International Symposium on Spatial Data Infrastructure (SDI)
University of Melbourne, Australia

Building Spatial Data Infrastructures: Case Studies and issues

Rob Atkinson and Maurits van der Vlugt

Abstract

This paper examines the role of Spatial Data Infrastructures in enabling the provision of spatially relevant information into the broader community. This issue goes far beyond simply "Online GIS" into examining how spatially explicit metadata is required for online information in documents, directories, transactions and notification channels.

The paper examines the types of services that an SDI must provide to allow for geographic organisation of the Web in general, and identifies the mechanisms available to do so. Some case studies to illustrate implementation of these principles in New South Wales are presented.

Introduction

It has always been hard to build effective geospatial applications. Technically, it demands a wide range of skills, software that's expensive to buy, expensive to learn, or both, and some decent equipment. But that's the easy part. The accepted benchmark is that most projects spend 85% of time and money budgets acquiring the data required. Frustratingly, it's usual to negotiate a copy of contextual data, without any update facility. Consequently, there are very few operational systems, and they are expensive to build and maintain.

The role of a Spatial Data Infrastructure is to increase the availability and effectiveness of applications. A mechanism is to increase accessibility of data – buts that's not the end game!

This paper looks at the role of Web based services in effecting an SDI, drawing upon several year's experience with implementing and researching SDI concepts and Web based technologies.

Infrastructure

If we accept the role of SDI as facilitating more effective uptake of spatial data within operational applications, then we can immediately start to identify the imperatives. In the same way that the Internet provides a communications infrastructure, SDI must provide a spatial data **access** infrastructure. This means that it must provide:

- Data that would have a large cost in terms of time, money, usefulness or risk to duplicate individually
- Documentation about that data so that others can judge its fitness for purpose
- Clear and simple rights of access, consistent with underlying social principles (Government has accountancy issues, business needs secure cost basis and freedom from subsidised commercial competition, community has moral right to free access)
- A means to build applications that can exploit the data
- A means to disseminate data to those applications
- The possibility to locate data and discover the access parameters

There is obvious progress towards all of these goals, however the last three remain adrift because they are where the "rubber hits the road" – it's not that they are technically such a problem so much as they really become a test of the maturity of other parameters.

It's tempting to therefore continue to say "we must get the framework data right before we consider using it" but this is naïve to the extreme. It is analogous to the architects of the Web saying: "we must have all human knowledge encoded as HTML pages before we release the interface protocols". In reality, only by exercising the entire system will it have an opportunity to gain momentum to generate the business requirement to continually improve the usefulness of the data.

Usefulness of data

The usefulness of data is a function of many factors. Accuracy and completeness are often cited as the most important factors to get right. However the practical impact of this is often poorly understood. In most cases it's a function of uncertainty regarding liability – its better to give people no data than risk bad data. A role of SDI is to get over this with a consistent policy framework that recognises that no data is ever perfect and what is required is a means to disseminate it in a potentially useful fashion without false expectations. Here is some news: the public is already used to dealing with accuracy issues. Road maps go out of date. Most of the time you can rely on them, when they are wrong or out of date it's not a big deal. They are judged to be **useful.** They are used.

The most important factor is accessibility – without it the data cannot be used, and its potential usefulness may not even be evaluated. This is the good news about current SDI initiatives: the focus on identifying the data custodian has been a vital step. Empowering them with policy-, legal- and technical frameworks remains to be effectively implemented, and the lack of understanding about how that data might be used is a key limitation.

Why Web based services?

The choice of a technical platform for online access has been effected. The use of the Web (i.e. HTTP) as an access protocol (over TCP/IP and the Internet) has been universally adopted as the candidate for implementing large-scale heterogeneous infrastructures. Earlier protocols (e.g. gopher, Z39.50, ftp, CORBA) have tended be good at supporting specific implementations, but have a relatively high cost of compliance and low levels of flexibility. In fact, they are not irrelevant, but gateways can be readily built between HTTP and other systems.

The OpenGIS Consortium Web Services Testbeds represent the coming together of both SDI "sponsors" and vendors. Thus the requirements and the ability and willingness to meet these converge and the focus on Web based services is confirmed.

The use of Web based services has a number of interesting implications:

- Most of the technical infrastructure can be borrowed, including security, e-commerce etc
- Moving small packets of data is the norm
- Advertising service access points (amongst billions of others) is an issue
- There are opportunities and expectations to interact with non-spatial systems abound

Case Study: NSW Natural Resources

The 12 agencies within the NSW Government with natural resource information management within their portfolios jointly commissioned a system that, in today's world, would be seen as the technical layer of an SDI. A project that exploits this infrastructure is called "Community Access to Natural Resource Information (CANRI)" and consists of a suite of applications, data servers and a catalog component.
This system was deployed as a network of data servers using a common open (but pre-standard) protocol. The system design assumed the emergence at some stage of a standard protocol and implemented a data model that assumed "binding" of services to protocols. Currently the system includes both OGC compliant Web Map Servers and prototype Web Feature Servers.
CANRI has a range of Web applications configured to use different sets of underlying data services. New applications can be created by choosing data from the CANRI catalog of data sources, or by configuration against non-advertised services. The configurations can then be exported (as XML) and reloaded as required.

A wide range of application paradigms have been explored, and tools evolved to enable configuration of these basic patterns. Some examples include:

1. State of the Environment reporting: integration of interactive maps into Web documents and data dissemination. A single application with multiple configurations (1 per environmental indicator) and shortcuts to local government views.
2. Wildlife Atlas: an application for navigating through a huge collection of records, allowing aggregations to be reported (e.g. counts of sightings in a particular area)

as well as mapping of the actual locations. This include a taxonomy service for species names.

3. FaunaNet: a similar solution that accesses a network of museum collections and displays simple spatial aggregation models.
4. River Internet Information System – real-time access to river flow data
5. SPADE – Access to soil profile archives

Case Study: Regional Services

Outside of the CANRI program, many of the same data sources are re-used in a program to develop regionally focussed delivery of government services. The ability to deliver maps in the same way that news items can be published is part of a multi-media portal based on content syndication from government (and eventually community and business) services. This exploits the same SDI as CANRI, and adds value-added capabilities of integration with X.500 directories, document searches and the emerging business directory technologies (UDDI).

Case Study: NSW Government Planning Portal

This project is another case of the ability to exploit and extend SDI once it is deployed. In this case the set of agencies with Planning responsibilities (which overlaps the Natural Resources Sector) is able to create a portal that incorporates Local Government as well as State agency data. The use of common access standards means however that any map view can be easily augmented by data served via the SDI infrastructure, and that the new planning information published can likewise become available within the original Natural Resources Sector.

This project also significantly extends the CANRI deployed technology platform by:

- introducing Gazetteer service functionality
- introducing a point-identify-navigate paradigm where a range of spatial objects can be identified (e.g. land parcel, catchment, local council) and used to navigate the map (i.e. show me Grose River catchment)
- linking between maps and documents – so that documents can be used to invoke maps at any time (“show me the planning zone this document refers to”) and also relevant documents located by querying the maps. This is implemented in a fashion that allows the maps, documents and applications all to be distributed across different agencies and technologies.

Lessons learned

The case studies here represent the way in which an SDI can be built and evolve in breadth of data, functionality and institutional buy-in, without undertaking any massively costly, high-risk projects. Our experience also highlights a number of issues and hurdles, some of which are yet to be overcome.

Wins

- The distributed architecture and HTTP/XML technology base has become widely accepted as the basis for online service delivery within SDI contexts
- As expected a standard protocol set has emerged and is continuing to extend its capability (OpenGIS Web Services)
- It has proven possible to incrementally migrate services through good SDI component design and implementation
- It has proven easy to create new applications once data has actually been served
- Flexibility and configurability of applications has been required and achieved
- New sectors have drawn from and added to the SDI as per theory
- Major vendors are providing support for the OpenGIS interface standards

Hurdles

- Interface design for usability is difficult and expensive – and necessary. Clunky GIS interfaces don't solve real world problems.
- New and appropriate technologies take a long time to be tested – nobody wants to pay for extending core functionality.
- Nobody wants to pay for shared services – e.g. Gazetteers
- Everybody needs shared services to implement effective solutions – e.g. Gazetteers
- It takes time to introduce new technologies due to yearly budget cycles. Hence it can take two years to implement a new or improved capability, and in the meantime a new option may emerge, which merely introduces new delays.
- There needs to be a significant ongoing budget to handle the process of evaluating new options and creating deployment strategies.
- New projects are not sufficient or appropriate sources of funds to cope with the burden of testing new paradigms. Projects want guarantees of success, which are incompatible with evaluating new technologies.

Conclusions

Many of the goals of Spatial Data Infrastructures can be only implemented effectively by online service delivery. The goal of sustainability of SDI efforts is probably tied up with the usefulness of online data provision services. Practical experience shows that, with a little skill and perseverance, SDI's can be grown incrementally through application of best practice. International standards efforts are, in practice, suitable and accessible to meet the fundamental requirements of SDI implementation.

Effectiveness and cost-efficiency would be, in our opinion, dramatically improved by application of robust funding arrangements for shared infrastructure. The funding requirements are however quite modest compared to the cost of developing new technology or data sets. Lack of shared facilities condemns a jurisdiction to costly replication and repeated mistakes.