White Paper

Schema Translation for SDI (Spatial Data Infrastructure)

Drivers for Change

Providers of geographic information are facing an ever escalating range of demands from their existing customers, new customers and legislation.

Public sector bodies are increasingly faced with pressure to enhance the use of information they capture and maintain. This can come from the commercial pressure of cost recovery targets or from the need to show value for the public money invested in the data. To make the information into a truly commercial product and appealing to new markets, it has to be offered in different forms and through different channels.

A range of initiatives at regional, national and European levels seek to provide data to consistent technical and quality standards across borders. INSPIRE[1] related projects, for example, seek to harmonise data across Europe for a number of themes and so require agencies from member countries to provide information in a consistent way. Many countries have similar initiatives at the national level to harmonise access to information across federal or local government boundaries.

In addition to the changes in content of data, there are also demands for improvement in delivery. Data users need the information not only to be accurate but also up to date. This drives demand for access to data through internet and web services, where users can be sure of accessing the latest available data. These services also free the users from the burden of having to manage large holdings of data locally. The fact that these services can also be shared by many users makes them the key building blocks of a spatial data infrastructure.

Limits of the "One Size Fits All" Approach

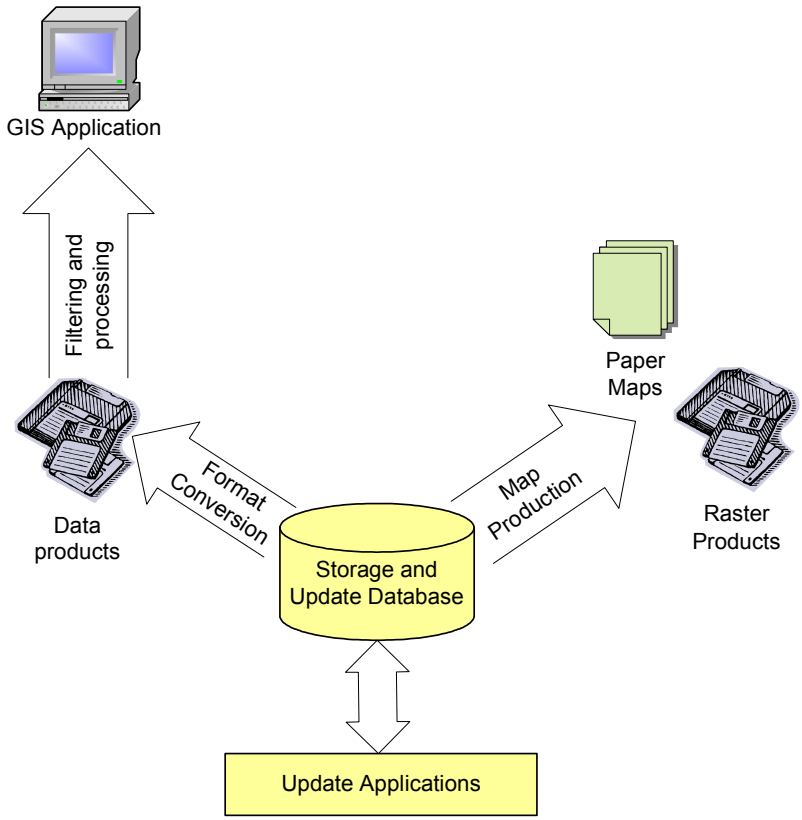
At present most data producers provide their output data in a single data model: the model which they store and maintain. Although the data may be supplied in a variety of formats (either standards based or proprietary GIS), the underlying data model is the same. The decisions about which objects constitute a geographical feature and what attributes those objects have are the same in all the supply formats.

To make use of the data in a particular application, the consumer must process the data to transform it into a form which suits their purposes. A rich data model containing many feature types and detailed attribution may be suitable for an expert user who wants to carry out complex analysis. That same information may well be of use to a more general user needing to provide context to other data or to illustrate a report, but the level of detail and technical jargon in the data model would inhibit them from using it. The more general user needs a simplified view of the information which suits their purposes.

[1] Infrastructure for Spatial Data in Europe

Taking road information as an example: routing applications require drive restriction information and road centre line geometry to carry out route planning. Highways asset management applications require detailed road geometries and area measures but have no need for turn restriction information. These different applications require quite different views of the road network.

The processing of data by the consumer, therefore, places overheads of time, specialist software and expertise on the consumer. Whilst the benefits of using the data outweigh the costs for many users, there are many more for whom the overhead is prohibitive. This 'one size fits all' approach to the data model effectively restricts the use of the data to 'expert' users and prevents 'casual' access to the data.



In addition to needing different content, different groups of users have varying requirements for supply. Paper plotting is popular because it provides a simple and highly accessible way of distributing information. The most common form of digital data supply is the transfer of files via media or download. While this is a robust and reliable form of supply, it does require a degree of data management effort by the consumer and may fall short of the user's requirement to see the most up to date information.

Figure 1 - At present, many data producers export a single data model for all customers

System Architecture for Electronic Delivery

The emerging commercial and legislative environment for data suppliers requires a new architecture for the systems used to supply data. This architecture needs to support a wide and changing set of channels for the distribution of information.

The storage and update database is the central repository for all the information which will ultimately be supplied to users. This database must therefore have a large data model capable of holding all the information which may be supplied. The model is designed and optimised to support an efficient maintenance process. Each piece of information is recorded only once even though it may be presented in different forms to different users. A number of applications supporting data capture, survey, quality control etc. are connected to this database to keep it up to date.

The storage and update database and applications may be existing systems which support the established data collection processes. This architecture therefore re-uses existing systems and allows current business processes for data maintenance to continue undisturbed.

Connecting to the database are the supply systems which contain two key types of functionality: a schema translation engine and a range of delivery mechanisms.

The schema translation engine carries out an 'on-the-fly' translation of data from the storage data model into the various models for supply to customers. This is the key part of the architecture since the storage model has been designed around the needs of the data maintenance processes: it is therefore not a good product model - since the needs of customers are secondary to its design. The use of a schema translation engine a) allows the supply of appropriate content to different channels and b) guarantees that all channels receive consistent information. All channels are getting the most up to date, definitive information direct from the database and in a form which suits them best.

The supply systems will contain a range of delivery mechanisms to support the needs of different users. For example:

- Web Feature Servers to provide internet access to the most up to date information with no data management overhead on the user.
- Web Map Servers to provide simple and rapid access.
- File supply on media supporting expert users with traditional desktop GIS software.
- Database synchronisation using Change Only Updates (COUs) to maintain databases for clients with specialist performance requirements on their database such as very high query rates or high resilience for business critical applications.
- Paper mapping, which remains a simple and portable means of distributing geographical information for the general user.

Benefits of this architecture include:

- Supports a wide range of users by providing the information in easy to use forms.
- Provides access through supply methods to suit different users.
- Enables all users access to the most up to date information regardless of supply method or data model.
- Ensures extensibility to meet the demands of new legislation.
- Can be augmented to new distribution channels for new markets with minimal investment.
- Protects and enhances existing investment in data maintenance systems.
- Prevents disruption of existing business processes

GO Publisher

GO Publisher is a suite of software products designed to support the architecture described in this paper. Common to all the products in the GO Publisher line is a schema translation engine which connects to any standard (JDBC compliant) database – Oracle, SQL Server, DB2 etc. It can therefore connect to existing databases which will take the role of the storage and update database.

Products in the GO Publisher suite offer a range of data supply methods including file transfer and internet/web services. The supply methods are built on international ISO and OGC standards such as XML, GML, and WFS. Once a translation is configured that configuration is portable across the product suite, so the same data model translation can be re-used in different supply methods.

The product suite contains robust and scalable server side components to handle high data volumes and query levels. It also contains graphical configuration tools for high productivity. Data model translations are configured using point and click interfaces which remove the need for complex scripting languages in the translation. New translations can therefore be added rapidly and reliably, since the addition of a new translation is considered a configuration task rather than a software development project.

With tools such as GO Publisher, data suppliers can easily build the system architectures they need to meet growing demands for access to information.

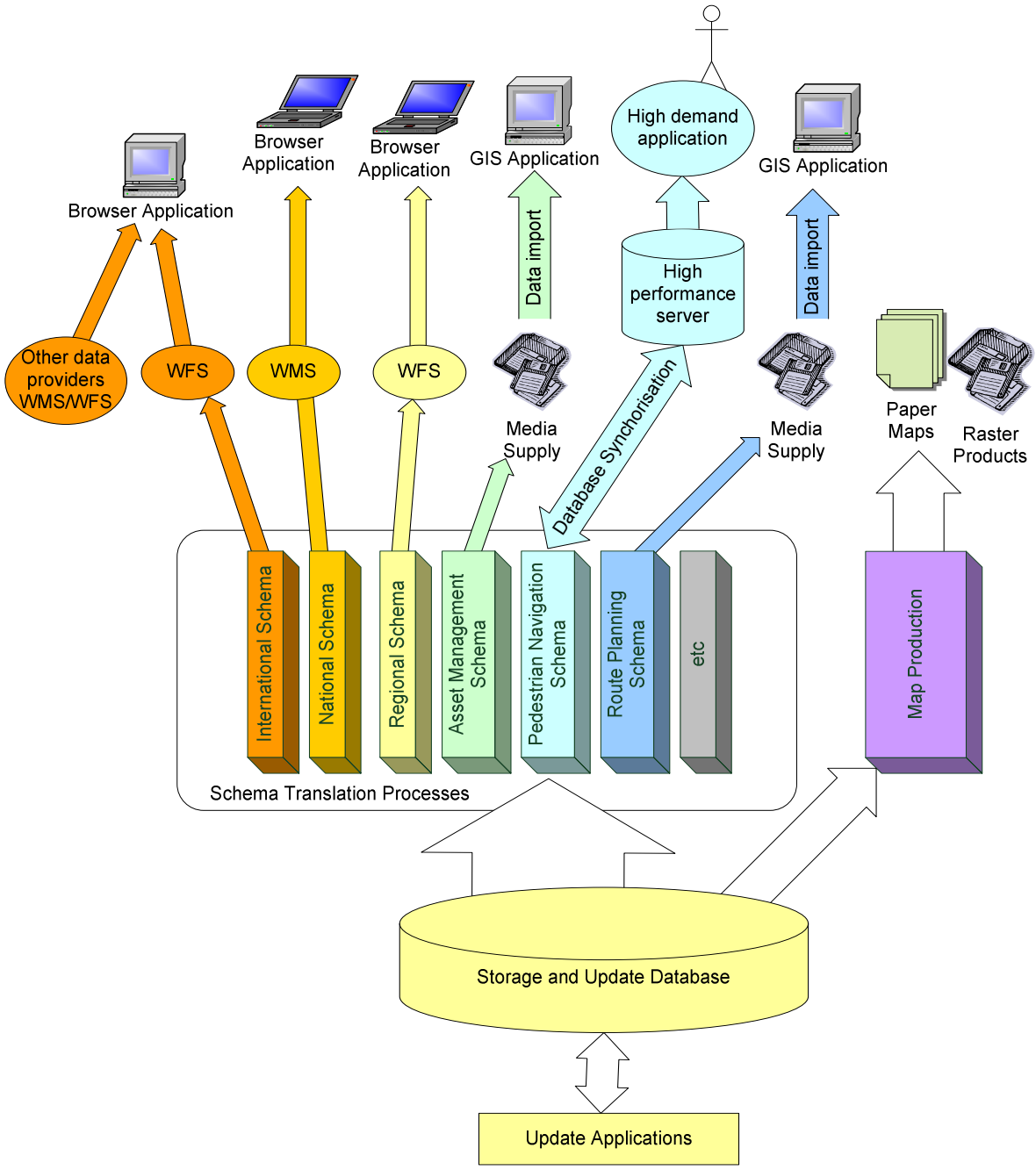


Figure 2 – The Emerging System Architecture for Electronic Delivery of Information

Contact Information
Snowflake Software
Eastgate House
Town Quay
Southampton
SO14 2NY
T: +44 (023) 8023 8232
F: +44 (023) 8023 8233
E: info@snowflakesoftware.com
www.snowflakesoftware.com