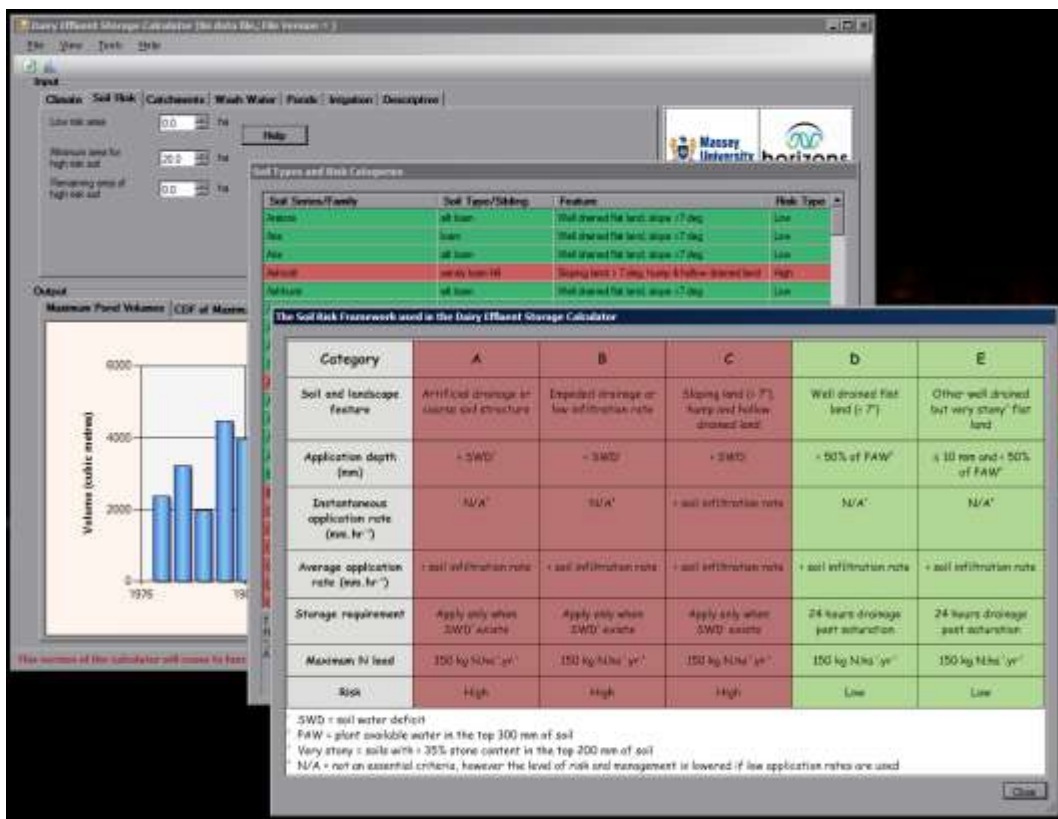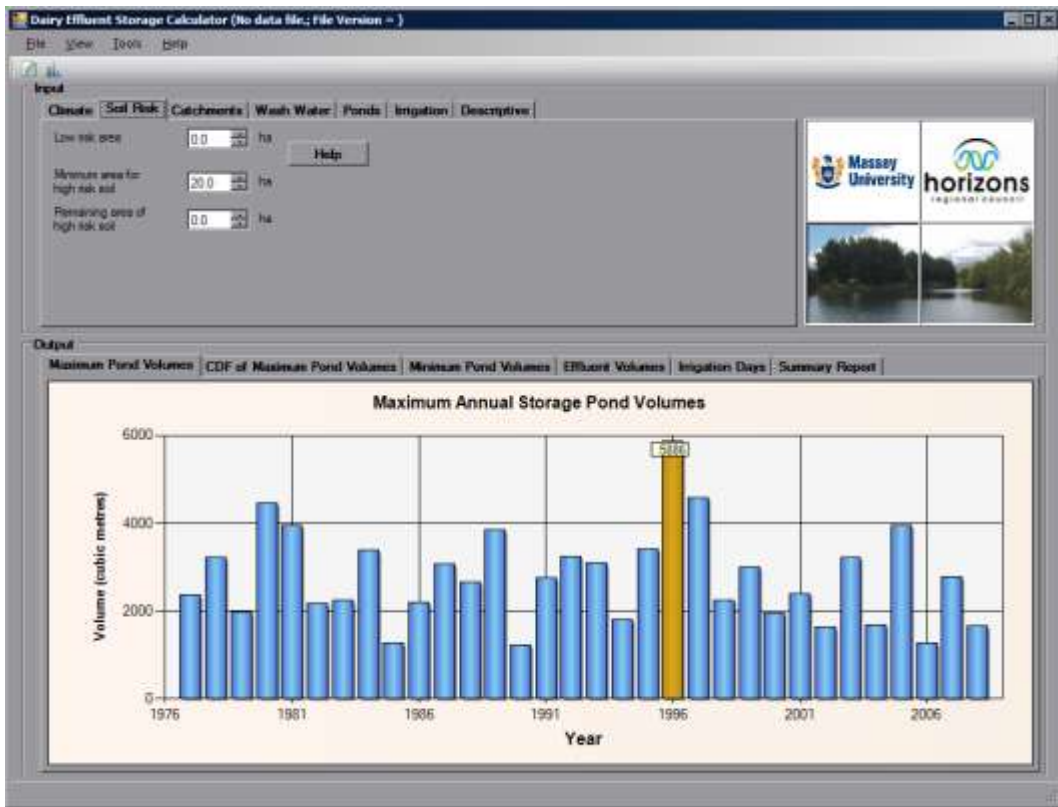
The are several problems with this approach:

- *Accuracy*: It requires estimates of the areal extent of the soil types to be made in S-map Online – a tool not designed to make these measurements – and then be manually entered into the DESC tool. This introduces potential errors due to inaccurate measurements and subsequent errors in data entry.

- *Currency*: The soil-type and risk-category mappings are specified in a static table that may be out of date at the time of calculation.

- *Cost:* It takes time to capture these data and transcribe them into the new application, time that a farmer could best use elsewhere, and errors can result in costs related to the construction of either inadequate or excessive storage ponds.

By providing real-time access to the Landcare Research soil data these issues can be addressed. The application can incorporate tools to accurately establish soil type and risk areas using appropriate spatial operations[3], automatically capturing the result; and the best available soil mapping and risk estimates will be available. All of this can be automated, reducing the time required to use the system.

---

[2] http://smap.landcareresearch.co.nz

[3] Note that accuracy is relative and this section assumes that the soil data have been captured at a scale appropriate for use in the DESC.

Figure 2 (a) Screen shot of the pond storage calculator showing the soil risk data entry tab. (b) Soil type and category tables, accessed via the Help button, which map the soil types to soil risk categories.

4   Related environmental modelling tools

Any soil data service provided by Landcare Research should support a variety of environmental modelling applications. This will minimise the number of systems maintained by Landcare Research, with corresponding cost savings, and, more importantly, will ensure that all modelling activity is undertaken using the same authoritative data sources, improving consistency between models. It is therefore useful to consider other tools, and the inventory of environmental models compiled for the NIWA-led interoperable freshwater modelling project provides a sample of existing applications[4]. Of the 39 listed models 25 require soil data as an input parameter, and of these, Overseer and IRRICALC are included as specific examples below because their developers or sponsors are discussing more efficient access to soil data with Landcare Research. The summaries are taken from the NIWA model inventory.

4.1   Overseer

Overseer is a model for farm-scale nutrient budgeting and loss estimation developed by AgResearch and the Ministry for Primary Industries, in collaboration with NIWA, Plant & Food Research, Landcare Research, and a number of farm industry agencies. It uses farm production and management data, and soil and weather information (currently embedded in the database), to provide models of internal N, P, K, S, acidity and greenhouse gas balance, leaching and greenhouse gas emissions. The tool enjoys widespread use by farmers, fertiliser company representatives, scientists and policy researchers.

Future development activity is looking at the capability to support web-enabled soil data. Specifically Overseer will require soil data for key attributes (e.g. drainage class, profile available water) for a specified parcel of land, such as one or more management blocks, or the whole farm.

4.2   IRRICALC

IRRICALC calculates soil moisture, water use, and drainage for irrigation systems in New Zealand. The model is based on the FAO-56 single-layer water balance model. It has been developed by Aqualinc and is widely used by irrigation users and the dairy industry. The model uses daily rainfall, potential evapotranspiration, profile-available soil water, and crop factor time-series data with irrigation rules to produce time-series models of soil moisture and water use. Environment Canterbury has used IRRICALC to generate maps of water demand for various crops and soils.

Irrigation New Zealand is promoting IRRICALC as a tool, along with tools like the DESC, to define irrigation codes of practice and enforce regulatory compliance. It has approached Landcare Research about the provision of real-time soil data to support a web-enabled version of the model.

---

[4] https://teamwork.niwa.co.nz/display/IFM/Compilation+of+models+and+their+attributes

5 Web delivery of soil data

5.1 Standards-based interoperability

Efficient delivery of up-to-date and reusable spatial data requires a standardised approach. Not only must the structure and content of the data be stable, but the way in which the data is accessed must also be well-defined and not subject to ad hoc changes. This way, the providers of client applications can develop tools confident that these will be supported for the foreseeable future.

Standardising data delivery and access requires some sort of context. Traditionally this has involved agreement around a common software package and its mandated data formats. This does not, however, guarantee stability of access – data models may still vary; organisations may still end up using different, perhaps incompatible, versions of the software; and new users may not be able to purchase or use the agreed package.

A preferable solution is to allow each participating organisation to maintain its own software environment and instead exchange its data through a standardised exchange layer that mediates between the systems via a common set of data-access protocols and agreed data models. These solutions are referred to as 'interoperable' as they allow a network of heterogeneous systems to interoperate, or work together to exchange data.

Networks implemented along these lines are known as Spatial Data Infrastructures (SDIs). They are deliberately designed without reference to any particular technology, instead focusing on definition of the behaviour of the systems. This gives free reign to the data providers and client applications in the system to implement the solution using any technology they see fit. Provided a client can request or submit data to a data service using the published protocols, and the servers provide the mandated response, both are valid participants in the system.

While the DESC does not itself require the deployment of an SDI – an application specific agreement to acquire soil data from Landcare Research would suffice – Section 4 shows that it is not the only environmental modelling application to require soil data. It is much more efficient for Landcare Research to deploy a single data source that can be used for multiple-purposes. Conversely the DESC benefits from the higher stability and availability that comes with data sources that have a larger client base.

A number of communities have defined standards for web-based data infrastructures; however, this report recommends the use of geospatial standards defined by the International Standards Organisation (ISO) and implemented by the Open Geospatial Consortium (OGC). The recommendation is based on the fact that the ISO/OGC standards provide explicit support for encoding spatial data and the operations that may be performed on spatial data; and they are increasingly applied to scientific and environmental data exchange, both internationally and within New Zealand.

The specification for the New Zealand SDI has been developed by the New Zealand Geospatial Office to meet the access requirements defined in the New Zealand Geospatial Strategy (LINZ 2007). The Minister for Land Information has directed state-sector and Crown agencies to comply with the standards of the SDI. While not explicitly nominated as a

participating agency, it is desirable for Landcare Research to align itself with this SDI, not least because a significant amount of complementary environmental information will be made available in this way.

5.2   ISO and OGC standards and services

The OGC is 'an international industry consortium of 456 companies, government agencies and universities participating in a consensus process to develop publicly available interface standards'[5]. It works closely with Technical Committee 211 of the International Standards Organisation (ISO/TC211) to implement their standards for a suite of geospatially oriented web services. These services support the delivery of maps, features (raw vector data), coverages (gridded data) and catalogues (that enable the publication of the other services). The consortium has also implemented ISO/TC211 standards to define practices for the modelling and encoding of geospatial information.

Two OGC standards are applicable to the web-enablement of the DESC: the Geography Mark-up Language (Portele 2007) and the Web Feature Service (WFS) implementation specification (Vretanos 2010).

5.2.1   Geography Mark-up Language

Geography Mark-up Language is a grammar for the encoding of geographic features using eXtensible Mark-up Language (XML) – a structured data format that is intended to be both human- and machine-readable. It is based on the ISO/TC211 General Feature Model (ISO 2002) – a conceptual basis for describing real-world features – and includes an extensive model of basic, spatial and temporal data-types. GML uses the definition of application schemas – so-called because they apply the general model to a particular field of interest – to extend GML into well-documented, domain-specific data models. GML application schemas are initially documented at a conceptual level using the graphical Unified Modelling Language (UML) and these are then converted in a physical data model encoded as XML schema documents (XSD).

The International Union of Soil Sciences (IUSS) has formed a working group to define the Soil Mark-up Language (SoilML) as an application schema for soil information exchange (Montanarella et al. 2010). It will reconcile a number of existing models, including the European INSPIRE soil conceptual model (INSPIRE 2011), the Soil and Terrain Mark-up Language (Pourabdollah et al. 2012) and the Australian OzSoilML. This is not a schema that will be ready in a suitable time frame for the DESC and may also not cover derived soil attributes like dairy effluent risk. As a result it has not been considered as a data model for use in this report. It is, however, recommended that its development be followed as it may be an application schema used by future versions of the application.

---

[5] http://www.opengeospatial.org/ogc

5.2.2 Web feature service

A Web Feature Service is a web service that provides GML representations of a geospatial feature to clients as a sequence of HTTP transactions. Depending on the nature of the service it will support basic query and discovery transactions (Basic WFS), or can be extended to allow transactions that support the creation, modification and deletion of data (Transactional WFS). The structure of the data returned as GML is defined in an application schema. Related to the WFS is the Web Map Service (de la Beaujardiere 2010) that provides precompiled map layers as spatially located images and can be used to efficiently display feature data from WFSs in GIS applications and web portals.

5.3 Controlled vocabularies

The standardisation of data is not restricted to models for the structure of data. Consideration must be given to ensuring that there is a shared understanding of meaning of qualitative data provided by the services. Controlled vocabularies are valuable for ensuring content is well documented with definitions for terms used, and consistent and error free (no spelling mistakes or transcription errors).

Landcare Research is also investigating ways of tailoring these web-enabled definitions to suit particular audiences. As a result, definitions of concepts, such as 'dairy effluent risk class A', may be available in multiple contexts, e.g. with a formal scientific definition or using a definition written from the perspective of a farmer.

GML defines dictionaries that are collections of definitions of terms and concepts and may be used to provide controlled content. Many communities are, however, favouring the use of the Simple Knowledge Organisation System (SKOS) over the GML Dictionary model and are integrating SKOS content into GML application schema web services (e.g. GeosciML).

SKOS is a Semantic Web[6] standard for encoding vocabulary items and is more expressive when it comes to interrelating codes, either into hierarchies within a vocabulary or mapping between equivalent values in different vocabularies. More importantly it identifies entries in a vocabulary with Uniform Resource Identifiers (URI) that can double as locations for definitions of terms on the Web.

---

[6] The Semantic Web is a set of standards created by the World Wide Web Consortium (the international community that develops the standards underpinning the web) to enable progress of the web from a web of pages to a web of data. (See: http://www.w3.org/standards/semanticweb/)

6   S-map web service

This section demonstrates the use of an ISO/OGC-based spatial data infrastructure to the web enablement of the DESC. It will use a Web Feature Service and an S-map GML application schema to serve relevant information from the S-map database. While the following should not be treated as a final specification, it is intended to define a robust means of providing data to the DESC and to form the basis of future work culminating in deployment of a production system[7].

6.1   Use case

The S-map web service must meet a very simple use case focused on the reporting of relevant soil property information for an area defined by a client application. The client application is any web agent that can invoke a web feature service request and it may be operated in any context (e.g. from a web or desktop GIS or within a modelling application).

Step One: Client requests data

A client application requests polygons attributed with the soil type(s) and associated dairy effluent risk class(es) present relative to a particular location.

The location may be specified using a geometry defining any of the following (these are for illustrative purposes, the data service should support requests spatially constrained by any two-dimensional geometric primitive):

- The area within a simple, rectilinear bounding box
- A polygon enclosing a farm property boundary
- A user-defined irregular polygon
- A point location

Step Two: S-map web service returns data

The S-map web service returns the collection of attributed soil polygons that intersect the geometry provided in the client request.

For each polygon the service must provide the soil classification (according to the New Zealand Soil Classification) and dairy effluent risk category

Each property value is provided with meta-information describing the:

- Percentage of the polygon where the value can be said to apply

---

[7] To ensure that the application schema does not become a defactor standard before it is thoroughly tested, the GML schema document will not be published until testing begins.

- The uncertainty related to the assertion of the value

- The role the property value plays if it is a member of a collection of multiple values for the same property

- The order the property occurs in an array of multiple values of the same property, playing the same role (if applicable)

The definitions for qualitative property values will be readily accessible by the client application with definitions appropriate to the target user-groups (e.g. a technical definition and an explanatory definition for non-scientists).

6.2   Constraints

To ensure that the S-map web service aligns with the LINZ New Zealand SDI the following constraints have been placed on the development work.

- The solution must reuse existing standards wherever possible, in particular an information model based on the ISO/TC211 General Feature Model and its OGC implementation specifications (principally ISO19136 – the Geography Mark-up Language).

- The solution must not adopt data formats or structures unique to a particular software vendor or implementation of their tools (physical file format or application data artefacts).

- The solution must use OGC Web Feature Services that support version 2.0 of the WFS implementation specification.

- Content must, wherever possible, be restricted to structured quantitative values and references to members of SKOS controlled vocabularies; free-text property values are a last resort.

- The model must support data products or raw data as required by stakeholders, but this requirement is balanced against a need to allow sufficient flexibility to apply the data to a number of applications. As such, some post-processing of raw data may be required by client applications to get the derived data that they alone need.

6.3   Architecture

Figure 3 shows a high-level architecture of the proposed solution, as constrained by Section 6.2, and the relationships between components of the solution and client devices and modelling applications. The figure is greatly simplified for explanatory purposes and each component may represent a number of components and parties in reality. The components are:

S-map Database: the data primary store of for the New Zealand national soils database – created and maintained by Landcare Research soil and information scientists. The database incorporates digital soil surveys where these are sufficient; new mapping; a range of core, derived and interpreted soil properties; and linkages with the National Soils Database (NSD).

S-map Online: the web portal maintained by Landcare Research and designed to provide entry-level access to S-map data to as broad an audience as possible.
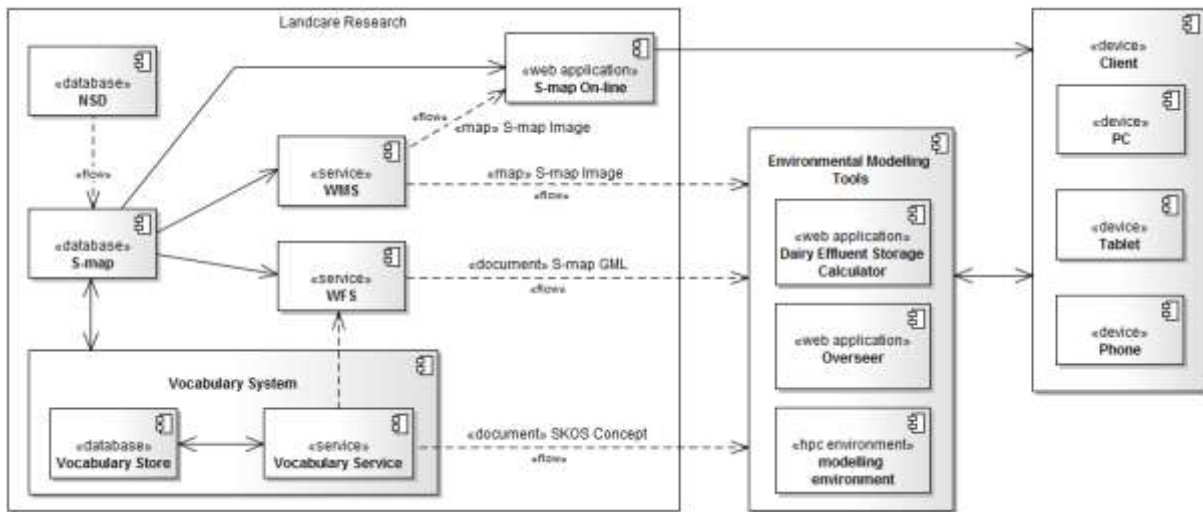

Figure 3 Simplified map of the interaction of modelling tools with the Landcare Research soil data web services. Components are described in the accompanying text. Solid lines represent connections between components and the direction of data or transaction flow between components. Dashed lines represent the flow of information between web services and client applications and are annotated with the type of data being transferred.

WFS – Web Feature Service:See Section 6.1.2: the web service that delivers raw S-map data as GML features structured according to the S-map GML schema.

WMS – Web Map Service:See section 6.1.2: a complementary web service that provides maps of the S-map data for use as layers in the client application's map view.

Vocabulary system: See Section 6.1.3: a complementary web service that provides values for vocabulary terms used in the S-map GML document for the WFS.

Environmental model tools The suite of client applications that may use data from the soil and vocabulary web services. An application may be a specific tool (e.g. DESC or Overseer) or may be a data repository in, for example, a high-performance computing environment where data are cached for processing by large or complex models.

Clients: Devices that allow end-users to interact with other components in the system. They typically bind to web applications over the Web (using the HTTP protocol). A client device may take a number of forms including, but not limited to, desktop PCs, tablets or smart phones.

6.4   Information model

A GML version 3.2 application schema has been defined using the AuScope FullMoon GML modelling tool (Golodoniuc & Cox 2010). Figure 4 is a graphical view of the data model and definitions of classes and properties are included in Table 2. Figure 5 shows the external data models reused by the application schema. In summary, the model defines a single feature type – SoilMappedFeature – to be returned by the WFS[8]. Each instance of a SoilMappedFeature describes a single S-map soil polygon or grid cell and will have one or more SoilProperty values. These may consist of a single asserted property value, or an array of values for the same property.

The schema makes use of the Simple Components defined in the Sensor Web Enablement (SWE) 2.0 Common Data Model (Robin 2011). SWE 2.0 was chosen as it provides data types to describe single and range Quantity and Category (controlled vocabulary reference) values for soil properties. The data types also include a facility for specifying simple estimates of uncertainty (quality) for each property value. No attempt has been made to restrict the available SWE data types, so temporal, count and Boolean data-types are available if necessary.

The model follows a soft-typing approach to allow flexibility of content, like the provision of new properties, without the changes to the structure of the data being served. This is possible as soft-typing does not explicitly encapsulate all possible features and their properties in distinct model classes; instead it uses a few simple classes that may be reused for multiple entities and properties. In this model the SoilMappedFeature is the only feature of interest, and each SoilProperty type is distinguished using a 'property' attribute whose values reference to a formally defined vocabulary of property types. The vocabulary – encoded as SKOS – shall list all available property types and hold their definitions.

---

[8] Note that use of GML 3.2 requires that the data be served from version 2.0 WFS.
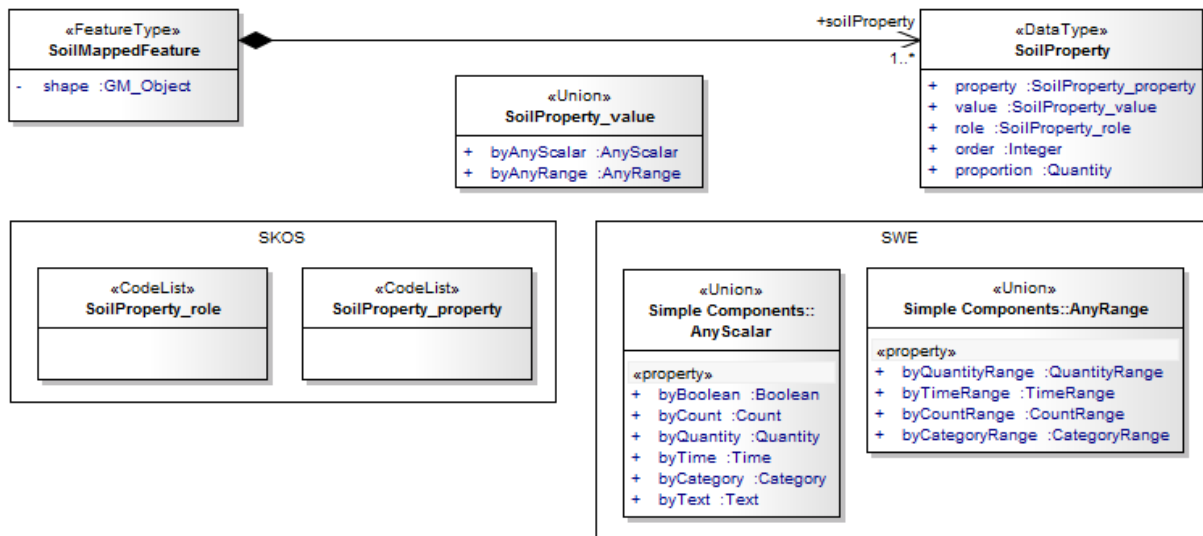
Figure 4 UML diagram for the proposed S-map GML application schema. Class, property and stereotype (text in double angle quotation marks, e.g. «FeatureType») definitions are included with Table 1. The SKOS system boundary encloses controlled vocabularies maintained in a separate vocabulary repository exposed using Semantic Web services. The SWE system boundary encloses the Sensor Web Enablement data type groups used for soil property values.
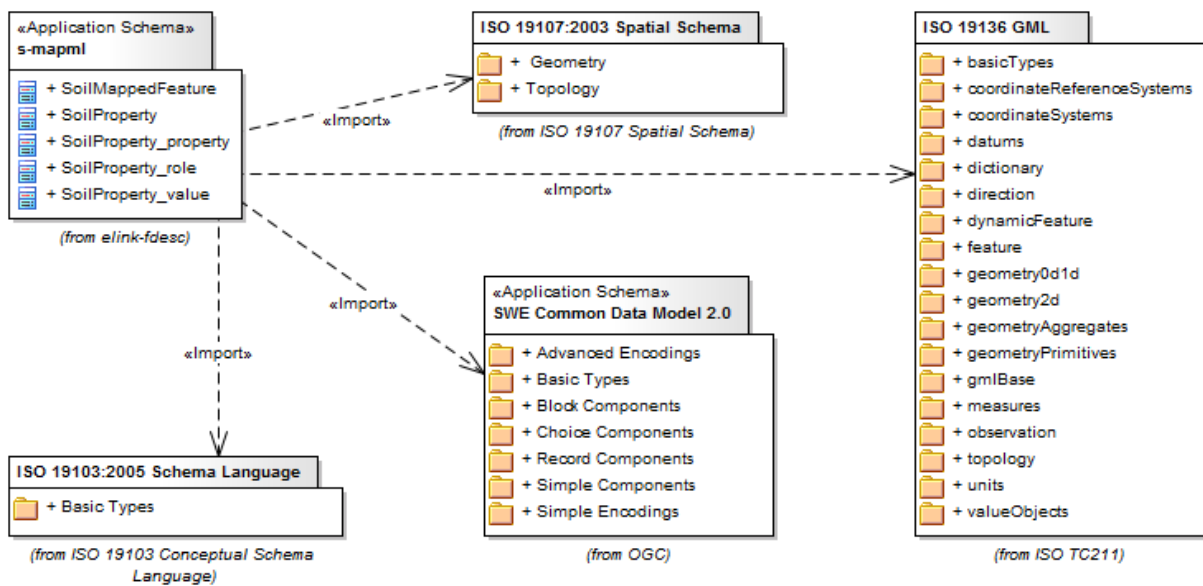

Figure 5 UML package dependency diagram showing the external packages reused by the S-map GML application schema. It is a GML 3.2 application schema (package ISO 19136 GML) using geometry primitives as defined in the ISO19107:2003 Spatial Schema and encoding property values using the Simple Components from the SWE (Sensor Web Enablement) 2.0 data model.

Table 2 Summary of S-map GML classes and properties.

Element	Type	Description
SoilMappedFeature	«FeatureType»[1]	A spatial feature enclosing a mapped, interpreted or observed occurrence of one or more soil units
soilProperty	SoilProperty	A collection (optionally ordinal – see SoilProperty.order) of values for Soil properties that have been measured or inferred for the extent of the SoilMappedFeature
shape	GM_Object (GML)	A spatial feature enclosing a mapped, interpreted or observed occurrence of one or more soil units
SoilProperty	«DataType»[2]	A data-type holding structured values of SoilProperties
property	SoilProperty_property	The soil property (e.g. 'dairy effluent class'). Taken from a catalogue of feature property types
value	SoilProperty_value	The value of the property. May be quantitative (a single value or value range pair) value encoded using as SWE Quantity; or qualitative (reference to a dictionary/vocabulary concept) encoded using a SWE Category. SWE quality values, a property common to both SWE Quantities and Categories, are used to provide a simple representation of the uncertainty for the value
role	SoilProperty_role	Where multiple instances of the same property are present for a SoilMappedFeature this describes the role the property instance plays – e.g. 'average' or 'worst'
order	Integer (GML)	Multiple instances of the same property, playing the same role form arrays. For ordinal arrays, the order property specifies their position. Default value is 0 – no ordering
proportion	Quantity (SWE)	The proportion of the area of the geometry that the property value may apply to
SoilProperty_property	GML «CodeList»[3]	Placeholder for a SKOS ConceptScheme containing terms defining available soil property types
SoilProperty_role	GML «CodeList»[3]	Placeholder for a SKOS ConceptScheme containing SoilProperty.role terms
SoilProperty_value	GML «Union»[4]	Reusable group providing a choice of data components from SWE
byAnyScalar	GML «Union»[4]	Group of references to SWE scalar data types: Boolean, Count, Quantity, Time, Category and Text
byAnyRange	GML «Union»[4]	Group of references to SWE Range data types: QuantityRange, TimeRange, CountRange and CategoryRange

[1] «FeatureType»: a General Feature Model/GML feature type (this is the entity served by the WFS).

[2] «DataType»: a structured data type for property values.

[3] «CodeList»: an enumeration of string values and a pattern – empty CodeLists represent externally maintained vocabularies and individual property values are specified by reference to their published location.

[4] «Union»: a choice group whose members are GML Objects or Features, or objects corresponding to DataTypes defining alternatives for a property value.

6.5   Interaction between client and WFS

An example of how the DESC would interact with the S-map WFS is outlined below. The WFS interface only specifies possible data transactions and makes no demands on the order in which they are invoked. This is left to the client to determine based on the broader task at hand, and the following is therefore only an example of many possible workflows.

1. The user defines area of interest in client application, e.g.

   - By selecting a property polygon on a map; or

   - Drawing a polygon around the area of interest.

2. The application submits a GetFeature request to the WFS, specifying a spatial filter using the selected area of interest. See Appendix 1 for a selection of sample spatial filter requests.

3. The WFS responds with an S-map GML document containing the features that met the spatial filter criteria (or an empty feature collection if no match was made). See Appendix 1 for a sample response.

4. The DESC application parses the response and incorporates the data into the model. How the document is parsed is entirely up to the client .Two likely (but by no means the only) options are:

   - DESC uses a data transformation tool to convert the GML into the data format most appropriate to the application, e.g. a spatial database table or GIS file format. This will usually involve picking only a subset of the data provided in the response. The process may involve requests to the vocabulary service to get labels or definitions for any terms used as (SWE Category) property values.

   - DESC uses an XML parser to traverse the document, targeting and retrieving property values using the XPath/XQuery XML querying techniques.

7   Conclusions

There is a clear need for the provision of soil data to a suite of environmental modelling tools that includes the DESC. This increasing demand, coupled with the ongoing growth of the S-map database, makes the benefits of real-time access to data clear. Having the most up to date available soil information will significantly improve the quality of output from the tools.

The conversion of modelling tools like the DESC to web applications places an onus on Landcare Research to provide their data in a web-accessible manner. Rather than working towards providing a suite of application-specific data products, Landcare Research proposes to deploy reusable data services that will be stable over long periods and integrate with other environmental data services. Stability, reusability and integration are best achieved by providing standardised query interfaces and data models for the ensuing responses.

The ISO/OGC suite of data and web service standards provide a context for standardisation, and are being increasingly adopted by environmental data communities as the basis for spatial data infrastructures (including the LINZ-defined New Zealand SDI). Communities of

academic and government researchers have developed soil information models in this context and these should be reused wherever possible. Any data delivery project undertaken by Landcare Research should monitor the development of GML application like SoilML and adopt them if they prove suitable to the needs of the organisation and its stakeholders.

Despite the current unavailability of a finalised community data model, this project has shown that a simple information model can be developed to deliver qualitative and quantitative soil property information. Example responses have been manually created and encode the necessary information to the satisfaction of Landcare Research soil and information scientists. They should be able to be parsed by client applications who have made spatially constrained requests for information. However, this can only be proven by a trial by DESC developers so it is recommended that a test deployment is undertaken in the near future.

The deployment and maintenance of nodes in the infrastructure comes at a cost for both human resources and hardware. Careful consideration must be given to the governance and funding models – if they are not addressed the infrastructure will be unmanaged, unstable and unsupported, and will decay to the point that it will ultimately fail. Ensuring that the services are deployed in a widely used SDI will help increase the case for ongoing support, but client agencies must be aware of the costs of ongoing service provision, and may need to make ongoing maintenance contributions in addition to funding the initial creation of the data services. There must also commitment to an ongoing dialogue between service providers and consumers to ensure the continued provision of relevant data. It is unclear how these issues will be resolved but they must be promptly addressed.

8   Recommendations

The solution has only been tested through peer analysis and must be tested against the needs of the DESC itself. It is recommended that a test deployment of the web data delivery platform is implemented, either in the context of Envirolink or another body of work. Appendix 2 summarises aspects of the proposal that are known to need evaluation. They include:

- An evaluation of the suitability of the data types used in the information model and the associated representation of uncertainty
- The consequences for the soft-typing approach adopted by this solution
- The possibility of returning data using formats other than GML
- The publication of alternative vocabularies to help explain the terminology to non-technical users of the system

Should the test deployment be successful then a production platform should be implemented, either under the auspices of an Envirolink tools project, or as a project co-funded by Landcare Research and relevant research and industry partners.

9 Acknowledgements

10 References

de la Beaujardiere J 2010. OpenGIS Web Map Server Implementation Specification http://portal.opengeospatial.org/files/?artifact_id=14416

Golodoniuc P, Cox S 2010. Geospatial information modelling for interoperable data exchange - application schema modelling: from concept to implementation. Proceedings of the 2010 IEEE Sixth International Conference on e-Science. Pp. 102–105.

Hewitt A, Barringer J, Claydon J, Campbell S, Carrick S, Gray C, Lynn I, Medyckyj-Scott D, Ritchie A, Webb T 2012. National Soil Database utility: a review and process for development. Landcare Research Contract Report LC1033 prepared for the Marborough District Council.

Horne D. Bretherton M, Hanly J, Houlbrooke D, Roygard J 2011. The FDE storage calculator – an update. In: Currie LD, Christensen CL eds Adding to the knowledge base for the nutrient manager. Occasional Report No. 24. Fertilizer and Lime Research Centre, Massey University, Palmerston North. http://flrc.massey.ac.nz/ publications.html.

Houlbrooke DJ, Horne DJ, Hedley MJ, Snow VO, Hanly JA 2004. A review of literature on the land treatment of farm dairy effluent in New Zealand and its impact on water quality. New Zealand Journal of Agricultural Research 47: 499–511.

Houlbrooke DJ, Laurenson S, Carrick S 2011. Categorising the environmental risk from land application of liquid wastes based on soil properties. AgResearch report prepared for the Marborough District Council. 49 p.

INSPIRE Thematic Working Group SOIL 2011. D2.8.III.3 INSPIRE data specification on SOIL – draft guidelines http://inspire.jrc.ec.europa.eu/documents/Data_Specifications/INSPIRE_DataSpecification_SO_v2.0.pdf

ISO 2002. ISO 19101. Geographic information – Reference model. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26002

LINZ 2007. Understanding our geographic information landscape – A New Zealand geospatial strategy. http://www.geospatial.govt.nz/sites/default/files/assets/Geospatial-Strategy/nz-geospatial-strategy-2007.pdf

Montanarella L, Wilson P, Cox S, McBratney A, Ahamed S, MacMillan B, Jacquier D, Fortner J 2010. Developing SoilML as a global standard for the collation and transfer of soil data and information. Geophysical Research Abstracts 12 http://eusoils.jrc.ec.europa.eu/esdb_archive/eusoils_docs/Poster/montanarella_EGU2010_XML.pdf

NZGO 2011. Spatial data infrastructure cookbook (version 1.1) http://www.geospatial.govt.nz/sites/default/files/SDI_Cookbook+V1-1+171111_0.pdf

Portele C 2007. OpenGIS Geography Markup Language (GML) encoding standard. http://portal.opengeospatial.org/files/?artifact_id=20509

Pourabdollah A, Leibovici DG, Simms DM, Tempel P, Hallett SH, Jackson MJ 2012. Towards a standard for soil and terrain data exchange: SoTerML Computers and Geosciences. http://dx.doi.org/10.1016/j.cageo.2011.11.026.

Robin A 2011. OGC SWE Common Data Model Encoding Standard. http://portal.opengeospatial.org/files/?artifact_id=41157

Vretanos P 2010. OpenGIS Web Feature Service 2.0 interface standard http://portal.opengeospatial.org/files/?artifact_id=39967

Appendix 1 – Sample web service requests and responses

The following are examples of requests that may be made to a S-map WFS for Soil Occurrence data. The requests assume a version 2.0 WFS using version 2.0 of the OGC Filter Encoding Specification. Note that WFS 2.0 allows the publication of Stored Queries for commonly used requests.

WFS Request One: Bounding Box

Filtered WFS request for smap:SoilMappedFeatures having extents (smap:shape values) that intersect the bounding box specified by the gml:Envelope.

Option 1. XML Request (submitted to the WFS as an HTTP POST request):

```xml
<?xml version="1.0" encoding="UTF-8"?>
<wfs:GetFeature service="WFS" version="2.0.0" outputFormat="application/gml+xml; version=3.2"
                xmlns:wfs="http://www.opengis.net/wfs/2.0" xmlns:gml="http://www.opengis.net/gml/3.2"
                xmlns:fes="http://www.opengis.net/fes/2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:smap="http://soil.example.org.nz/xmlns/s-map/0.1"
                xsi:schemaLocation="http://www.opengis.net/wfs/2.0 http://schemas.opengis.net/wfs/2.0/wfs.xsd
                                    http://www.opengis.net/gml/3.2 http://schemas.opengis.net/gml/3.2.1/gml.xsd">
  <wfs:Query typeNames="smap:SoilMappedFeature">
    <fes:Filter>
      <fes:BBOX><gml:Envelope srsName="urn:ogc:def:crs:EPSG::2193">
        <gml:lowerCorner>177.465 -38.335</gml:lowerCorner><gml:upperCorner>177.495 -38.305</gml:upperCorner>
      </gml:Envelope></fes:BBOX>
    </fes:Filter>
  </wfs:Query>
</wfs:GetFeature>
```

Option 2. URL Key Value Pair Request (submitted to the WFS as an HTTP GET request – split across multiple lines for clarity):

```
http://{host}/ows?service=WFS&version=2.0&request=GetFeature&typeName=smap:SoilMappedFeature
                &namespace=xmlns(smap=http://soil.example.org.nz/xmlns/s-map/0.1)
                &BBOX=177.465,-38.335,177.495,-38.305
```

WFS Request Two: Dairy Effluent Class

Filtered request for smap:SoilMappedFeatures that have dairy effluent risk property values equal to 'D'.

Option 1. XML Request (submitted to the WFS as an HTTP POST request):

```xml
<?xml version="1.0" encoding="UTF-8"?>
<wfs:GetFeature service="WFS" version="2.0.0" outputFormat="application/gml+xml; version=3.2"
                xmlns:wfs="http://www.opengis.net/wfs/2.0" xmlns:gml="http://www.opengis.net/gml/3.2"
                xmlns:fes="http://www.opengis.net/fes/2.0" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
                xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:smap="http://soil.example.org.nz/xmlns/s-map/0.1"
                xmlns:xlink="http://www.w3.org/1999/xlink"
                xsi:schemaLocation="http://www.opengis.net/wfs/2.0 http://schemas.opengis.net/wfs/2.0/wfs.xsd
                                    http://www.opengis.net/gml/3.2 http://schemas.opengis.net/gml/3.2.1/gml.xsd">
  <wfs:Query typeNames="smap:SoilMappedFeature">
    <fes:Filter>
      <fes:And>
        <fes:PropertyIsEqualTo>
          <fes:ValueReference>smap:soilProperty/smap:SoilProperty/smap:property/@xlink:href</fes:ValueReference>
          <fes:Literal>http://soil.example.org.nz/propertytype/s-map/0.1/dairyEffluentRisk</fes:Literal>
        </fes:PropertyIsEqualTo>
        <fes:PropertyIsEqualTo>
          <fes:ValueReference>smap:soilProperty/smap:SoilProperty/smap:value/swe:Category/swe:identifier</fes:ValueReference>
          <fes:Literal>http://soil.example.org.nz/concept/s-map/dairy-effluent-risk/d</fes:Literal>
        </fes:PropertyIsEqualTo>
      </fes:And>              [... continues over page ...]
```

```
        </fes:Filter>
    </wfs:Query>
</wfs:GetFeature>
```

Option 2. URL Key Value Pair Request (submitted to the WFS as an HTTP GET request):

XML `<fes:Filter>` elements may be appended to a URL key value pair request (see previous example) however, due to the length of the resulting filter string they become unwieldy and are not recommended.

WFS Response

An example response to a WFS GetFeature request for a smap:SoilMappedFeature. It is retuned as a member of a WFS Feature Collection. There is a single soil name value and two dairy effluent risk values (one domiant and one worst possible value). All are encoded as swe:Category values. All swe:identifier and xlink:href URI values refer to values in a SKOS Vocabulary and the URI should dereference to a SKOS Concept fully defining the term - the choice of a raw data or HTML (for requests made in a web browser) representation should be available.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<wfs:FeatureCollection numberMatched="1" numberReturned="1" timeStamp="2012-05-01T00:00:00.000Z"
  xmlns:smap="http://soil.example.org.nz/xmlns/s-map/0.1" xmlns:gml="http://www.opengis.net/gml/3.2"
  xmlns:swe="http://www.opengis.net/swe/2.0" xmlns:xlink="http://www.w3.org/1999/xlink"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:wfs="http://www.opengis.net/wfs/2.0"
  xsi:schemaLocation="http://soil.example.org.nz/xmlns/s-map/0.1 http://soil.example.org.nz/schema/0.1/s-map.xsd
                      http://www.opengis.net/wfs/2.0 http://schemas.opengis.net/wfs/2.0/wfs.xsd">
 <wfs:member>
  <smap:SoilMappedFeature gml:id="smap.SoilMappedFeature.987654321">
    <gml:identifier codeSpace="http://www.landcareresearch.co.nz">
        http://soil.example.org.nz/feature/s-map/SoilMappedFeature/987654321</gml:identifier>
    <smap:soilProperty>
      <smap:SoilProperty gml:id="smap.soilproperty.987654321.1">
        <smap:property xlink:href="http://soil.example.org.nz/propertytype/s-map/0.1/soilName" xlink:title="soil name"/>
        <smap:value>
          <swe:Category id="smap.value.1">             [... continues over page ...]
```

```xml
        <swe:identifier>http://soil.example.org.nz/concept/nzsc/templetonf-9</swe:identifier>
        <swe:label>Templetonf (Sib 9)</swe:label>
        <swe:quality>
          <swe:Category id="smap.value.quality.1">
            <swe:identifier>http://soil.example.org.nz/concept/s-map/confidence/low</swe:identifier>
            <swe:label>low</swe:label>
          </swe:Category>
        </swe:quality>
      </swe:Category>
    </smap:value>
    <smap:role xsi:nil="true" nilReason="inapplicable"/>
    <smap:order>0</smap:order>
    <smap:proportion>
      <swe:Quantity id="smap.proportion.987654321.1">
        <swe:uom code="%25" xlink:href="http://www.opengis.net/def/uom/UCUM/0/%25" xlink:title="percent"/>
        <swe:value>70</swe:value>
      </swe:Quantity>
    </smap:proportion>
  </smap:SoilProperty>
</smap:soilProperty>
<smap:soilProperty>
  <smap:SoilProperty gml:id="smap.soilproperty.987654321.2">
    <smap:property xlink:href="http://soil.example.org.nz/propertytype/s-map/0.1/dairyEffluentRisk"
                   xlink:title="dairy effluent risk"/>
    <smap:value>
      <swe:Category id="smap.value.2">
        <swe:identifier>http://soil.example.org.nz/concept/s-map/dairy-effluent-risk/d</swe:identifier>
        <swe:label>D</swe:label>
      </swe:Category>
    </smap:value>
    <smap:role xlink:href="http://soil.example.org.nz/concept/s-map/soil-property-role/dominant" xlink:title="dominant"/>
    <smap:order>0</smap:order>
    <smap:proportion>              [... continues over page ...]
```

```xml
          <swe:Quantity id="smap.proportion.987654321.2">
            <swe:uom code="%25" xlink:href="http://www.opengis.net/def/uom/UCUM/0/%25" xlink:title="percent"/>
            <swe:value>70</swe:value>
          </swe:Quantity>
        </smap:proportion>
      </smap:SoilProperty>
    </smap:soilProperty>
    <smap:soilProperty>
      <smap:SoilProperty gml:id="smap.soilproperty.987654321.3">
        <smap:property xlink:href="http://soil.example.org.nz/propertytype/s-map/0.1/dairyEffluentRisk"
                       xlink:title="dairy effluent risk"/>
        <smap:value>
          <swe:Category id="smap.value.3">
            <swe:identifier>http://soil.example.org.nz/concept/s-map/dairy-effluent-risk/b</swe:identifier>
            <swe:label>B</swe:label>
          </swe:Category>
        </smap:value>
        <smap:role xlink:href="http://soil.example.org.nz/concept/s-map/soil-property-role/worst" xlink:title="worst"/>
        <smap:order>0</smap:order>
        <smap:proportion>
          <swe:Quantity id="smap.proportion.987654321.3">
            <swe:uom code="%25" xlink:href="http://www.opengis.net/def/uom/UCUM/0/%25" xlink:title="percent"/>
            <swe:value>30</swe:value>
          </swe:Quantity>
        </smap:proportion>
      </smap:SoilProperty>
    </smap:soilProperty>
    <smap:shape>
      <gml:Polygon gml:id="smap.shape.987654321" srsName="EPSG:2193" srsDimension="2">
        <gml:exterior>
          <gml:LinearRing>
            <gml:posList> 177.47 -38.33 177.47 -38.31 177.49 -38.31 177.49 -38.33</gml:posList>
          </gml:LinearRing>                [... continues over page ...]
```

```
            </gml:exterior>
          </gml:Polygon>
        </smap:shape>
      </smap:SoilMappedFeature>
    </wfs:member>
  </wfs:FeatureCollection>
```

Appendix 2 – Issues to be addressed during testing

The following aspects of the proposed S-map web data service need to be addressed in any test implementation.

- The suitability of the SWE data types as carriers of soil property values. They may not be appropriate if the scope of the services move beyond the reporting of data for a particular problem, nor may they provide a suitable means of representing the uncertainty associated with the provide property values. An evaluation of UncertML[9], and its integration into the ISO/TC211 suite of standards is recommended.

- The consequences for the soft-typing approach adopted by this solution. Soft-typing runs counter to the concept of self-describing services and data models as the specification of properties is deferred to the feature catalogue, as opposed to the XML schema. This has inherent weaknesses as what a service provides is not made explicitly clear in the information model, and while it is documented, this is done using standards that are not formally implemented. These weaknesses must be evaluated against the strength of soft-typing: that new properties may be introduced without changing the information model used to provide the data.

- The possibility of returning data using the Java Script Object Notation (JSON, or its spatially enabled variant, GeoJSON) or the Resource Description Framework (RDF) data formats. While the WFS specification mandates the provision of GML as an available output format, there are no restrictions on the use of other encodings. RDF or JSON may be more appropriate for some applications.

- The publication of scientific vocabularies using SKOS. This should include the development of alternative versions of the vocabularies to help explain the terminology to non-technical users of the system.

---

[9] http://www.uncertml.org/