INSPIRING SCOTLAND

"An ideas forum and open competition to encourage and promote innovative thinking connecting the places, spaces and faces of Scotland"

SPATIAL DATA INFRASTRUCTURE TO INSPIRE SCOTLAND

S Ramage

1Spatial, Cavendish House, Cambridge Business Park, Cambridge, CB4 0WZ
Email: steven.ramage@1spatial.com

Member of the AGI Working Group for INSPIRE, OGC Europe Forum and ePSIplus contributor

Abstract

Scotland is uniquely positioned to become the leading regional entity in Great Britain. The importance of federalism is self-evident to those in the European Community. This paper examines the opportunity to use spatial data to create an effective regional entity in Scotland. The paper examines the experiences in Australia and Northern Ireland. It attempts to draw on the experiences particularly from Northern Ireland and draw some generic conclusions.

Keywords

Reuse
Spatial Data Infrastructures (SDI)
Spatial Data Quality
Semantics

The Context

Public services need reliable, accurate and timely information to manage services and account for performance.[1] Although the Audit Commission refers to data in general, the comments they make are pertinent to spatial data. The Commission goes on to state that a great deal of time and money is spent collecting data of dubious quality. Since the underlying data are used for major planning decisions as well as performance measurement, the need to demonstrate that the data are reliable is critical.

In recognising the volume of resources already spent on spatial data collection, the Commission has identified a key issue. These existing spatial data have been collected over the last 10 or 15 years. They will have registered to the Ordnance Survey base map of the time, almost certainly Landline, and will have included data from both public and non-governmental bodies. In all likelihood they will have been collected to produce maps bounded by sheet edges. They will have been collected by jurisdictions, some of which have passed into history, using data schemas that in some cases are, at best, partially understood by successor bodies. It is also likely that they will have been collected in proprietary geographic information systems. The re-collection cost of these data sets

[1] Improving Information to support decision making: standards for better quality data. *A framework to support improvement in data quality in the public sector.* (Audit Commission, March 2007).

across Europe was estimated in 2000[2] at €36bn. This is the challenge facing the Scottish Government as it seeks to create regional data fit for managing the economy of Scotland.

The Challenge

The challenge is similar to when the Digital Chart of the World (DCW) was produced. This project involved a team of people, reported to be between 200 and 400 at various stages of the project, bringing together data collected by a large number of bodies with different objectives. This was interoperability long before the term was ever used and on a scale that was unique and probably will never be seen again. Then DCW started to be used outside of the specification from which it was created, namely operational navigation charts. Why not? It was public domain. New issues became apparent about this exciting new resource. It was not fit for purpose and there was the subject of the ongoing maintenance for this tool. Indeed why should a dataset created for the purposes of medium altitude en route navigation by *dead reckoning visual pilotage* be of any value in assessing drainage basin characteristics?

These realisations led a committed group of people to create what has become ISO19113/19114 (for a neat summary of these see Chapter 15 in 'Spatial Data Quality', Shi et al. (2002); ISBN 0-415-25835-19). The opening sentence of ISO19113 sets the tone:

> '*Geographic datasets are increasingly being shared, interchanged and used for purposes other than their producers' intended ones'*.

The standard work was carried out under the auspices of ISO/TC 211 and then applied as a case study to the DCW project. We have the Scandinavians to thank for much of this work (see: http://www.nlh.no/ikf/gis/dcw/). By 1995 the initiative had run out of steam because only non-quantitative assessments of geographic datasets could take place. Quantitative assessments were not possible for large GIS datasets because of the lack of processing power. It is quantitative assessment that is really valuable for assessing logical consistency and positional accuracy, and as Jakobsson (ibid 2002) eloquently stated:

> '*Combining data sets that have no quality information can be very difficult or impossible*.'

Today the processing power required is available. The Open Geospatial Consortium (OGC) has carried out significant work on standards to complement the ISO work described above. The ISO work produced a set of measures that would form the basis of quantification of spatial data quality. They are reproduced in Table 1 (adapted from reference 1). The framework is in place. What remains? The flow of information across departmental, regional and national institutions is rapidly growing in size and complexity as legacy data is re-worked and as different data types (CAD/GPS/RFID) are included, causing larger data integration problems. The data are increasingly being subject to legal contract (the digital rights management requirement) and service level agreements and are required for performance measurement. There is a hugely complex supply chain in which spatial data is playing an escalating role serving these connected consumers and suppliers to form an SDI.

The role for the Scottish Government in this challenge can be obtained from the calls for

[2] PIRA

the current *eContentplus* programme. Projects will be supported, which build a collaborative community or consensus to deliver solutions that:

> *'…improve the usability and quality of existing digital content…and are geared towards innovation in organization and in deployment.*[3]*'*

Measure	Description
Accuracy	Data should be sufficiently accurate for its intended purposes, representing clearly and in sufficient detail the interaction provided at the point of activity. Data should be captured once only, although it may have multiple uses. Accuracy is most likely to be secured if data is captured as close to the point of activity as possible.
Validity	Data should be recorded and used in compliance with relevant requirements, including the correct application of any rules or definitions. This will ensure consistency between periods and with similar organisations.
Reliability	Data should reflect stable and consistent data collection processes.
Timeliness	Data should be captured as quickly as possible after the event or activity and must be available for the intended use within a reasonable time period.
Relevance	Data captured should be relevant to the purposes for which it is used. It may be necessary to capture data at the point of activity which is relevant only for other purposes, rather than for the current intervention. Quality assurance and feedback processes are needed to ensure the quality of such data.
Completeness	Monitoring missing, incomplete, or invalid records can provide an indication of data quality and can also point to problems in the recording of certain data items.

Table 1 Measures of Data Quality

Second phase SDIs

The majority of SDI's are at the first stage. A typical example is that described to the EuroGeographics - EuroSDR - AGILE joint workshop on quality assurance in geographic data production, February 2006 by Hunter et al.[4] from the Australian state of Victoria. Victoria is part of another federal entity. The Victorian state government already has an online property report. There are 80,000 downloads per month, the level of complaints concerning the quality of the information is so great that complaints are no longer being logged. One of the main issues concerns the dimensions of land parcels in the online version which do not coincide with the official certificate of title. The reason is that there are two versions of the parcel database, contravening the ISO accuracy measure.

There is a new generation starting to appear, based on creating a single source of the truth. Northern Ireland is one such entity.[5] The geographic information strategy for Northern Ireland has been branded as GeoHub NI. The Ordnance Survey of Northern Ireland (OSNI) has the key strategic aim to keep the mapping of Northern Ireland up-to-date in a timely, accurate and effective way. OSNI's Information and Communications Technology (ICT) strategy has determined that to enable OSNI to meet its business

[3] http://ec.europa.eu/information_society/activities/econtentplus/programme/workprogramme/index_en.htm
[4] http://www.eurogeographics.org/eng/documents/WSQ_4_Usability.ppt
[5] Building European Spatial Data Infrastructures. I. Masser (2007). ESRI, ISBN-10 1589481658

objectives, OSNI must provide mapping products and other data electronically to customers in a format that meets their specific needs, rather than only providing standard products. Northern Ireland's development of use of Geographic Information (GI) is generally believed to be slower than in other regions. There are a number of reasons that contribute to this situation:

- General lack of resources and cost of adopting new techniques
- Huge backlog of non-digital data and the availability of large-scale printed maps
- Centralised nature of government decision making
- Lack of awareness of the benefits of spatial data
- Existence of a culture that does not readily accept change
- Small geographic area.

In order to respond to this situation, Northern Ireland has become the first region within the United Kingdom to develop a GI strategy; although the Location Strategy is out for review at the time of writing this paper. The starting point was to address the lack of awareness. The next step, a typical first phase was to create a portal and make information available through the portal. The current phase is to build a collaborating community with an objective to increase the frequency with which information is updated. A key component in the implementation of this phase of GeoHub NI is the creation of a Geographic Information Hub or GeoHub, which will hold, make available and permit users to access government-held spatial information. The GeoHub provides centralised access to and management of key data and online GIS services through centralised, standards based storage and ubiquitous access. It also supports policy making through the availability of cross-departmental geographic information and is helping to develop the infrastructure to support the Northern Ireland Mapping Agreement (NIMA). It is based on OGC WMS/WFS standards.

Phase 1 of the Web Access project enabled OSNI to electronically push their spatial data to their customers. The second phase SDI is to enable OSNI customers to pull data from the OSNI database, streaming mapping on request to their desktop and enable other Government departments to hold their data in an OSNI Hub. This GeoHub must also integrate and exchange data with other systems within the Northern Ireland Civil Service (NICS), with Public Sector bodies outside the NICS, and with private sector organisations. This activity positions the OSNI Hub as a true Geographic Information Hub for Northern Ireland and enables OSNI to take the lead in the implementation of GeoHub NI. Digital Rights management capabilities allow the originator of the data to be credited. This provides the basis for OSNI to update their detailed topographic map base, founded on known business rules and quality measures. This produces up-to-date data for the benefit of the community.

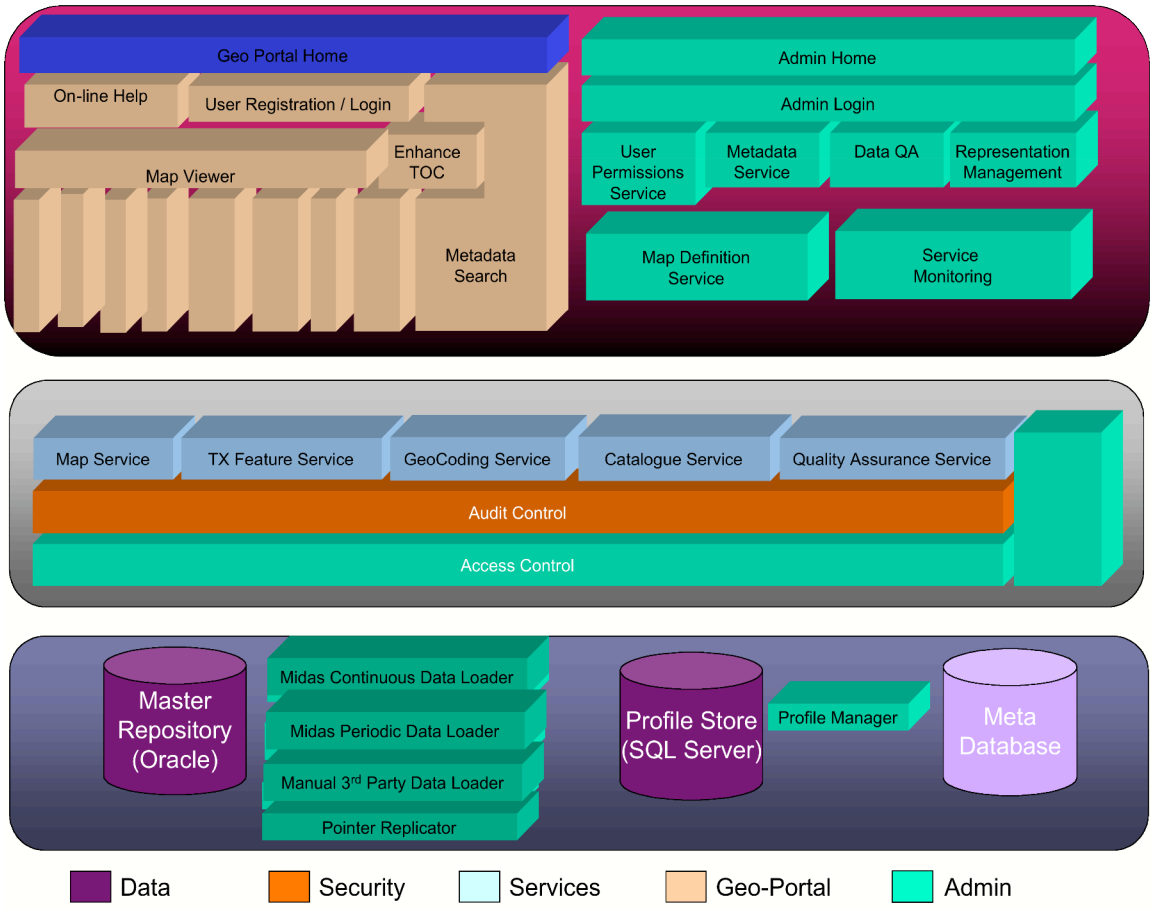

Figure 1: GeoHub architecture

In the GeoHub programme, members of the community do not need to own their own GIS, they can use the GeoHub's facilities based on OGC compliant technologies for Web Mapping, Web Feature Serving, and Web Registry (cataloguing) Services with an emphasis of metadata standards, management and storage spatial data. Functionality is now being provided for validating data, defining layers on the data, setting the access rights to data, and setting a default representation for the data. In line with the common approach of SDIs, GeoHub has been designed to be capable of accommodating future change to provide a central portal and a repository (both virtual and physical) for storage and exchange of data, to be accessible via a variety of web browsers and enable streaming and data downloading to user desktops. At the heart of the GeoHub is a rules engine. To enable the development of a large community, the OGC standards have been deployed so that GIS users in the jurisdictions can utilise the hub. The functionality is being delivered using components in today's Web 2.0 architecture, but with a view to upgrading as IT architectures evolve.

GeoHub has addressed a number of SDI problems, not least the under-exploitation of the benefits of sharing spatial data due to poor access, a lack of understanding and awareness of the applications for spatial data, the perceived costs and difficulties of using geographic information technology and the typical silo approach to storing and managing spatial data. This affects a number of business process areas, such as policy development and service delivery, but also hampers communication, meaning that decision-making is impeded and service delivery efficiencies are not realised. It can also result in an ill-informed public and business community, investments that have already made in areas like data capture being squandered and wasted resources in maintaining (or duplicating) spatial data and systems in multiple locations.

Spatial Data Quality and Semantics

These second phase SDIs must solve the spatial data quality problem as they seek to integrate data and provide a single source of truth for decision-making. Initiatives from the OGC in particular have helped solved the problem of transporting spatial data from silos, but methods for assessing the quality and fitness for purpose of the datasets are still being developed to allow for effective comparison and integration.

Often overlooked, and a stumbling block, is the semantic quality of the data. Combining datasets from different sources and different databases often means integrating similar datasets but from differing schemas. This can be hugely problematic. The use of ontologies is beginning to help provide a framework for describing the meaning or purpose of spatial objects and datasets, which gives a basis from which to work. But even then, measuring quality and assessing fitness for purpose still remains an issue.

It is important that quality assessment goes beyond simple geometry checks looking for overlaps, for overshoots etc and makes use of this semantic information. Combining data for SDIs compels us to check that the descriptions of features match and make sense if we are to use these combined datasets for decision-making purposes. Is this type of valve permitted on this pipe? Does this planning application make sense for this building type? Both are valid questions that go beyond checking that the geometries of such objects coincide. Measurement and maintenance of such criteria is imperative to the usefulness of a SDI.

Recent advances in rules based spatial data quality measures are gaining acceptance and proving an effective solution. The OGC itself is now investigating this approach in an attempt to formulate standards in this area and there is now an international Data Quality Working Group[6]. It will attempt to define a framework and a grammar for the certification and communication of spatial data quality.

[6] http://www.opengeospatial.org/projects/groups/dqwg

The Future

A lot of work has been carried out in Great Britain on creating a Digital National Framework (DNF)[7]. This work set of principles, concepts and methods evolved through best practice with the intention of improving data integrity, promoting greater reuse and faster and easier sharing of application information. It is neither a standard nor a database. The aim of the DNF is to improve business value and returns on investment through adopting more reliable and consistent methods of data connectivity. It is an attempt to create a supply chain, but more importantly, it provides a schema for the ongoing update based on Master Data Management principles.[8] The DNF has no effective ownership in Great Britain, the opportunity is for the Scottish Government to take ownership of the DNF for Scotland. It can then as defined in the One Scotland - One Geography strategy[9], moving towards the time when information about the physical land surface such as local authority boundaries or electoral wards, is known to be accurate and be a leader as defined below:

> *"The fundamental requirement for the ongoing, interoperability needed to sustain high tech is accurate and honest exchange of information. Your partners need it, your distribution channel needs it and must support it and your customers demand it".*[10]

---

[7] http://www.dnf.org/Pages/home/default.asp
[8] McGuffog – http://ukpeb.org/diamond/diamond1.pdf
[9] http://www.scotland.gov.uk/Topics/Government/Open-scotland/OneScotland/Spaces
[10] Crossing The Chasm. G.A. Moore (1999). Harper Business, ISBN-10 0066620023