

LAND ON THE WEB

HOW E-COMMERCE IS RESHAPING LAND DATA MANAGEMENT AND DISTRIBUTION IN VICTORIA

Mark Zangari

SPATIALinfo
Level 14, 459 Little Collins Street
MELBOURNE Vic. 3000
Australia

INTRODUCTION

There is no question that the emergence of the internet has spawned a revolution in the IT industry, if not the economy as a whole. And, if anything, the pace of change is increasing with improving technology. While much of the public face of the internet is centred on consumer oriented sites on the World Wide Web, the foundations of the revolution, and probably its most sustainable outcomes, come from the internet's impact on the way companies are doing business. Even today, the as yet untapped potential of the internet to revolutionise business to business, or "B2B", information sharing is immense.

This is particularly true in the area of spatial data management, where progress towards "web-enabling" large GIS systems has in general not kept up with cutting edge e-commerce systems. At first glance, this seems somewhat at odds with the potential apparently offered by web technologies, given the inherently graphical nature of GIS. However, because of the richness and complexity of the data in typical high-end GIS systems, building e-Commerce models for GIS presents unique challenges not found in many other IT disciplines.

This paper addresses these challenges by considering various e-commerce models for GIS that complement the strategic direction that many utility businesses and land management authorities are taking, namely:

- Outsourcing of design, data management, and data distribution functions to other companies that specialise in these activities, while retaining ownership of the master databases.
- Automating data transfer between entities without sacrificing application functionality or data quality assurance.
- Adding value to spatial map datasets by integrating them with data from other sources.
- Providing new business opportunities for spatial data distributors.

The way internet technologies have been successfully leveraged to achieve these objectives is illustrated by various initiatives under way in Victoria.

SPATIAL DATA MANAGEMENT

Most utilities and land management authorities manage the entities they are responsible for via large and complex spatial databases. Nowadays, spatially enabled relational databases, where both spatial and aspatial data are stored in a relational database, are replacing the classic geo-relational model, where spatial data is stored in an application specific repository, and linked to relational database records via some proprietary mechanism. In the past, businesses which relied on these databases used hand-drafted paper maps as the archival—indeed, often the only—repository for the information. With the advent of CAD systems, the map sets were converted to electronic files, often with no change other than what once used to be stored on paper was now stored on disk. The next transformation in spatial data management has seen discrete drawing-based systems upgraded to seamless GIS systems—a transition that is still in under way in the industry.

The introduction of progressively more sophisticated IT systems into mapping and facilities management applications has reduced the labour required to operate and manage these systems. This trend is continuing, albeit the next apparent transition is not merely the result of a technological upgrade, although it is dependent on new technology and the internet in particular. Instead, it is primarily a consequence of the organisations which rely on GIS systems redefining themselves. Many large commercial utility businesses and government authorities alike, are restructuring themselves to focus on their core business activity, and are outsourcing support functions that used to be provided in-house. Network operators no longer wish to run engineering and construction divisions, land authorities no longer wish to be data managers, etc. Rather, these functions are increasingly being performed by outside contractors.

Challenge: Merging and Synchronising Outsourced Data

The high-performance inter-business network connectivity made available by the modern internet, along with the effectiveness of security protocols, has all but solved the physical challenges of transferring data between data owners and data managers. A major challenge less easily solved is the synchronisation of data maintained by an outside contractor with the master database of the data owner.

The factors to be considered in this regard are the following:

- Data owners wish to maintain sophisticated databases containing spatial and aspatial data.
- Data owners wish to be able to leverage their existing software and data, and not have to implement entire new systems or undertake large data migration projects to enable outsourcing.
- Data owners should be able to outsource different subsets of the database, separated either by spatial extent or content type, to different service providers.

- Data service providers wish to be able to use software systems which best fit their business requirements and available resources, rather than be forced to replicate the systems of each of their customers.
- Both data owners and data service providers need to provide quality control checks before importing data into their operational systems.

A simple file-transfer protocol would not be able to meet the above criteria. Let us now consider and evaluate some other possible strategies.

Synchronisation Strategies: Direct Connection vs. Replication vs. Incremental Update

The data synchronisation challenge can be addressed by solutions which can be categorised into three broad alternatives:

1. The data services provider remotely logs directly onto the same system as the data owner. In this model, there is no need for explicit synchronisation, as the data manager is updating the owner's data directly.
2. The data services provider sends updated data back to the owner (and vice-versa) by transferring complete updated datasets in the format of the target database. Synchronisation is achieved by loading the data into the target system and replacing the old dataset by the new one.
3. The data services provider sends back platform-neutral incremental update files which specify only the changes in the remote dataset that have occurred during a specified time interval. These are processed and used to make the changes required to synchronise the source and target databases.

Under the first alternative, synchronisation ceases to be an issue because the changes are made directly into the data owner's system. This model is analogous to the way most commercial web-sites operate. As we discuss below, remote, web-driven interfaces find a natural application in the outsourcing of functions which involve little complex spatial editing and must be utilised by personnel without specialised computer skills. For more specialised applications, it is possible to securely connect clients to servers via the internet using technologies such as Virtual Private Networking (VPN).

Direct-log in, despite its relative simplicity, imposes a system administration overhead on the organisation which operates the servers. In addition, quality control and security become issues when outside operators are granted direct access to the owner's data.

The second and third alternatives are similar in that the data owner usually plays no operational role in running any part of the service provider's system. Also, in both cases, synchronisation is deferred and "batched", and an intermediate data transfer format is used, rather than an application writing directly to the database. The differences between them are that, in the second option the complete target dataset is sent back, including data that has not been altered, and the transport format is normally closely tied to the target database, usually employing a standard toolset for

data transfer provided by the database system vendor. Exchanging binary database dump files are an example of this kind of strategy.

By contrast, the incremental update approach uses a platform-neutral file format and does not attempt to define the new dataset, but records only changes made to the data within a specified timeframe (normally the time since the last update was loaded).

This immediately has several advantages:

- Since only changes are referenced, the amount of data transferred is smaller than transferring a complete dataset, and the load cycle is therefore shorter.
- Since the incremental update specification is platform neutral, there is no requirement that the service provider and the data owner use the same system configuration, or even the same data schema. Similarly, both the owner and the service provider can update their software without forcing a change in the other party. For organisations sharing data with many subcontractors, this is an important benefit.
- The load process can apply quality control screening before making changes to the target database.

Vicmap Digital Property Maintenance and Distribution

The Victorian Cadastral database resides in a system called Vicmap Digital Property (VDP) and is the responsibility of Land Victoria, a division of the State Department of Natural Resources and the Environment. However, for a number of years, it has been successfully maintained by another specialised data management organisation, Dataflow, who manages the day-to-day maintenance and update of the system as land boundaries change, data quality is improved, etc.

Land Victoria stores VDP data in a GIS system that is optimised for the query and analysis functions undertaken by the department. Dataflow's requirements as the data manager, however, are focussed not on query and analysis, but on efficient and controllable data entry with a high level of auditability and quality assurance. Hence, a different system architecture is used, using SDM/Oracle as the spatial repository with a CAD-based system as the front end edit application.

Data synchronisation between Dataflow and Land Victoria is achieved by incremental update output produced by Dataflow at regular intervals. This output format is well suited to electronic data interchange between the two organisations, consisting of human readable ASCII statements identifying which entities are to be replaced, the updated data values, and which versions of previous entities are being superseded.

SPATIAL DATA DISTRIBUTION

The world wide web (WWW) is an ideal technology for data distribution of almost any kind for the following reasons:

- It is widespread both in the consumer and business sectors, readily available, and cheap.
- It defines a set of standard protocols for data transfer.
- It is possible to send data with a high level of security.
- Software applications which provide access to the web (i.e. browsers, etc.) are very cheap and do not require specialised computer skills to use.
- It is extensible, as many development tools exist for building specialised applications to run within web browsers.

As a case in point, the Victorian State government has recently decided to allow data from various databases managed by the State Department of Natural Resources and the Environment to be distributed via the web. However, rather than developing the capability to do this in-house, the department has instead made available licences allowing third party commercial businesses to distribute data on the government's behalf. As well as competing for the supply of state-owned cadastral and land title data to established customers, some of the licence holders are attempting to exploit the wide reach and ease of use of the web to generate new business.

Four cases where the web provides a new and unique opportunity for GIS data distributors are:

1. Selling (or licensing) GIS data to both businesses and the public using a WWW-based "e-store" concept.
2. Using the WWW to provide direct access to GIS data for customers (or business partner).
3. Using a single web interface to provide seamless access to a number of disparate datasets, each residing on different systems.
4. Web-enabling GIS systems and using these as spatially navigable front ends to query data from other, aspatial systems.

Selling Spatial Data on the Web

One of the highest profile classes of web site is the "e-store" where consumers can purchase items by searching for and choosing the products they desire from the vendor's inventory, and then supplying their delivery and payment details. Typically, the ordering process is completely automated, handled exclusively by the vendor's software. Order fulfilment, on the other hand, is typically manual, from the packing of the goods to their delivery to the final customer.

A similar model can be applied to the distribution of GIS data, except that since the "goods" being delivered are data, the entire process, including delivery, can be automated. This process has the potential to greatly democratise access to complex spatial databases, as a well constructed web interface should allow almost any computer user to select and purchase the data they require, and generic delivery formats will make this data readily accessible to users who do not have high-end CAD or GIS systems.

From the vendor's perspective, however, the process raises software issues that are not present in the sale of discrete items from an inventory, as GIS "products" are generally qualified by spatial areas and content types. In addition, most government

authorities and utilities require a higher level of auditability and traceability of the data they distribute than typical web stores. Consequently, most commercially available e-commerce development kits are unable to deal with the specific requirements of setting up a map data distribution site and specific expertise in both spatial data and e-commerce is required.

The area where the web-based sale of GIS products differs most markedly from more traditional e-commerce is in the way the customer selects what they wish to purchase. Rather than the web interface providing a customer with the ability to search the supplier's inventory and select the items required for purchase, a GIS e-commerce web site must allow the customer to spatially navigate over a map of the area from which data can be purchased. A particular challenge in this regard is ensuring that, in providing the navigation map, the vendor does not inadvertently give away the ability to view the required data for free. Consequently, great care must be taken in specifying the content displayed in the navigation map to ensure that sufficient data is shown to allow efficient navigation, but not so much that potential customers will not have to make a purchase to receive the data they require.

Another challenge unique to selling spatial data via the web is the calculation of pricing. Again, since the data on offer is generally a continuum rather than a set of discrete items, simple cost-per-unit pricing is not sufficient. This is especially true given that cost calculations may be based on simple areas, entity counts per area, types of content provided, type of customer, or a combination of all of these. In addition, the distributor will typically have to pay a royalty to each of the owners providing the data in each sale, of which there may be several if data from several sources is being integrated (see below). Each data owner will typically have their own royalty pricing structure.

Streamlining and Outsourcing Customer Interfaces

Another major opportunity offered by the web is the ability for organisations to broaden access to the data maintained in GIS systems first throughout the enterprise, second to their business partners, and third to their customers.

Since many utilities and government authorities have outsourced much of their interaction with consumers via points of presence at third-party retail businesses, a need exists for these businesses to be able to access information from the utility's databases. The web provides exactly the infrastructure needed to enable such access, with dial-up-on-demand meeting the needs of smaller retail outlets. By adding a web interface with a sophisticated map viewer application to the utility's GIS, third-party retailers can easily view, navigate and run queries on graphical maps showing such features as property boundaries overlaid with existing and planned network. Similarly, a data entry page allows retail sales staff to order new services for the utility's customers and, in many cases, provide immediate quotes. This kind of capability will generally allow near real-time turnaround for many functions which would take days without direct on-line access, and dramatically reduces the utility's point-of-sale support costs.

Integrated Access to Disparate Datasets

The ease with which web servers can now be integrated with a variety of data sources makes the WWW an ideal technology for providing integrated and seamless access to several systems via a single interface. Since no single authority generally has responsibility for all of the data which may be required in a given application, integrating data access and presentation via a common interface is a major benefit offered by the web. In most cases, the user need not be aware that they are in fact interacting with several discrete and probably completely separate systems.

As a hypothetical example, consider a case where a query on a land titles database is to be run and must produce a table of values which refer to property boundaries on a map as output. This would usually require first a query on the titles database to retrieve the title data, and then a query on a cadastral database to produce the map. A single web interface to both systems, such as is under construction in Victoria at present, will allow a user to define both the textual and the map the information that they wish to see as the parameters to a single logical query. It is then left up to the system to determine which physical databases must be queried in order to fetch the data, and to determine how the data is to be presented. Both of these operations are transparent to the user, as are pricing and royalty calculations behind the various queries.

Spatially Enabled Presentation of Aspatial Data

Data associated with geographically located entities is most naturally represented in a graphical map display. Maps with appropriate symbology allow information-rich views, and especially spatial characteristics, to be interpreted much more efficiently than almost any other method of data display. However, much data about geographical entities resides in aspatial databases, and spatially registering this data is usually complicated and costly.

A more efficient approach is to use existing keys in both the spatial and aspatial datasets to link spatial and aspatial data in different databases. (This technique is effectively a meta-application of the geo-relational paradigm, so that instead of using link keys to relate graphical entities to their attribute data within a GIS system, the links are used to relate the GIS entities to other aspatial databases.) Such link keys, when coupled with the inherent ability of web servers to work with multiple data sources, provide a powerful mechanism integrating the display of data from aspatial databases with a base map.

This approach is being used in Victoria to link aspatial land titles records to property polygons in VDP via a common identifier. The existence of such a cross-database key enables web-based distributors with access to both the cadastral and Land Titles Office (LTO) datasets to use the cadastre as a “logical” spatial front-end to the LTO database, even the LTO data is purely aspatial. In such an approach, users can use the cadastral map to navigate to the required title records, and data sourced from the LTO database can be displayed on the map, as if it had been spatially registered.