Evolving Infrastructure: Growth and Evolution of Spatial Portals

Jeanne FOUST, USA, Winnie S.M. TANG, Hong Kong SAR, China and Jan SELWOOD, Japan

Key words: GIS, Global, Spatial Infrastructures, Internet/Web, Services, Integration, Interoperability, and Cooperation.

SUMMARY

The term Spatial Portal appeared in the mid to late 1990s to describe the interface to spatial clearinghouses and data infrastructure projects. Since then portals have developed dramatically and now play a central role in the way we discover and consume spatial resources. Their role will increase as developments in technology, design and policy consolidate SDIs and offer new and exciting opportunities for distributed computing. This paper reviews the evolution of spatial portals illustrating different approaches with case studies drawn from around the world, and considers the future and issues that must be overcome if they are to realize their full potential.

Issues considered include the construction and maintenance of resource catalogs, harvesting strategies, quality control and data validation. Of key importance, particularly in international portal initiatives, are the range, depth and consistency of metadata. Attention is also turning to cataloguing metadata of unstructured text, statistical or image documents so often excluded from traditional SDI. Drawing on experience of a number of portal initiatives, thoughts on achieving an effective balance in metadata policy are outlined.

As portals evolve beyond simple search engines, visualization becomes increasingly important. The paper goes on to consider techniques for visualization including current Web mapping capabilities and limitations, and the opportunities to visualize portal search results with traditional desktop applications and the move towards Service Oriented Architectures.

Finally the paper looks how spatial portals build relationships and strengthening communities within SDI initiatives. It considers management techniques and portal interface designs that have been developed to achieve this.

Evolving Infrastructure: Growth and Evolution of Spatial Portals

Jeanne FOUST, USA, Winnie S.M. TANG, Hong Kong SAR, China and Jan SELWOOD, Japan

1. INTRODUCTION

Since its introduction in the late 1960s Geographic Information Systems (GIS) has worked to improve access to geospatial information and analysis. It linked map graphics and related textual attributes, overcoming a problem that had plagued geospatial representations since man started drawing maps and compiling gazetteers. It provided new ways of encapsulating geographic knowledge and concepts so that they can be effectively stored, managed and analyzed – a new language for describing and studying the world (Dangermond, 2004). It simplified integration and overlay of geographic data from different sources and in different geographic projections, and by improving interoperability enabled exchange of information services in many different formats. It released the map from the physical constraints of the hardcopy map sheet, allowing multiple users to simultaneously access the same continuous, seamless datasets. It took maps and geographic data from the shelves and cabinets of map libraries, allowing users to work with spatial information on their own desktop computers, and share and access it across local and wide area networks.

Spatial portals are gateways through which users can disseminate, discover and access of geospatial information (Tang and Selwood, 2005). Their evolution can be viewed as a further, important step towards the goals that excited the first GIS practitioners back in the 1960s, and that have guided the development of GIS as a discipline since. Their role will increase as developments in technology, design and policy consolidate SDIs and offer new and exciting opportunities for distributed computing. This paper reviews the evolution of spatial portals, illustrates different approaches with case studies drawn from around the world, and considers the future and issues they must overcome if portals are to realize the full potential of SDIs.

2. THE EVOLUTION OF PORTALS

The computer industry adopted the word ‘portal’ in the mid 1990s to describe Web sites that either assemble many online resources and links into a single location to form easy-to-use products (e.g. AmericaOnline or CompuServe), or provide search tools that helped users find information on the Web (e.g. Yahoo! or Google). Derived from the medieval English *portle* meaning city gate and originally from the Latin *porta* meaning simply gate, these sites aim to be their users’ primary “point of entry” to the Web – their ‘gateway’ or ‘portal’. They proved extremely popular. As the number of users and the volume of content on the Web grew exponentially throughout the 1990s, portals provided a convenient way for the casual user to navigate what was otherwise becoming an impenetrable mass of information. By connecting Web user with Web content provider portals played a significant role in the development and popularity of the Web.

2.1 Portals and the GI industry

Portals were introduced to the Geospatial Information (GI) industry in the late 1990s as the gateways to Spatial Data Infrastructure (SDI) initiatives. Studies conducted in the late 1980s and early 1990s by organizations and governments around the world revealed that despite considerable progress in GIS technology and capturing digital geospatial information, significant barriers remained to its efficient distribution and use. 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.

2.1.1 Catalog portals

Spatial portals associated with SDI are generally what are commonly known as 'catalog' portals. They create and maintain indexes or 'catalogs' of metadata that describes the nature and location of resources in the SDI. Resource owners (also called 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. Users can use the catalog to search for services coming from any of the registered service providers. In most cases the provider continues to host the services to which the metadata refers and the portal simply connects the user with the service(s) in which they are interested. They are, as it were, the broker between users and service providers. Users have a single location through which they can access up-to-date, authorized information, and the service providers have a single location through which they can reach large numbers of users.

2.1.2 Application portals

Though catalog portals and the early development of spatial portals were associated with SDI, new types of portals have evolved. The rise of Web service technology as a robust vehicle for combining and serving complex geospatial data and functionality across the Web, has allowed organizations to establish dedicated Web-based mapping packages that undertake particular tasks or applications. These 'application' portals package a range of services to meet the needs of well-defined audiences or user requirements. In addition to the generic search tools of the catalog portals, they provide structured user interfaces that guide users to required services. As the users and user requirements are 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, and so on). The distinction between application and catalog portals is blurring as catalog portals increasingly add visualization and analysis tools.

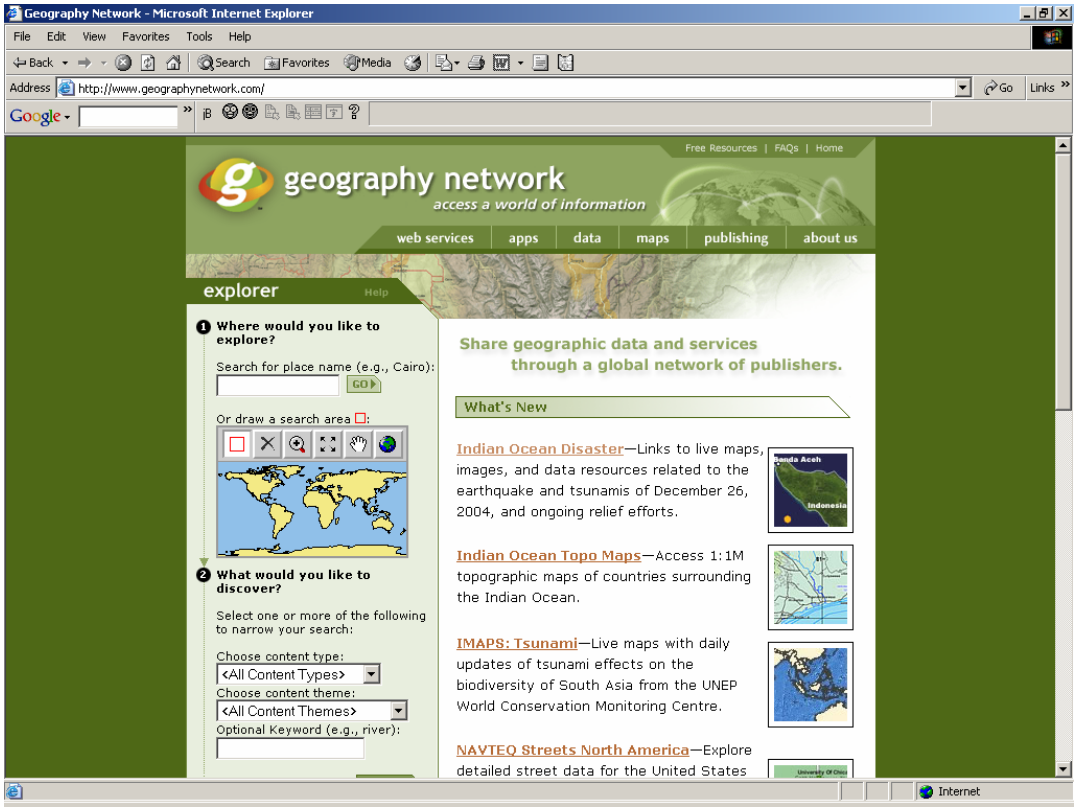

Figure 1: The Geography Network was one of the first spatial catalog portals to include graphic-based spatial search and visualization functionality.

2.1.3 Enterprise portals

A third type of spatial portal that is beginning to emerge is the 'enterprise' spatial portal that integrates spatial data and functionality with business enterprise solutions. Enterprise solutions appeared in the late 1990s to help large organizations manage distributed information resources. Such systems have concentrated 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. This allows users to switch easily from data in document or spreadsheet form to view related mapped data without having to leaving the corporate portal. There is a trend to not only bring spatial resources into the portal, but also 'spatialize' the portal. Such portals adopt location as the primary means of ordering and searching enterprise information. This recognizes that much spatial data is often stored in free text format - contract documents, letters, reports, emails, spreadsheets, images, Web pages and so on. Tools to search and index spatial information held within such 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.

3. EVOLVING METADATA

The development of spatial metadata has been a critical factor in the growth of spatial portals (Maguire and Longley, 2005). Metadata, structured documentation of information services, is an essential prerequisite for building catalogs in catalog portals. While application and enterprise portals do not necessarily depend on metadata archives, they often include them so that users know the provenance of the services they find. The SDI initiatives beginning in the 1990s raised the GI industry's awareness of the importance of metadata, and fostered a number of important metadata standards and the creation of large quantities of metadata.

The process of compiling metadata into catalogs, and the increased access to catalogs that spatial portals offered highlighted a number of important issues.

Firstly, differences in interpretation and use of the metadata standards became apparent as metadata records produced by different organizations were loaded into portal catalogs. Compiling catalogs, and attempting to create search tools that produced meaningful results, highlighted diversity in the way individual organizations (and often parts of the same organization) interpreted and applied the same standard. This experience has lead to standards organizations and industry groups tightening their specifications and issuing greater guidance on how certain fields should be interpreted and completed. Industry groups or large organizations are increasingly using gazetteers, keyword dictionaries and thesauri to help standardize entries for key fields.

Secondly, portals highlighted a distinct difference in the way users interact with metadata. Portal managers found that the searches undertaken by the majority of users focused on a relatively small number of key parameters. Comprehensive metadata records continue to be important in order to ensure services are fully documented, however, only a fraction of these fields are actively used for searching. This is important, as it helps refine guidelines for metadata entry. Attention can be focused on actively enforcing standards for those key parameters on which searches are most commonly undertaken and for which standardization is essential, while more flexibility may be allowed in other parameters that are less frequently used for catalog searches. Interest in slim metadata standards, such as the Dublin Core (which has only 15 core elements in comparison to Federal Geographic Data Committee's (FGDC's) 334), that can be used specifically for search and retrieval purposes is increasing (e.g. CEN, 2003). Such refinement helps to ensure metadata creation, maintenance and use is efficient.

Developments in portals and SDI are also expanding the range of services that metadata must describe. In addition to data services, portals now commonly catalog: spatial models, applications, component tools, methodologies, reports and so on. They serve new types of data services including large continuous stream image libraries and real time data feeds. Current metadata standards were not designed to document such a broad range of spatial services. Work is therefore required, and indeed is ongoing (FGDC, 2002) to refine metadata formats to better cater for the kinds of services now being registered at spatial portals. In addition, more work on indexing and providing metadata for unstructured spatial data records is required.

4. EVOLVING CATALOGS

A key design consideration within spatial portals is how catalogs are built and maintained. Portals generally provide tools for service providers to create, maintain or upload metadata in agreed formats that describe the services they offer. This manual procedure can become quite inefficient for service providers, as they must remember to update the portal whenever they modify or create new metadata records. For large information suppliers who register services on multiple portals this can get time-consuming. Portals have therefore developed automatic routines that can access remote metadata databases over the Web. Service providers can maintain their own metadata databases and the central portal can access these and automatically extract new or updated information.

Harvesting techniques have also evolved over time. Initially, particularly in early SDI initiatives, a distributed approach was adopted. In this, metadata was actually held at the service providers' sites, which the portal queries only when it receives a search request from a user. While an elegant approach that reduces duplication (and the resultant potential for inconsistency), this proved difficult to implement as search speed and reliability was entirely dependent on the robustness of the network and individual metadata databases. If one database or network connection fails, search results will not be complete. In addition, it is hard to ensure consistency throughout the metadata collection, as individual organizations are responsible for maintaining their own metadata archives with less central, coordination.

Such problems have prompted the development of a more centralized approach. In this, the portal routinely collects (or 'harvests') metadata from remote sources and consolidates it in catalogs stored within the portal database infrastructure. Creating central metadata databases that consolidate metadata from all registered service providers helps to ensure scalability and reliability of search performance. In addition, portal managers can more easily check and control the quality and completeness of metadata records. Most portals now adopt a combination of automatic harvesting and manual update to maintain their catalogs. Automatic harvesting techniques from the library and archive industry are increasingly being adopted (for example those based on the Z39.50 and Open Archive Initiative (OAI) protocols).

5. EVOLVING APPLICATIONS

Portals have taken advantage of general developments in the IT industry to expand range and complexity of the applications they offer. Initially conceived as basically permitting text-based query functionality of metadata catalogs, portals now provide sophisticated tools for not only discovering and exploring data, but combining, analyzing and using the services found. Tait (2005) points to the importance of Web services and Service Orientated Architectures (SOA) that have permitted such functionality to be delivered. Industry standards such as eXtensible Markup Language (XML), Simple Object Access Protocol (SOAP) and Web Services Description Language (WSDL) underpin the interoperability that portals require as central gateways to diverse services, and permit services to be combined, integrated and shared not only between different geospatial technologies, but with other IT technologies.

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 portals 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 over the Web. 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 translation applications supplied by companies such as SAFE software that allow on demand translation between numerous different formats.

Depending on the type of portal, such mapping clients range from those providing 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.

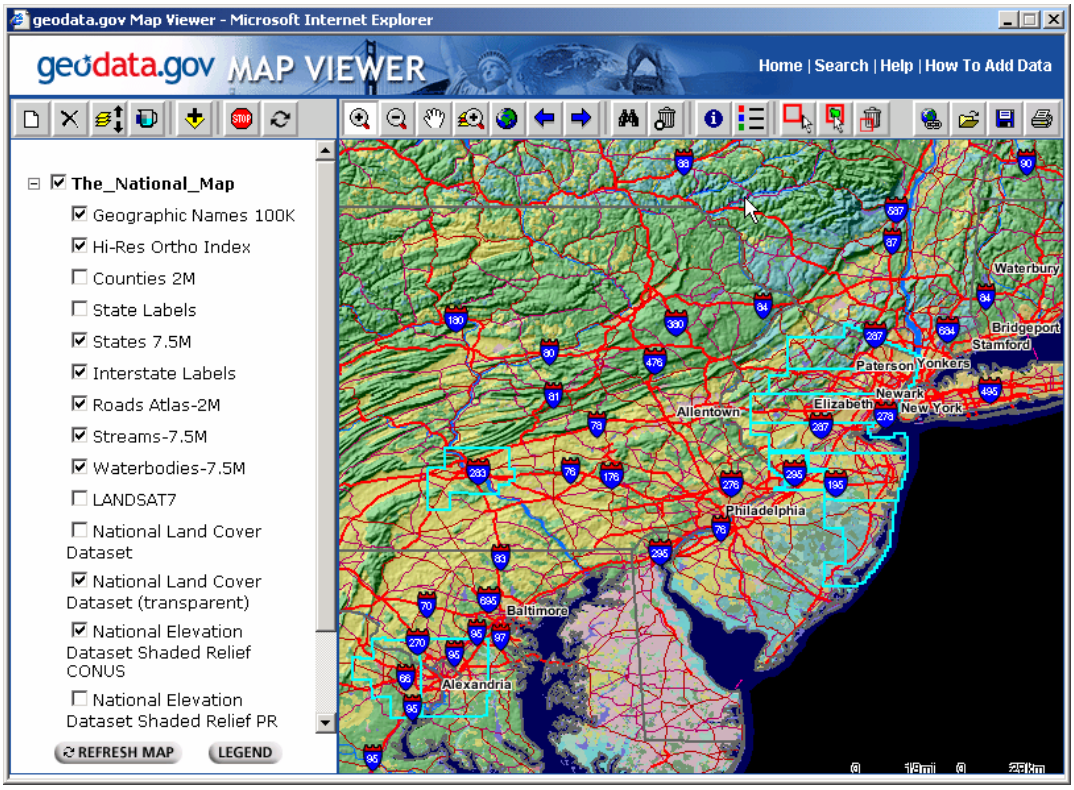

Figure 2: The Map Viewer of the Geospatial Onestop allows users to add, view and work with data from multiple remote services within a user-friendly environment.

Applications available through enterprise portals are also becoming more sophisticated as server-based GI technology appears, providing comprehensive GI functionality that can be deployed across and embedded within enterprise information systems.

6. EVOLVING COMMUNITIES

One of the key goals of many SDI initiatives is the expansion and development of user communities. This may include: promoting awareness and use of geospatial resources, defining and supporting industry-wide standards, and enhancing coordination and collaboration within the geospatial industry. SDI provide a forum for information exchange, raise awareness of the importance of metadata and availability of resources. Portals have contributed towards this aim. They greatly simplify the ease and user friendliness of search and access activities. Combining text and spatial search functions, and the ability to view and work with the resources these find makes the information discovered through the portal far more immediate and accessible to a wider audience. Techniques that guide users to relevant information, as for example the design of application portals that target particular audiences, or the introduction of dedicated 'channels' that collect links and information on a particular area or subject help to make the search process more efficient. Maguire and Longley (2005) note direct access to online services (in addition to metadata), and access to such services through both Internet Mapping applications (thin clients) and standard desktop applications (thick clients) as being particularly significant.

Links within the geospatial community can also be strengthened by collaborative work on the design and development of the portal. This may take the form of distributed stewardship of parts, or all, of the portal. So for example, different industry segments may organize expert groups to maintain channels dedicated to their field, a process that may include designing the look and feel of the channel, searching and vetting channel contents, and undertaking regular updates and 'channel-cleaning' operations. Channels (and portals in general) provide a forum for information exchange and discussion that is no longer restricted by distance.

7. FUTURE DIRECTIONS

SDI and spatial portals take a further step along the route that GIS has been charting since the 1960s of improving access to, and use of, geospatial information and analysis. They expand the user community, improve communication within that community and make it easier to find, evaluate and use geospatial resources. Much work, however, remains to be done. This includes encouraging greater participation and involvement. Work is needed to make metadata development and consolidation more efficient, and to develop crosswalks and translations between standards used in different industries or regions. Portals that permit multilingual search or registration of services remain rare and require more attention. The increased ease with which data can be accessed and integrated, highlights the urgent need to standardize the way in which data or applications are conceptualized and modeled. SDI, spatial portals, and progress on standardizing data format and sharing datasets in different formats have overcome many barriers to information exchange. Semantic differences – differences in how a land cover or soil type class is defined, or the way an algorithm models a particular process – now present far greater barriers to interoperability than physical access to data or the data format in which it is held. It is important also to recognize that the framework on which SDI and portals are built assumes reliable access to the constant, reliable Internet access and reasonably high bandwidth. In large parts of the world such an environment does

not yet exist, and ways of fostering it are urgently required. All of these tasks require attention. However, while challenges remain, spatial portals and SDI provide an increasingly robust, extensive infrastructure through which the GI community working together can address them.

REFERENCES

Dangermond J, 2004, "*GIS and Geography: A new language for society*", Schermerhorn Lecture presented at Institute for Geo-Information Science and Earth Observation, Enschede, The Netherlands

European Committee for Standardisation, 2003, "*Mapping Between Dublin Core and ISO19115, Geographic Information Metadata"*, CEN, Brussels.

Federal Geographic Data Committee, 2002, "*Content Standard for Geospatial Metadata: Extension for Remote Sensing Metadata*", FGDC, Virginia.

Maguire DJ and Longley PA, 2005, "The emergence of geoportals and their role in spatial data infrastructures", *Computers Environment and Urban Systems*, Vol. 29:1, 3-14.

Tait MG, 2005, "Implementing geoportals: application of distributed GIS", *Computers Environment and Urban Systems*, Vol. 29:1, 33-47.

Tang W and Selwood JR, 2005, "*Spatial portals: gateways to geospatial information services*", ESRI Press, Redlands CA

BIOGRAPHICAL NOTES

Ms Jeanne Faust: Ms Faust is ESRI's Global Manager of Spatial Data Infrastructure. She is an active member of a number of US and international committees and industrial associations promoting SDI and GI interoperability. Ms Faust has a MSc. degree from the University of South Carolina.

Dr Winnie Tang: Dr. Winnie Tang is the founder and CEO of ESRI China (Hong Kong) Limited. She is recognized as one of the highly regarded leaders in IT industry in Asia, and sits on a number of academic and industrial associations. In 2001, she was awarded the 'Ten Outstanding Young Digi Persons Selection' in Hong Kong in recognition of her contributions and achievements. Dr Tang and Mr Selwood are co-authors of "GIS Web Services: Connecting our World" published in 2003, and are currently working together to prepare "Spatial Portals: Gateways to Geospatial Services" due out in early 2005.

Mr Jan Selwood: Jan Selwood has worked in the GI industry for over 15 years many of which have been spent leading system design and implementation projects in Asia. He is an independent GIS consultant and works extensively with ESRI China (Hong Kong) as a consultant. He has an MSc. degree in GIS from Edinburgh University, UK. Mr Selwood and Dr Tang are co-authors of "GIS Web Services: Connecting our World" published in 2003, and are currently working together to prepare "Spatial Portals: Gateways to Geospatial Services" due out in early 2005.

CONTACTS

Ms Jeanne Faust
Global Manager of Spatial Data Infrastructure
ESRI
380 New York Street
Redlands, CA 92373
USA
Tel. + 1 909 793 2853
Fax +1 909 793 5953
Email: jfaust@esri.com
Web site: www.esri.com

Dr Winnie S.M. Tang
Chief Executive Officer
ESRI China (Hong Kong) Ltd.
Level 10, Cyberport 2
100 Cyberport Road
HONG KONG
Tel. + 852 2730 6883
Fax + 852 2730 3772
Email: wtang@esrichina-hk.com
Web site: www.esrichina-hk.com

Mr Jan Selwood
Consultant
ESRI China (Hong Kong) Ltd.
Level 10, Cyberport 2
100 Cyberport Road
HONG KONG
Tel. + 852 2730 6883
Fax + 852 2730 3772
Email: jselwood@esrichina-hk.com
Web site: www.esrichina-hk.com