

SPATIAL PORTALS: ADDING VALUE TO SPATIAL DATA INFRASTRUCTURES

W. Tang, J. Selwood

ESRI China (Hong Kong), 10F Cyberport 2, 100 Cyberport Road, Hong Kong-
(wtang, jselwood)@esrichina-hk.com

KEY WORDS: Spatial Infrastructures, Web based, Internet, Metadata, Value-added, Interoperability

ABSTRACT:

This paper examines how well designed spatial portals bring value and return on investment to Spatial Data Infrastructure (SDI). Spatial portals are generally the visible front-end to SDI, they are the gateways (or brokers) through which users access the geographic services made available through SDI. Until recently the geographic information (GI) community focused effort on the practicalities of building infrastructure. These include: the development and implementation of the data, metadata, policies, standards, interpretations, networks, skills, database and application resources. SDI would not exist without this essential work. However, such emphasis has perhaps distracted attention from the spatial portal – the final, critical bridge between the user and the spatial resources offered through a data infrastructure. Over the last three or four years this beginning to change with greater attention being given to the design and functionality of spatial portals. This paper highlights some of these developments and argues that spatial portals play an important role in the success and continued viability of SDI.

1. INTRODUCTION

1.1 Spatial Portals – Critical Gateways to SDI

Access and distribution of geographic information is at the heart of every single SDI project. The United States National Research Council's Mapping Sciences Committee that first coined the term "spatial data infrastructure" in 1993, defined SDI as "the means to assemble geographic information that describes the arrangement and attributes of features and phenomena on the earth. The infrastructure includes the materials, technology, and people necessary to acquire, process, and *distribute* such information to meet a wide variety of needs" (Masser, 2005:7) (authors' italics). The first sentence of the GSDI's more recent and considerably longer definition (GSDI, 1999) is even more explicit, stating the aim of SDIs is to "support *ready global access* to geographic information."

Ultimately, users evaluate the success or otherwise of an SDI by the ease with which it allows them to locate and access spatial resources. As such, spatial portals, the gateways through which users interact with SDIs, play a vital role. Well-designed spatial portals are critical to securing tangible benefits from the considerable effort and investment involved in building SDI. Equally, poorly functioning or unstable portals will not be accepted by the user community, disengage service providers, and risk undermining the underlying SDI regardless of how good or comprehensive it is. Spatial portals are critical to the achieving return-on-investment and sustainability of infrastructure initiatives.

1.2 Aims and Structure

This paper aims to draw attention to the importance of spatial portals in the success of SDI, and highlight a number of recent developments that improve the speed, flexibility and functionality that portals offer SDI user communities. The paper is divided into four sections. The first discusses the evolution of portal technology from the late 1990s and identifies distinct types of portal. The second looks at a number

of design considerations relevant to contemporary spatial portals. In the third section we illustrate these with reference to portal projects drawn from around the world. These serve SDIs of varying size and nature, have wide relevance, and highlight elements of good portal design. In the final section we draw conclusions and outline a number of future developments and challenges that still remain.

2. PORTAL ORIGINS AND TYPES

2.1 Portals within the IT Industry

Since the mid 1990s, the computer industry has used the word 'portal' or 'web portal' to denote Web sites that either assemble numerous online resources and links into a single site (such as AmericaOnline or Compuserve), or provide search tools that help users locate information on the Web (for example, Yahoo! or Google). Derived from the medieval English word, portle, meaning city gate and originally from the Latin porta meaning simply gate, 'portal' sites seek to be their users' primary "point of entry" to the Web – their "gateway" or "portal". As the number of users and content on the Web grew exponentially throughout the 1990s, portals allowed the casual user to navigate an otherwise impenetrable mass of information. They played a significant role in the evolution and popularity of the Web by linking Web users with Web content providers.

2.2 Origins of Spatial Portals

By the late 1980s and early 1990s many government and commercial organisations had created sizable digital spatial data holdings and resources. At the same time the geospatial community recognised that difficulties in digital data access and dissemination presented major barriers to realizing the potential of these GI investments. Few standards governed the development of datasets. Differences in data models, nomenclature, and technology frustrated efforts to share and integrate data, models, and processes. The geospatial community had no consistent way of describing the source,

quality, scale, intended use, and other basic details about information services. It was difficult to compare different datasets of the same feature or region, or appreciate their intended use and limitations. As GIS usage expanded, large data providers such as national mapping agencies spent more and more time explaining datasets, responding to requests for data, and circulating updates. Keeping data up-to-date and consistent was an increasingly time-consuming task for data provider and user alike.

Studies by governments and organizations illustrated the extent of the problem worldwide. Though conducted independently, a number of studies undertaken by governments and organizations around the world (for example: those conducted by the Canadian Government, 1986; British Government, 1987; National Research Council in the United States, 1990; the Dutch government, 1995; and Australia and New Zealand, 1996) identified the need for better coordination and sharing of spatial resources and improved access. This is well documented in recent literature (Masser, 2005; Maguire and Longley, 2005).

Attention of the GI industry started to shift from capturing and creating data to finding more efficient ways to share, distribute and use it. SDIs were widely appreciated as a key strategy for achieving this and have been established by many different organizations at local, regional, national and global scales. Spatial portals provided a single location at which users could explore the resources available through the SDI.

2.3 Types of Spatial Portals

There are three basic types of spatial portal.

Catalog portals: Spatial portals associated with SDI are often referred to as 'catalog' portals. They create and maintain indexes or 'catalogs' of metadata that describe the nature and location of resources in an SDI. Resource owners (or 'service providers') register their services at the portal and supply metadata descriptions. The portal arranges metadata records from service providers into a consistent, searchable catalog and makes this available to users. Through the catalog users can search for services coming from any of the registered service providers. In most cases providers continue to host their own service and the portal simply connects users to the service(s) in which they are interested. The portal is a broker between users and service providers. Users have a single location (portal) through which they access up-to-date, authorized spatial information, and the service providers have a single location (portal) through which they can reach large numbers of users.

Application portals: In the late 1990s the rise of Web service technology as a robust vehicle for combining and serving complex geospatial data and functionality across the Web, allowed organizations to establish increasingly sophisticated Web-based mapping resources that permit users not only to search, but to view, combine, query and operate on spatial data discovered on the Web. In addition to the catalog portal's generic search tools, application portals provide more structured interfaces that include specific tools and applications relevant to user's domain interests. As the users and user requirements are often well understood, application portals can be tailored to meet specific needs, and the interface designed to provide efficient access to those data and functional services needed. Often application portals store some, if not all, of the data and functional services at the portal site. Application portals provide Web mapping tools to allow users to view and work with the

data they find (for example, geo-processing tools such as: route finding, geo-coding, printing, complex query and perhaps even redlining and edit/update functions). The distinction between application and catalog portals is blurring as this kind of functionality is increasingly being added to catalog portals.

Enterprise portals: A third type of spatial portal is emerging as the 'enterprise' spatial portal that integrates spatial data and functionality with business enterprise solutions. Enterprise solutions software appeared in the late 1990s to help large organizations manage distributed information resources. Such systems have tended to concentrate on office automation, enterprise-wide resource planning, and document handling and, until recently ignored spatial information. This is changing as many are now integrating GIS functionality and data into the portal environment. Users can switch between viewing data in document or spreadsheet form to analyzing the same data within a mapping environment without leaving the corporate portal. There is now a trend not only to bring spatial resources into the portal, but to 'spatialize' the entire portal. Enterprises are beginning to recognise location the primary common field around which information is ordered and searched. Valuable spatial data is often stored in free text format - contract documents, letters, reports, emails, spreadsheets, images, Web pages and so on. Tools to identify and accurately index spatial information held within unstructured documents are now being developed. Spatial search routines work with electronic document management systems, to index references to locations found within them. Building such indexes allows users to search for all documents that mention a particular location, regardless of type or format and are becoming increasingly important within enterprise portals.

3. DESIGN CONSIDERATIONS

Key design requirements of successful spatial portals are:

- Search accuracy;
- Speed;
- Simplicity/ease of use; and
- Interoperability and integration.

As experience in the development and operations of portals has grown these objectives have greatly influenced the evolution of portal design.

Search accuracy: The ability of the portal's search functions to return results that meet users' search criteria. This is a function of a number of elements including: the completeness and consistency of the catalog index, the robustness of the search process, and the design of query builder. Perhaps the most significant is the construction and maintenance of the catalog index. Early portals tended to search complete metadata catalogs. Users could search on any field within the metadata schema. However, it became apparent that, while this approach offered maximum flexibility to search on whatever field they wished, it produced unsatisfactory results. This is because organisations interpret metadata standards in different ways, and produce metadata with different levels of completeness - some complete all fields, others only compulsory ones. Catalog indexes were also often populated in a federated manner with little centralised control on quality and consistency. As a result, inconsistency crept into metadata catalogs which meant that searches processed on them produced highly unreliable results. Attempts to impose strict rules on the interpretation and

completeness of metadata records were perhaps feasible in strongly localised portals (say within a single organisation or department), but were unmanageable in larger regional, national or international infrastructure initiatives.

In response, portal designers managers now often establish dedicated search indexes in addition to metadata catalogs. Tracking user behaviour on some of the larger portals shows that the majority of users search on relatively few metadata fields (ESRI, 2004). Portal managers can build consistent indexes that cater for the majority of users' searches by strictly controlling the interpretation and quality of these relatively few, fields. Service providers may still submit full metadata records that can be used for reference. This approach makes it far easier for portal managers and service providers alike to maintain consistent catalog indexes. There is also increasing interest in adopting slim metadata standards such as Dublin Core (which has only 15 core elements in comparison to Federal Geographic Data Committee's (FGDC's) 334) for search purposes (e.g. CEN, 2003). In addition, update tools and procedures (automatic harvesting routines such as those based on the Z39.50 and Open Archive Initiative (OAI) protocols) help service providers ensure that records are up-to-date and consistent.

Another key consideration in the reliability (and speed) of search results is the flow of the query process. Another result of the adoption of distributed or federated catalog design used in early portals was that search requests had to be passed to multiple remote metadata index sites at the time of query. Though again elegant in theory, this proved difficult to implement in practice, as search speed and consistency were entirely dependent on the reliability of the networks connecting the metadata indexes. If a network or remote server had technical problems, search results would be incomplete. This is now changing, and most portals now store and process search request against centralised catalogs. This means that users' searches can be processed within the portal architecture, and only when a user needs to access more detailed information about a search hit will an external request be sent to the service provider. This greatly reduced network traffic, and increases scalability, speed and reliability of portal searches.

Finally, increasing attention is being paid to the search interface. Initial systems were strongly orientated towards the design and nomenclature of metadata schema. These often had little or no meaning for casual users. Defining search criteria could therefore be a confusing and frustrating process. Interface design techniques can help this. Portals now frequently include natural language search capability that hides technical jargon from the user, search parameters value domains, and combined spatial and textual search tools. In addition, channels help to identify and provide quick access to information on specific themes or topics.

Speed: The ability of the portal to quickly retrieve and display search results and related information. For portals to be successful they typically must return search results within 2 to 5 seconds. Achieving this relies on the shift to storing catalog indexes centrally within the portal infrastructure rather than in a distributed, federated model. A centralised model makes it far easier for portal managers to scale portals to accommodate growth in search requests and/or index size. Speed of display and operation may be enhanced by central storage of commonly referenced geospatial layers (such as those layers used in orientating users in the graphical search interface).

Simplicity/Ease of use: The ability of the portal to support a wide range of users who are likely to have little or no specific training in GI or portal operations. Users must be encouraged to explore the portal through simple, powerful interface tools that guide users to the information they seek. Examples include: creating clear search flows as in Geospatial One-Stop's "When, What, When" approach, and providing context sensitive dropdown boxes that are continually refined as the search progresses. Segregation of content into channels based on industry or theme also helps the user to rapidly navigate to data and information of relevance to them. Distributed management of portal information so that, for example focus groups manage channel design and content, help to ensure that channels are focused on requirements of the users they serve (note the management of information is distributed not the information itself). Intensive testing of portal design and continued monitoring of user navigation through the portal pages also help refine design.

Application portals are becoming more and more powerful (and common), and help users to easily move from finding information to using it either within the portal itself or in the user's own mapping clients. Application portals now now provide powerful mapping functions that range from relatively simple pan, zoom and identify functionality, to highly customized mapping interfaces that permit focused query and analysis. If service providers charge for the information services they offer, portals may also provide e-commerce and accounting functions that allow users to pay for services.

Interoperability: The ability of the portal to retrieve and present services in different formats and integrate them directly with 3rd party solutions. This includes being able to bring together data and functionality that may be hosted on completely different systems into a single application or map window. It also involves the ability to embed portal elements within 3rd party software solutions, a development that has the potential to bring the power of spatial searches to a much wider audience. Many portals now provide map visualization tools that permit integration of live map services from multiple remote service providers. User can access services discovered through the portal either through light Web-based mapping clients provided by the portal, or (increasingly) directly from their own desktop applications. Such tools allow users to combine, view and work with multiple remote services – pulling a GeoMedia Web service from one site, base mapping stored in ESRI, ORACLE and MapInfo formats from others. Different services can either be developed to conform with a single standard format (such as the Open Geospatial Consortium's WMS or WFS formats), or increasingly use automatic server-side translation applications (for example those from SAFE software) that allow on-demand translation between numerous different formats.

Industry standards such as eXtensible Markup Language (XML), Simple Object Access Protocol (SOAP) and Web Services Description Language (WSDL) underpin these developments and portals' ability to integrate not only different geospatial technologies, but between spatial technology and wider IT systems.

4. PUT INTO PRACTICE

4.1 Geospatial One-Stop

Geospatial One-Stop (GOS), an interdepartmental initiative managed by the United States Department of Interior, aims to increase access to spatial information held in government departments. Originating from early work on spatial data warehouses and metadata standardisation by the Federal Geographic Data Committee (FGDC), the GOS concept has evolved from distributed metadata search functionality, into one of the leading examples of catalog portals. A initial version of the portal launched in 2003 provided solid search functionality as well as, certain tools for content management, channel creation and content viewing. GOS2 (in development at the time of writing), focused on addressing a number of the issues raised in the preceding section.

In particular, attention focused on improved usability, interoperability and speed. With regards to usability it was recognised that even the three step What, Where, When search flow could be confusing or off-putting to many users. As a result GOS2 will follow the route taken by text-based search portals of fronting a sophisticated search with a simple, minimalist interface. A ‘Google-style’ natural English search bar will allow users simply to type in search strings that will be parsed against metadata records and combined with spatial searches. Thus for example “Alaska bridge” may be used to retrieve services related to bridges within or including Alaska. In addition, GOS has been ported to technology that will allow its functionality to be directly integrated into 3rd party portals or tools so that, for example organisations can imbed a GOS search bar or news and information feeds into their own Web pages. Users can therefore search for spatial information using GOS without leaving their own familiar environment. In addition, users can customise their own GOS interface that can be invoked every time they return to the portal and store predefined screen layout, search criteria, and relevant news and channel content. GOS may be found at <http://www.geodata.gov/gos>.

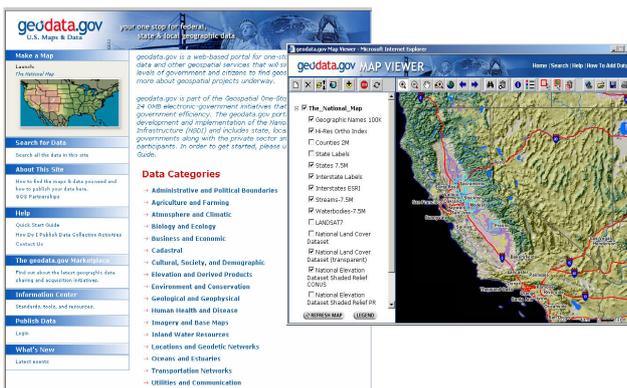


Figure 1. GeoSpatial One-Stop

4.2 MAPSTER

The MAPSTER portal established by the Department of Fisheries in Canada allows staff throughout the organisation access to distributed fisheries, marine and fluvial habitat data. This provides a good example of an application portal. Much of the data is served from DoF’s own central data warehouse, though the portal pulls data from selected data resources hosted

by a number of state and federal organisations. The data served, and user requirements are fairly clearly identified allowing the interface to be very heavily customised and even assumes users have some knowledge of the areas and datasets. There is, for example, no generic metadata search functionality. Data is presented in a hierarchy of folders based the project in which it was collected or the area it covers. For those familiar with the projects or regions covered, this is often more efficient than a generic search. Similarly the mapping interface provides very focused functionality targeted at users within the organisation. A modified version is available to the public at http://www-heb.pac.dfo-mpo.gc.ca/maps/maps-data_e.htm.

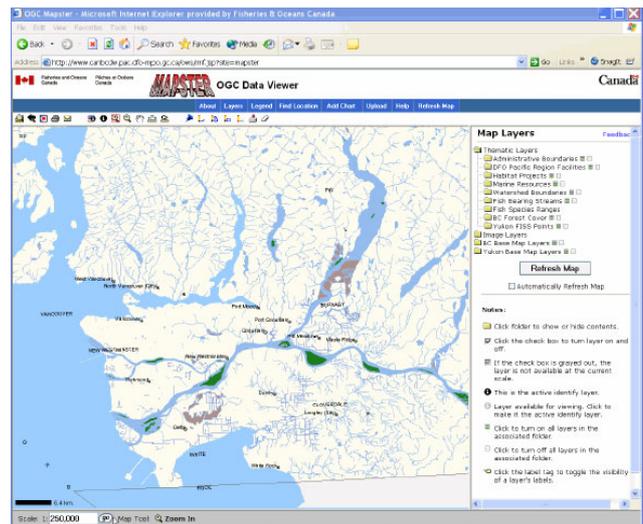


Figure 2. Department of Fisheries, Canada’s MAPSTER portal

4.3 Transport Direct

Transport Direct, a national traveller information system for the UK, illustrates how an application portal can mask extremely complex search flows with a simple, intuitive interface. The portal provides a highly customised environment to allow multi-modal travel information, route planning, ticket purchasing across the UK’s public transport infrastructure, including information on local and national buses, trams, trains, planes, ferries, taxi stands.

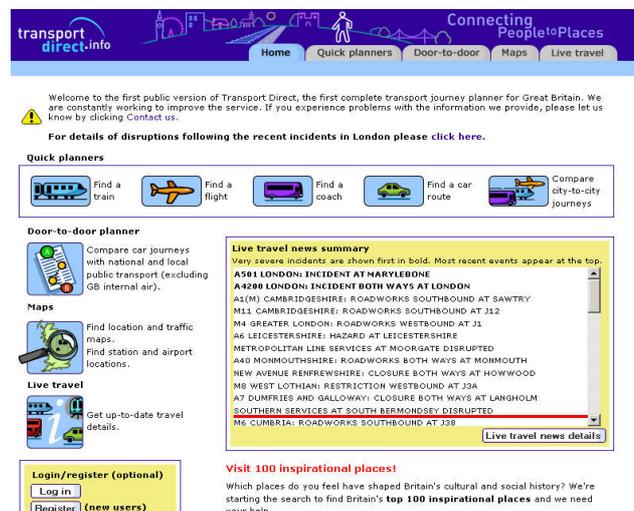


Figure 3. UK Department of Transport’s Transport Direct portal

It brings together a large number of remote data sources and applications from different transport companies, regional and national government, police and emergency services and a number of independent travel and consumer organisations.

Given the volume and dynamic nature of regional travel information attempting to centralise this information was impossible. Transport Direct therefore establishes a central portal which users see, and the portal then interacts with a large number of independent services to build route plans and execute individual parts of the service. Central data and application servers focus on system administration, managing communication with remote services, and the user interface, while remote servers are used to process travel queries.

The Transport Direct may be found at: <http://www.transportdirect.info/TransportDirect/en/Home.htm>.

5. CONCLUSIONS

Portals expand access to existing spatial resources. Each of the examples discussed above builds on existing resources and functionality, making them available to a wider community. Expanding access and usage is critical to achieving return on investment on SDI and spatial services development. This will become more important as commercialisation of certain spatial information services (for example: navigation and tracking technologies, and initiatives from key industrial players like Microsoft and Google) stimulates public demand for spatial data and services. Portals are evolving rapidly to meet these

needs, and focusing on usability for power and casual users alike, accuracy, interoperability and speed to do this.

Certainly the time when data and applications can be shared seamlessly across the Web is not yet with us, but developments such as W3C's Semantic Web are beginning to build the standards and structures that may make this feasible in the not too distant future. Spatial portals are beginning to illustrate this potential, and to sustain and strengthen demand and justification for the underlying frameworks and standards that will make the goal of seamless data and application sharing possible.

REFERENCES

- ESRI, 2004. *GIS Portal Technology*. ESRI Whitepaper. www.esri.com/library/whitepapers/pdfs/gisportal.pdf (accessed 12 Jul 2005).
- GSDI, 1999. *The Strategic Imperative of a Global Spatial Data Infrastructure*. <http://www.gsdi.org/docs1999/stratim.html> (accessed 12 Jul 2005).
- Maguire, D. and P. Longley 2005. The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems* 29: 3-14.
- Masser, I. 2005. *GIS Worlds: Creating Spatial Data Infrastructures*. Redlands: ESRI Press.
- Tang, W. and Selwood, J., 2005 (forthcoming). *Spatial Portals: Gateways to Spatial Information*. ESRI Press, Redlands.

