EPA STRIVE Programme 2007-2013

Systems Analysis and Design

(2007-S-CD-1-S1)

ISIS Final Technical Report 14

Prepared for the Environmental Protection Agency
by
Teagasc and Cranfield University

Authors:

Hallett, S.H., Fealy, R.

ENVIRONMENTAL PROTECTION AGENCY

An Ghníomhaireacht um Chaomhnú Comhshaoil

PO Box 3000, Johnstown Castle, Co.Wexford, Ireland

Telephone: +353 53 916 0600 Fax: +353 53 916 0699

Email: info@epa.ie Website: www.epa.ie

ACKNOWLEDGEMENTS

This report is published as part of the Science, Technology, Research and Innovation for the Environment (STRIVE) Programme 2007-2013. The programme is financed by the Irish Government under the National Development Plan 2007-2013. It is administered on behalf of the Department of the Environment, Heritage and Local Government by the Environmental Protection Agency which has the statutory function of co-ordinating and promoting environmental research.

This project involved many staff and students based at Teagasc Johnstown Castle, Teagasc Athenry, Cranfield University and University College Dublin. Many thanks go to all the staff, PhD students, placement students and FAS placements who contributed to this project. We also wish to thank the Steering Committee, Dr Alice Wemaere (EPA), Ms Niamh Connolly (EPA), Dr Mark Gibson (Teagasc), Prof Peter Loveland, Dr Erika Micheli and Dr Arwyn Jones.

DISCLAIMER

The EPA STRIVE Programme addresses the need for research in Ireland to inform policymakers and other stakeholders on a range of questions in relation to environmental protection. These reports are intended as contributions to the necessary debate on the protection of the environment.

EPA STRIVE PROGRAMME 2007-2013

Published by the Environmental Protection Agency, Ireland

Details of Project Partners

Teagasc - Environment, Soils and Land-use Centre,
Johnstown Castle, Wexford, County Wexford, Ireland.
Tel.: +353 (0)53 9171200
E-mail: rachel.creamer@teagasc.ie

Teagasc – Spatial Analysis Group, Rural Economy and Development Programme.
Ashtown, Dublin 15, Ireland.
Tel.: +353 (0)1 8059953
E-mail: reamonn.fealy@teagasc.ie

Cranfield University - National Soil Resources Institute
Cranfield Campus, Cranfield, Bedfordshire, MK43 0AL, UK
Tel.:+44 (0)1234 750111
Email: t.mayr@cranfield.ac.uk

University College Dublin
School of Biosystems Engineering, Agriculture and Food Science,
Belfield, Dublin 4, Ireland.
Tel.: +353 (0)1 7162808
Email: nick.holden@ucd.ie

Executive Summary

The Irish Soil Information System (ISIS) project was established in 2008, following a comprehensive inventory of Irish soil data compiled by Daly and Fealy (2007) which highlighted that soil data coverage of Ireland was incomplete in both detail and extent. The ISIS project is funded under the Environmental Protection Agency STRIVE Research Programme 2007-2013 and co-funded by Teagasc. It was led by Teagasc with the participation of researchers from Cranfield University (UK) and University College Dublin. The overall objective of the ISIS project was to conduct a programme of structured research into the national distribution of soil types and construct a soil map, at 1:250,000 scale, which will identify and describe the soils according to a harmonised national legend. This map is now available in digital format and forms the basis of a new soil information system for Ireland (http://isis.teagasc.ie).

The ISIS project has utilised existing data and maps from the previous National Soil Survey (NSS) conducted by An Foras Talúntais (forerunner organisation to Teagasc). The NSS produced: mapping at 1:126,720 scale for 44% of the country; a General Soil Map of Ireland and a National Peatland map, both at 1:575,000 scale and other miscellaneous large scale mapping of experimental farms. In addition, more recent map products have been included such as the Indicative Soil and Subsoil mapping (Fealy and Green, 2009) with national coverage using GIS and remote sensing techniques.

Comparison of soil information at European scale has led to the requirement for the harmonisation and coordination of soil data across Europe, and, in light of the demands for soil protection on a regional basis within member states there is a growing need to support policy with a harmonised soil information system. The European Soil Bureau Network (ESBN) Technical Working Group dealing with Soil Monitoring and Harmonisation recommended a soil map of Europe at a scale of 1:250,000 as an economically feasible intermediate scale that can identify specific problems at regional scale (Montanarella and Jones, 1999).

The ISIS project adopted a combined methodology of utilising novel predicted mapping techniques in tandem with traditional soil survey applications. This unique combination at a national scale has resulted in the development of a new national soil map for Ireland. Building upon the detailed work carried out by the An Foras Talúntais (AFT) survey (known as *Terra Cognita*), the ISIS project generated soil-landscape models at a generalised scale of 1:250,000 for the counties of Carlow, Clare, Kildare, Laois, Leitrim, Limerick, Meath, Offaly, Tipperary South, Waterford, Westmeath, Wexford, West Cork, West Mayo and West Donegal. These soil-landscape models (also referred to as soilscapes) were used as the baseline data for statistical models (random forests, Bayesian belief networks and neural networks) to predict soil map units in counties where there was no map available (referred to as *Terra Incognita*). To validate the methodology, this work was supported by a 2.5 year field survey, in which 11,000 locations were evaluated for soil type, using an auger bore survey approach. These data were used to check the predicted soil mapping units (associations) for counties: Cavan, Dublin, East Cork, East Donegal, East Mayo, Galway, Kerry, Kilkenny, Louth, Monaghan, Roscommon, Sligo, Tipperary South and Wicklow, where a detailed soil survey map was not available. Where new soil information was generated, due to previously unknown combinations of soil-landscape units, profile pits were selected at representative locations across the country. These 225 pits were described and sampled in detail and were used to generate a new soil classification system for the country. The final product is a unique combination of new and traditional methodologies and soils data from both the AFT and the ISIS project. The final, soil association map of Ireland consists of 58 associations (excluding areas of alluvium, peat, urban, rock or marsh) that are made up from 213 soil series. Associated representative profile information is available in the online soil information system.

A key component of the ISIS project has been the development of a soil and land information system and associated public web site. This system has been designed to hold the complete set of information deriving both from the ISIS field programme and modelling activity, as well as the previously existing legacy soils information available for Ireland. Drawing on this information system, the web site is designed to hold and disseminate this information online both in cartographic and tabular form to stakeholders. Prior to this development, there has been no harmonised computerised system in place to hold and manipulate national Irish soils data. The information system therefore addresses the pressing need and requirement for a publicly-accessible, integrated IT framework based upon contemporary informatics standards to serve the many and varied stakeholders having an interest in soils information in Ireland.

THE IRISH SOIL INFORMATION SYSTEM (ISIS)
Work Package 5

System Analysis and Design

Version 5.01

Stephen Hallett[1], Réamonn Fealy[2], Adnan Younas[1], Caroline Keay[1], Ian Truckell[1], Liam O'Connor Hannon[2]

[1] National Soil Resources Institute
Department of Environmental Science and Technology
School of Applied Sciences,
Cranfield University, Wharley End,
Cranfield MK43 0AL, England UK

[2] Teagasc
(Irish Agriculture and Food Development Authority)
Spatial Analysis Unit
Rural Economy and Development Programme
Ashtown Campus, Dublin 15,
Ireland

Cranfield University and Teagasc

November, 2011

This work is part of the Irish Soil Information System (ISIS) Project, managed by Teagasc (the Irish Agriculture and Food Development Authority) and co-funded by the Environmental Protection Agency (EPA) of Ireland through their Science, Technology, Research and Innovation for the Environment (STRIVE) Programme, as part of the National Development Plan 2007-2013.

Executive Summary

ISIS Work Package 5 focuses on the design and implementation of the central 'core' Soil Information System for Ireland (SIS-Core), plus the ancillary data management tools developed to capture and present soils data to this repository. The SIS-Core will also hold and manipulate the data holdings following from the other wider project activities. The SIS-Core further represents the principle source from which users and stakeholders will draw upon for their future use of the results of the overall project activities. The design and implementation must therefore not only be 'fit for purpose' for the immediate project requirements, but to facilitate interoperability of the datasets, must also be consistent with the principles of Directive 2007/2/EC, Infrastructure for Spatial Information in the European Community (INSPIRE). The resultant system will reside within a dedicated computer system suitable for the management and administration of the information, and which will be able to provide a range of packaged data-oriented services to end-users.

This document lays out the technical systems analysis and design options open for the implementation of the soil information system itself, together with the preferred route forward. Included here is the anticipated database schema for the SIS implementation, and the identification of technical systems options for the underlying architecture and environment as well as their procurement and commissioning within the project office.

Table of Contents

Citation. This report may be cited as follows:
Hallett, S., Fealy, R., Younas, A., Keay, C., Truckell, I., O'Connor Hannon, L. (2011) The Irish Soil Information System (ISIS): System Analysis and Design, Version 5.0. Work Package 5 Deliverable. 21pp. NSRI/Teagasc.

Glossary

DMZ – Demilitarised Zone (area for public access webservers)
RDBMS – Relational Database Management System
SMU – Soil Mapping Unit
OGC – Open Geospatial Consortium
DAFF – Department of Agriculture, Fisheries and Food
INSPIRE – Infrastructure for Spatial Information in Europe
SIS – Spatial Information System
ISRIC - International Soil Reference and Information Centre
LandIS – Land Information System
NSRI – National Soil Resources Institute
WMS – Web Map Service
KML – Keyhole Markup Language

Introduction

ISIS Work Package 5 focuses on the design and implementation of the central 'core' Soil Information System for Ireland (SIS-Core), plus the ancillary data management tools developed to capture and present soils data to this repository. The SIS-Core further represents the principle source from which users and stakeholders will draw upon for their future use of the results of the overall project activities. The design and implementation must therefore not only be 'fit for purpose' for the immediate project requirements, but to facilitate interoperability of the datasets, must also be consistent with the principles of Directive 2007/2/EC, Infrastructure for Spatial Information in the European Community (INSPIRE). The resultant system will reside within a dedicated computer system suitable for the management and administration of the information, and which will be able to provide a range of packaged data-oriented services to end-users.

This document comprising, the component of the ISIS project deliverable document "WP5 D1 - Analysis and Design", and lays out the technical systems analysis and design options open for the implementation of the soil information system itself, together with the preferred route forward. This document is necessarily an evolving feature of the project and has been released in several versions as the designs iterate. The elements of the analysis and design phase include the following components:

1. Technical Systems Architecture options [network, hardware and software];
2. Database Schema [for both spatial and aspatial data];
3. Data Discovery and Metadata management;
4. File management for project files;
5. Procurement and commissioning.

Technical Systems Architecture

The SIS Technical Systems Architecture describes the technical environment and equipment necessary to operate the SIS satisfactorily. Where budgeted, this will necessitate the procurement of equipment. This will involve close collaboration with the wider ICT group in Teagasc, as well as with the ICT and data service groups in the EPA, given the need for dissemination of ISIS data via various existing and intersecting systems operated from there.

The system will be built on the ArcGIS Server GIS platform from ESRI, following the recommendations as per the report of Daly & Fealy (2006). This offers a range of design and application advantages, whilst also providing the basis for future extension and incorporation of advanced data interrogation and modelling tools. Underpinning the SIS system will be a spatially-enabled relational database management system (RDBMS). This will serve to hold and manipulate the variety of data types, scales and sources required, including attribute characteristics tables to hold the descriptive property elements required in the data schema. The RDBMS will also be used to store the spatial mapping tables holding the geometric map layers and associated legends.

The SIS configuration will entail three interconnected and dedicated hardware platforms. Firstly the database server, which will hold the Relational Database Management System (RDBMS). Secondly the GIS services server which will hold the 'ArcGIS Server' spatial software technology from ESRI Inc. Together this computing environment will store, manipulate, manage and disseminate the spatial and aspatial datasets generated and used by the SIS activities. Related to the key dissemination activities, a separate web-enabled GIS server will also be placed in the DMZ. A one-way data upload will take place from the back-office GIS server to place data on this public-facing server. It is considered that this configuration will allow an enhanced approach to security while minimising the complexity that might be involved in implementing alternative solutions such as a reverse proxy approach.

Activities within the Work Package Five extend beyond solely the soil information system alone. Appendix One lays out the set of these wider activities.

EPA Dissemination Tools

Once operational, the SIS significant datasets will be presented or passed to appropriate intersecting sites within the Irish EPA for dissemination. Transfer of data will be decided by the joint technical steering group as envisaged by recent inter-agency planning meetings and agreed in the proposed MOU between the EPA and Teagasc. The key resources there of interest include:

EPA Service	URL	Description
ENVISION	http://maps.epa.ie/	EPA ENVision online environmental map viewer. A web mapping tool for environmental data in the EPA. The likely destination for the 'significant' ISIS datasets.
ERC Secure Archive for Environmental Research Data (SAFER)	http://erc.epa.ie/safer/iso19115/displayISO19115.jsp?isoID=7	NSDB Datasets and GIS Maps - available from ERC's Secure Archive for Environmental Research Data (SAFER)
NSDB	http://erc.epa.ie/nsdb/	The Irish National Soils Database – a national baseline database of soil geochemistry including data point maps and spatial distribution maps of major nutrients, major elements, essential trace elements, trace elements of special interest and minor elements.

Provisional Definition of Significant ISIS datasets

The ISIS project will generate and involve a variety of transitory working, derivative and output datasets. From these a subset of core 'significant' datasets will be developed and finalised for dissemination for viewing by the public and for further research purposes. The composition of the significant data subset requires confirmation, but will to incorporate at least the following:

- Irish National Soil Map at 1:250,000 (attributed spatial map units);
- National Soil Map Legend;
- Modal soil profile descriptions.

Department of Agriculture Data Centre System Configuration

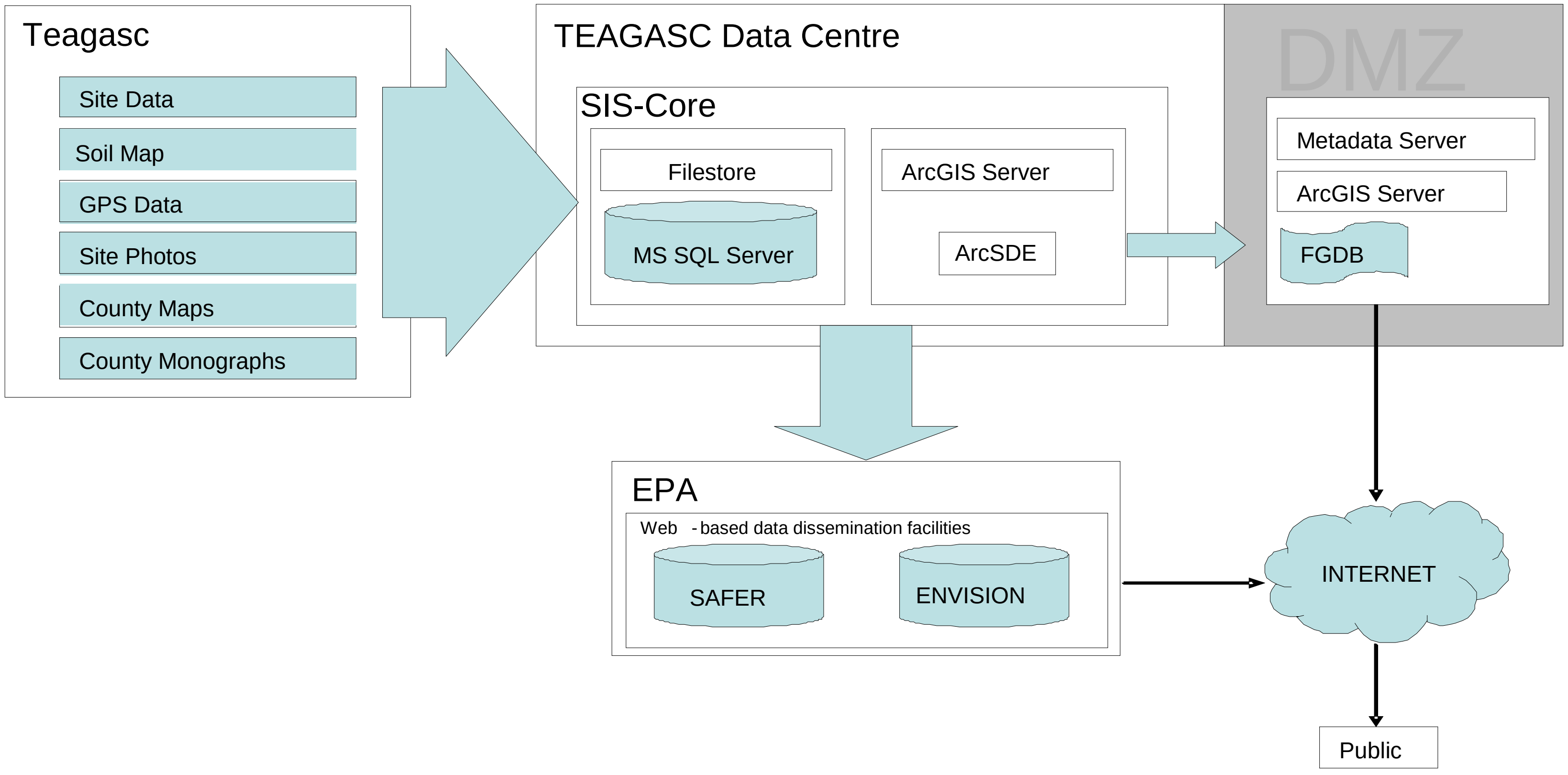
The proposed infrastructure will be housed in the Teagasc Data Centre. It is considered that an installation with the RDBMS and ArcServer software (which includes ArcSDE) installed on the same computer server as the ArcGIS Server installation would not be ideal in terms of capacity planning, data security or load-balancing. The development of the final design configuration for the ISIS has resulted from extensive consultation between the parties leading to agreement on a final deployment option. The preferred option for a national system such as the SIS requires the procurement and utilisation of two separate computer hardware servers for each of these tasks for the back-office implementation and a third server to be placed in the DMZ for the provision of datasets for public consumption over the Internet (Figure 1). A suitable specification for hardware is presented in Appendix Four. A further decision was required as to the specific RDBMS that will be procured and installed. There are three possible viable options available:

- Microsoft SQL Server Express
- Microsoft SQL Server Enterprise
- Oracle database from Oracle Inc.

Each of these options has technical, performance and financial implications. However, as Teagasc operates a Microsoft SQLServer operation this has been deemed the most strongly preferred option. This option is supported by the Teagasc ICT Department, who also have strong SQLServer capabilities.

The final system shall have a series of data layers built within the core SIS database and ArcGIS Server configuration, providing a series of data access and discovery tools to facilitate its usage, consistent with the principles of Directive 2007/2/EC, Infrastructure for Spatial Information in the European Community (INSPIRE), which came into force on May 15, 2007. A web mapping service (WMS) will be able to publish simple information feeds to Internet users of the core data themes held in the SIS. Such data can be fed directly into client GIS applications. For users without GIS, a simple web-based mapping application will serve to provide simple web-based mapping and summary tabular data from the SIS datasets. A further enhancement will be a data feed to produce output compatible with the popular Google Earth or Goole Map applications from Google Inc.

Fig. 1 Proposed architecture for the development of the SIS



Database Schema

The data structures presented follow the design principles of the Land Information System (LandIS), operated in the UK by Cranfield University as well as data schema derived from ISRICs standard World Soil Database structure (Keay *et al*, 2009). This information structure will be designed to hold the results of the field programmes, modal profile attributes for representative soil profiles, linkages to soil laboratory results for field samples and the spatial extents of regions of similar soil properties, or soil mapping units (SMUs). The overall design will further be guided by the outline data structure as described in Daly & Fealy (2006).

Appendix Two lays out a provisional data schema used to capture the modal soil profile descriptions from the existing published Soil Monographs accompanying the county soil maps. The profile descriptions are entered into individual county databases by different project team members. The databases are then collated into a single master database which forms the core of SIS. This data structure is designed to represent the Irish data circumstances, in terms of the data descriptions that exist. The overall design also upon the experience of operating the Land Information System held within the National Soil Resources Institute (NSRI), Cranfield University, which in turn is linked to wider European definitions of soil profile determinations for site and horizon. In other words, the design for the SIS is flexible enough to suit the application of the data captured in Ireland, whilst also making the datasets comparable to other national datasets at a European scale.

One of the challenges for the ISIS project is to seek to reduce time and potential errors for transferring observations of soil properties made in the field with the core soil information system. It is seen that field data recording will serve as a means to capture and validate data in the field on-site, for both auger and representative profile sites. Such data can then be transferred digitally to the main information system database. The Work Package 5 team, in consultation with members of the other relevant work packages have investigated two potential approaches. Firstly a handheld GPS unit running Windows Mobile 6 and running the ESRI ArcPad software tool with a customised data capture form, and secondly a ruggedized field laptop or tablet capable of running Windows XP Tablet edition and able to run a customised data capture tool written in MS Access. Both configurations have been tested and priced. The latter option was selected, and has been adopted due to perceived screen legibility, battery life and customisation options. As a result a number of 'Getac' units have been procured for the field teams (Fig 2). These will be used in combination with handheld bar code scanners (Fig 3) for logging soil samples taken specifically from representative profile sites, for which a series of pre-printed bar code labels have been prepared (Fig 4).

Fig. 2 Ruggedised Field Tablet computers

Fig.3 Field Barcode scanners

Fig.4 Field Barcode labels
Yellow=Bulk Density; White=Composite

Data Discovery and Metadata management

The SIS data repository will be designed to store and disseminate metadata, associated with the wider project datasets, and structured according to appropriate international standards including ISO19115 and Dublin Core.

The EPA 'SAFER' system employs a standardised template for the representation of metadata. SAFER will be used as one of the final means by which metadata will be disseminated, therefore the SIS metadata structure must be compatible with the SAFER schema.

Given the complexity of environmental data schemes such as that for soils, the design of the spatial data infrastructure will require linkage into various other thematic disciplines (e.g. climate, vegetation, land use, hydrology, pollutant emissions, geology) to facilitate its further application and usage. Depending on the specific data strategies, linkage particularly between regional or national data platforms will be needed to exchange harmonized information using OGC standards, and in fulfilment of metadata standards. In light of this the OGC standards-based 'GeoNetwork opensource catalog application' (GeoNetwork, 2009) is being evaluated and if appropriate, it is envisaged that this platform will serve as the Teagasc ISIS project SIS metadata repository.

Datasets held in the SIS will be grouped and identified for anticipated end-uses and user communities. Discovery-level Metadata will be collected for these groupings and published to appropriate 'clearing house' metadata servers, consistent with the requirements of ISO19115, to ensure the widest possible appreciation of the data holdings as well as the most appropriate usage of the data. Metadata captured in this way will be consistent with the wider international standards emergent for such description and discovery resources.

File Management for Project Files

There have built up a vast number of data files, documents and associated materials in the conduct of WP1,2 and 3 to date. Cranfield have provided an online repository, accessible via WebDav for ISIS Teagasc and UCD staff. Comprising now some 814Gb of data, this repository represents the progress of the project to date. These project files should be ordered and preserved for posterity as a record of the work undertaken. Accordingly, the design of the SIS Core architecture has had included a file server capacity. A recent mission to Ireland saw a copy of the current state of key documents transferred to this new facility. It should be noted that the magnitude of these files had not been forseen at the project outset – however, the Cranfield IT facilities kindly provided this capacity and whilst that was in place the specification of the new Teagasc servers was amended accordingly to receive these files. Full metadata descriptions will be drawn up for sensible groupings of these folders, ordered hierarchically on a work package basis.

Procurement and Commissioning

As advanced prototypes are put in place, so a series of system and deployment tests for the production system can be undertaken by selected representatives of the end-user community. Testing will be conducted both in-house, and via an invited group of expert users from external agencies, designed to achieve both an educational function to key stakeholders as well as a technical feedback route. Appendix Four presents the collated views of the stakeholder consultation on the final preferred implementation plan.

System Data

System data will be input within the project by the field survey teams and system operators, and will be made available ultimately as output via Internet mapping services, as well as in data digest form on CD-ROM. The implementation will provide tools to manage and manipulate the data holdings within the system. Datasets will be hosted on a suitable spatially-enabled relational database management system using ArcSDE technology from ESRI. The database system, through interaction with the ArcGIS Server, will provide a repository for spatial and aspatial data derived from the wider SIS activity.

The SIS shall contain data structures to permit the handling and management of the spatial and aspatial soils data gathered in the other work packages. The use of the spatially enabled relational database management system will enable the coupling within the data schema of the locational information with the representative descriptive data structures.

System Usage and Interoperability

In order to design and implement the SIS, a number of criteria need to be fulfilled:

- connectivity with inventory and monitoring systems at various spatial and temporal scales;
- connectivity and exchange of interoperable data with other regional, European and global services;
- link with specific projects related to data assimilation, modelling, regionalisation and methodological development;
- data quality control mechanisms and data harmonisation;
- assessment of soil data (about the state of soils → indicators);
- further development of existing data with improved future data availability.

Currently, INSPIRE develops the rules for developing geospatial infrastructures. These rules, oriented on the basis of Open Geospatial Consortium (OGC) standards, adapted in various International Standards Organisation (ISO) and European Committee for Standardization (CEN) rules about the management of geographic information, serve as the framework for developing geospatial web-based infrastructures such as SIS. The SIS will accordingly be capable of providing simple Web Mapping Services (WMS) (OGC, 2004) to reveal the data holdings. The simple web-mapping application proposed will act as a ‘consumer’ of the WMS to provide web-users with the means to interrogate and visualise the data. This will be integrated within the wider project website.It is anticipated that the core SIS shall include the following user-interaction capabilities:

SIS Component	**Description**
Viewer	Simple web-based tools providing access to mapping and tabular SIS datasets. Internally, these will draw upon the WMS component noted below.
WMS	Web mapping services providing Open Geospatial Consortium (OGC) compliant data for external users/consumers for the core SIS datasets via a data feed expressed using Geography Markup Language (GML).
KML	A service to export proprietary Keyhole Markup Language (KML/KMZ) for use of SIS datasets within the Google Maps, Google Earth application.
Metadata export	Metadata service provision of geo-spatial metadata (ISO19115 format) and aspatial metadata (Dublin Core format) for core SIS datasets.

References

Daly, K. & Fealy, R (2006) Digital Soil Information System for Ireland (2005-S-DS-22), Final Report. EPA. 66p.

GeoNetwork (2009) GeoNetwork Open Source catalog application. http://geonetwork-opensource.org/ [Accessed 18/5/09].

Keay, C.A., Hallett, S.H., Farewell, T.S., Rayner, A.P. and Jones, R.J.A. (2009) "Moving the National Soil Database for England and Wales (LandIS) towards INSPIRE Compliance", International Journal Of Spatial Data Infrastructures Research, 134-155.

Hallett, S.H., Jones, R.J.A., and Keay, C.A. (1996). Environmental Information Systems Developments for planning sustainable land use. Int. J. Geographical Information Systems, 10, 1, 47-64

OGC (2004) OGC Web Map Service Interface. Ed. Jeff de la Beaujardiere. OGC 03-109r1. 84pp.

Appendix One: Associated Documents and Actions

Work on ISIS WP5 has led to the development of a series of software tools and associated documentation outwith the scope of this document. These are recorded here, however, to provide a basis for recognising the breadth of activity associated with WP5 work. These are summarised in a separate document 'WP5 Deliverables' and mostly each have a separate document outlining the specifics.

1. **Project TWiki 'Intranet' site** at http://isis.dmz.cranfield.ac.uk. This web portal, restricted to computers within the Cranfield, DMU and Teagasc domains provides the resources for managing the project objectives, actions and calendars etc., plus providing a useful document sharing tool. Early developments on the TWiki also crucially provided tools for capturing and registering soil series from the county monographs.
2. **ISIS Legacy representative profile collection**. An early stage in the project involved capture and recording of representative profiles from each of the counties from the published county monographs. A database system was designed and deployed that allowed a file to be drawn up for each county. Once complete a further tool in MS Access allowed the collation of all the datasets so collected into one coherent database. This data will be extracted and placed into the SISCore information system.
3. **ISIS Soil series registration**. Another early tool was to extract and collate all the soil series defined in each of the county monographs. Accordingly a tool was devised on the TWiki to allow soil series to be recorded and categorised. This data will be extracted and placed into the SISCore information system.
4. **Handheld soil profiler.** A clear early objective was to be able to collect soils information directly in the field in electronic format. An early prototype system was developed in ESRI's ArcPAD tool to investigate how well as handheld unit could perform this task. Ultimately discarded due to the small screen size, this tool helped inform the development of the auger bore and representative profile systems on the Getac ruggedized laptops.
5. **Soil Auger field recording system**. A comprehensive tool for capturing site and horizon data for soil auger sites has been developed and documented in full in the associated system manual, describing in full the associated data schema.
6. **Representative profile field recording system**. A comprehensive tool for capturing site and horizon data for representative profile sites has been developed and documented in full in the associated system manual, describing in full the associated data schema.
7. **Photograph management tool.** A tool for managing and manipulating the many photographs taken at field locations has been provided together with full documentation, describing in full the associated data schema.
8. **Project Work - and Dataflows** are presented, documenting the flows of data and information around the project, considered a crucial task

due to the complexity of the programmes in place, the numbers of staff involved and the geographical disparities of the project.

9. **Data Schema Construction Tool**. Field data is captured in a series of MS Access databases – for auger bores and for representative profiles, as well as photographs etc. The ultimate residing place for this data is the SISCore database, which will sit in the Teagasc data centre, implemented on MS SQLServer. There is a need for a tool to facilitate the transfer of schemas and data from this collection of 'feeder' applications to the core database. The schemas may vary between those adopted in the standalone applications (described here) and the core information system. Therefore a mapping is required from one schema to the other, allowing for changes in data table design. Furthermore, this is a process that may need to be undertaken more than once, necessitating a clear audit trail. Accordingly a tool is developed to aid the construction of the SISCore data schema from these component parts as well as to then transfer the data.
10. **Metadata Management**. Provision of datasets and documentation has necessitated the production of a metadata recording system, in line with Deliverable D3 of WP5. Accordingly a system has been put in place on the TWiki tool to allow for INSPIRE-compliant metadata to be captured, recorded and reported on. This is available at: http://isis.dmz.cranfield.ac.uk/twiki/bin/view/ISIS/MetaData.
11. **Configuration of SISCore**. A series of three guidance documents are produced to facilitate the process of first adding spatial data themes into the SISCore spatial database (comprising MS SQLServer and ESRI ArcSDE) (document: AddingSpatialDataConnection), and then subsequently adding this data into a GIS project (documents: AddingArcGISServer and ImportingSpatialData respectively).
12. **Auger Bore Rule Processing - Queries Handbook**. Field data is captured in a series of MS Access databases – for auger bores and for representative profiles. The field data capture tool undertakes a degree of data validation at the point of data entry. However, a set of post-entry data verification rules need to be applied to check data integrity. This deliverable outlines the basis for undertaking this and presents a set of SQL scripts to effect a selection of rule checks.
13. **ISIS Auger Bore Table Preparation Tool**. The early county data was collected in what is now termed the Johnstown Castle database (or 'Julian database'). Some issues have become apparent in preparing these data for the validation checks for importing into the later 'Cranfield database'. These problems relate to the use of separator mask codes in the fields that store multiple values. On occasion these fields were completed whereby the separator codes were entered by hand, incorrectly. This document described the procedure for removing these issues in advance of the main set of validation checks. VBA code is also attached.

Further to these outputs, two other important services have been provided to, and facilitated within, the wider project. These are namely:

14. **File Server capacity at Cranfield.** A WebDav service has been provided by Cranfield IT facilities that is currently holding and serving up to Teagasc, UCD and Cranfield staff some 814Gb of project data. It is anticipated this will be transferred in full to Teagasc.
15. **Virtual Servers**. A number of virtual servers have been established at Cranfield to act as facsimiles of the anticipated Teagasc servers being procured and commissioned. The Cranfield servers have been installed with ArcGIS Server, ArcSDE and SQLServer and configured as 'mirrors' of the anticipated Teagasc environment.

Appendix Two: Data Dictionaries

The following data structure diagram and data dictionaries represent the provisional design employed for the soil auger and representative profile data capture system used in the field programme.

The following datafiles are drawn up as representing the provisional database schema for SIS-Core for receiving data from the field programmes. Initially, these tables relate only to the normalised schema for soil auger bore records and soil series descriptions. Other entities will be appended as the project progresses. These scripts have been implemented in the ISIS SIS-Core server.

HumanDetail.sql
HUmanID.sql
ISISIS_Data.MDF
ISISIS_Log.LDF
LandPosition.sql
LanduseclassDetail.sql
LanduseclassID.sql
landuseTypeDetail.sql
landuseTypeID.sql
ObservorGroupDetail.sql
ObservorGroupID.sql
PictureGroupDetail.sql
PictureGroupID.sql
Site_profile.sql
SlopeGroupDetail.sql
SlopeGroupID.sql
SubGroup.sql
SubSeriesCode.sql
VegetationClassgroupDetail.sql
VegetationClassGroupID.sql
Vuspceiesgroupid.sql
VuSpeciesGroupDetail.sql

SIS-Core Data Transfer tool

A tool is being developed based on a Java applet that allows the data elements originating from the field data capture tools to be mapped onto their equivalents in the SIS-Core schema. An XML file holds the mapping between the separate systems. This approach will allow and facilitate easy amendments to be made as the design evolves.

Appendix Three: Hardware Specifications

The following configurations represent the hardware specifications that form the provisional system design for the SIS. Included are basic specifications that can form guidelines for procurement for the system servers, desktop workstations and handheld GPS/dataloggers for the field.

The recommended server specification is:
A DELL™ PowerEdge™ R710 (SV1R710), having

- 2 Quad Zeon processors
- 24 Gigabyte RAM
- 2 Gigabit Network Interface with failover
- Smart-array controller with failover
- Power supply with redundant failover

The recommended hardware specification for a GIS workstation is:
A Dell T5400 Workstation *or equivalent* running Windows Vista Enterprise x64bit Operating System with the following specification:

- Dual Intel Xeon E5440 (2.83GHZ, 1333Mhz, 2x6MB,Quad Core) Processor
- 8GB DDR2 Quad Channel Memory
- Minimum of a dual 256mb nVidia Quadro NVS290 Graphics Card
- Minimum of 2 x 300GB (15000rpm) Hard Drives
- 16xDVD+/-RW + 16xDVD Drive
- 2 x 24in Widescreen displays

The recommended hardware specification for a GPS handset is:
Mobile software development will be tested on a Trimble Juno SB handheld GPS unit running Windows Mobile 6.1 and ESRI ArcPad v8 software. The GPS unit has the following specification:

- 533Mhz processor
- 128mb Ram and flash data storage
- Integrated digital camera
- Integrated Bluetooth and wireless LAN
- Integrated high sensitivity GPS/SBAS receiver and antenna
- 12 channels
- 2 to 5 meter accuracy after differential correction
- NMEA and SiRF protocol support
- Micro SD (microSDHC) card slot

The recommended hardware specification for a ruggedized field laptop/tablet is:

- Rugged Tablet PC
- IP rating of IP54 preferably IP65
- Screen size between 8 - 12", preferably > 10"
- GPS
- Windows Operating System that supports Microsoft Access

- Long battery life and ability to use additional batteries (Hot swap preferable)
- Sunlight readable display
- Touch screen capability
- Communications required for extracting data (USB, Wifi, bluetooth, 3G etc.?)
- Preferably 2GB RAM
- Processor >= 1.0ghz
- To include additional batteries, car/office chargers, case/holders, docking units.

Appendix Four: Stakeholder consultation on final preferred implementation

The development of the final design configuration for the ISIS has resulted from extensive consultation between the parties leading to agreement on a final deployment option. Five meetings in total have been held between the stakeholders with the express aim of agreeing the final design specification The potential design configuration options were initially discussed at a meeting between Teagasc and Cranfield members of the project team, Teagasc ICT staff and EPA representatives on 18/03/2009. During discussions, three options were initially envisaged for structuring the delivery of the final SIS and its components. The first two options were primarily statements of potential configurations at opposite ends of the potential spectrum of options. The first option discussed would see Teagasc effectively "box and ship" the systems at project completion and cease all further interaction with the developed system. The second option was the conceptual opposite to option 2 and envisaged Teagasc retaining full and exclusive ownership and management of the system as built at end of project. The third option was perceived as a collaborative effort, comprising joint ownership, between the EPA and Teagasc. The three options as discussed at the first meeting are described in summary below.

Option 1
On completion of the project a full handover to the EPA of deliverables including the developed databases would take place. The EPA would integrate the developed databases into their existing dissemination systems eg EnVision and SAFER. Teagasc would cease their involvement in the SIS on completion of the project.

Option 2
The SIS developed to facilitate the project would be designed with the longer term view of transitioning towards a role as a public dissemination tool for the final delivered SIS components. In this option, Teagasc would formally integrate the SIS and its development into its core ICT systems architecture. Responsibility for public dissemination via INSPIRE compliant web mapping facilities would rest with Teagasc.

Option 3
On completion of the project the SIS would be incorporated into the Teagasc ICT systems architecture. However the developed project outputs would be passed to the EPA for inclusion on their dissemination platforms with databases most likely going to EnVision and shapefile/spatial output going to the SAFER platform. Access to the data would be via both entry points, Teagasc and EPA. Teagasc would maintain the data and provide amendments where necessary. Updated data releases would be provided to the EPA with formal protocols developed to ensure the smooth operation oft this transfer and to maintain the synchronicity of the disseminated data. This option also allows for the possibility of the EPA dissemination being handled

by an online linking mechanism to the Teagasc SIS. Possibilities range from the simple URL linking out to the Teagasc online system to the provision of map services which would be consumed by the EPA dissemination systems. The design of this mechanism is not currently within the scope of the SIS project specification but the development of SIS could proceed in a manner to facilitate its development if this option was viewed as a preferred.

Given Teagasc's mandated role in the collection and maintenance of national soils data and the EPAs mandated role in national environmental data dissemination along with the project financial contribution from both organisations, the first and second (as numbered above) options were not deemed viable solutions for the infrastructure development and management. The third option proposing joint ownership was deemed to be a realistic and sustainable model for system development. Option 3 was therefore the preferred solution choice and agreement on a modified form was reached between the project partners and the EPA at a joint technical and management meeting on 19/01/2010.

Agreements arising out of the meeting which have been designed to support the core technical design aspects of the SIS will be formalised in the Memorandum of Understanding drawn up between Teagasc and the EPA and ultimately approved at Director level.