Technical Architecture for GSDI

Doug Nebert, Technical WG Chair
U.S. Federal Geographic Data Committee, GSDI Secretariat
12201 Sunrise Valley Drive
Reston, Virginia 20192 USA
Phone: +1 703 648 4151
Fax: +1 703 648 5755
Email: ddnebert@fgdc.gov

Doug Nebert is the Clearinghouse Coordinator for the U.S. Federal Geographic Data Committee (FGDC) Secretariat in Reston, Virginia. He has worked for the past 17 years for the U.S. Geological Survey first on water resources applications of geographic information systems, then metadata standards and software, and finally, with the FGDC on standardized methods of GIS data dissemination to promote discovery and re-use. Doug is the Technical Working Group Chair of the Global Spatial Data Infrastructure (GSDI) activity and chair of the Architecture and Catalog SIGs of the OpenGIS Consortium. He holds a Bachelors Degree in Environmental Studies from Evergreen State College (Olympia, Washington, USA) and a Masters of Science in Geography from Portland State University (Portland, Oregon, USA).

Abstract

The Technical Working Group (TWG) of the Global Spatial Data Infrastructure (GSDI) was convened to evaluate practical techniques for creating a consistent trans-national spatial data infrastructure, to publish an implementation guide to promote adoption of recommended practices, and to develop and share reference software implementations that support the technical infrastructure of the GSDI. The SDI Implementation Guide, known as the GSDI "Cookbook," was published in March 2000 as documentation of a full suite of known and emerging practices, standards, specifications, and case studies from local, national, regional, and international sources. As a dynamic document, the Cookbook contents continue to introduce new discussion on recommended practices, and standards from various points of view. This paper presents several new views of an interoperable Internet-based technical architecture that accommodate the diverse supporting technologies of a spatial data infrastructure.

This paper seeks to describe a current, interpretive view of technical spatial information infrastructures that transcend data with networks of standards-based, data-enabled services that can be used to satisfy a broad variety of requirements.

Comparative View of Spatial Data Infrastructures

A review of five national spatial data infrastructure initiatives and the GSDI shows the use of a similar terminology and scope to define the bounds of the initiative. As these definitions are intended to guide broad activities, they are not expected to be precise in including or excluding features of SDIs.

	GSDI	Australia (ASDI)	Canada (CGDI)	United States (FGDC)	South Africa (NSIF)	United Kingdom (NGDF)
Technology	technologies	distribution networks	(implied)	technology	technologies	(implied)
Policy	policies, organizational remits,	(implied)	supportive policy environment	policies	policies	(implied)
Standards	standards	technical standards	standards	standards	standards	standards
Institutions	financial and human resources	institutional framework	partnerships	human resources	institutional arrangements, human resources	data providers
Data	data	fundamental datasets	data framework	geospatial data	spatial information	Standard Geographic Base
Supported Functions	delivery mechanisms	basis for value-adding services	access	acquire, process, store, distribute, improve utilization	exploit	discovery, access, integration, use

Table 1. Comparison of terms used to define national SDIs and GSDI [GSDI, ANZLIC, CGDI, FGDC, NSIF, NGDF]

Table 1 shows a clear pattern of similarities in the terms and coverage of the full activities known as SDIs in this sample of five countries (plus the GSDI). From a technical perspective, strong support for consistency in technology, standards, data, and supported functions is evidenced in the definitions evaluated here. In its current definition, the "GSDI encompasses the policies, organizational remits, data, technologies, standards, delivery mechanisms, and financial and human resources necessary to ensure that those working at the global and regional scale are not impeded in meeting their objectives" (http://www.gsdi.org). The GSDI Technical Working Group has been extending this definition and sees a major role of the GSDI and its membership to perform the role of an advocacy group for the deployment of recognized standards and practices. This commitment is required to encourage a truly global network of interoperable spatial data infrastructure member activities while recognizing local, national, and regional cultural and technical issues of implementation. In the sessions of the Steering Committee of the GSDI during this Cartagena conference, discussion of a revised scope and definition of the GSDI is planned to support the technical and organizational growth of GSDI and its members.

The definitions of SDIs above do not detail the specific types of data required or functions to be performed, but set a vision or scope for guiding broad often multidisciplinary activities. From a technical perspective, multi-lateral organizations interested in developing integrated systems appear to be identifying the same ingredients. Many of these infrastructural elements are not unique to spatial information but must be recognized for their specific role and implementation when applied in a geographic context.

GSDI Implementation Guide: The Cookbook

An implementation guide, known as the GSDI Cookbook was written by GSDI TWG members and contributors to:

- Provide a forum for discussion, definition, and reference for common GSDI elements
- Identify and recommend existing standards, specifications, and where available reference implementations of software that support common infrastructural deployment
- Support a continuum of readers through a multi-tier design to assist managers, project leaders, and implementers in understanding and interpreting GSDI recommendations

The Cookbook is organized in chapters that progressively reflect new capabilities that build on previous chapters' discussion. The progression is made from discussions of fundamental spatial data and its collection and specification, through its documentation (metadata), its discovery (catalog), and its use in a variety of services including the most popular today, web mapping. Additional chapters provide case studies and capacity-building issues of SDIs and SDI-like organizations and a terminology chapter drawn from the words used in the Cookbook. The Cookbook has undergone two minor revisions since its initial publication and presentation at the GSDI-4 conference in Cape Town, South Africa in March 2000. A new digital version is planned for distribution at the Cartagena conference.

Most countries, based on available resources and technologies, have focused on the initial issues of data development and discovery and are only slowly progressing towards more advanced dissemination and use of the data in the Web environment. The deployed systems, standards, and specifications that simplify this task are still quite new and are in the early stages of mainstream commercialization and availability. In addition, the constraints imposed by local and regional communications infrastructure still precludes citizen access to complex geographic information over the Web in many parts of the world.

Beyond Data

A data-centric focus is obvious even in the names we apply to these spatial "data" infrastructures, when in reality we really intend to state that data is critical but certainly not the only main component of a viable SDI. As discussed in the Cookbook, SDIs start with spatial data, but then include necessary services to make the data useful in various contexts. Organizations such as the OpenGIS Consortium and the International Digital Earth Initiative are working to identify and standardize on specific classes of geospatial information services, which the Cookbook likewise seeks to promote. Over time, mature SDIs will implement many types of geospatial information services that can be combined through their standard software interfaces into broader capable systems that can be applied for many uses.

In the Open Systems Interconnection designs of the 1980s, a tiered model of networking services was published that illustrated the interdependency of various communications protocols that layer on one another to form the computer networks that we count on today. Known as the OSI “stack” model, it established a precedent for organizing and describing what was otherwise a tangled matrix of standards, protocols, and interfaces into a coherent hierarchy in which levels were defined by the function they performed and the information types they work with. The OSI stack model provides a suitable metaphor for considering geospatial information services as well.

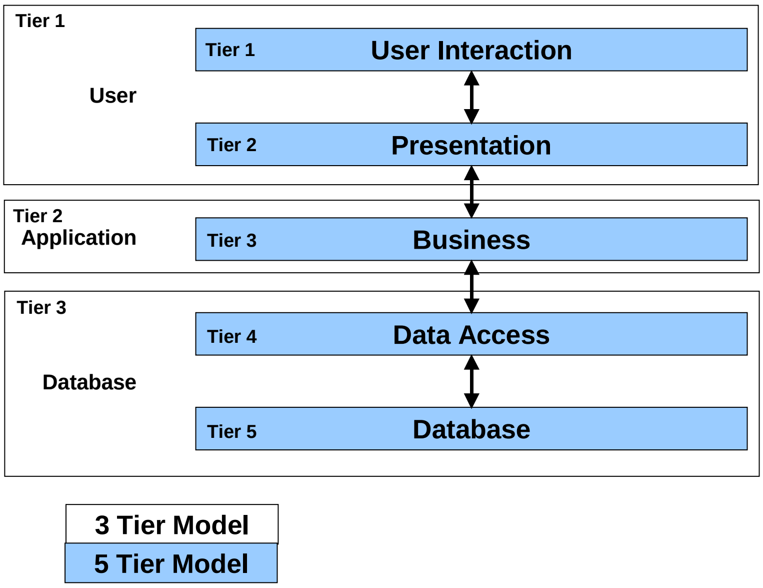

Figure 1. Generic 3 and 5 Tier Architectures

Figure 1 is a recent diagram from the Open GIS Consortium to help refine the interaction between different types of software interfaces (shows as arrows) that could be exposed in a Web services architecture. Instead of only showing data in a simple client-server environment, this model illustrates how a multi-tier architecture may function. This middle tier – what is popularly known in the Web environment as an “application server” – can play a role to discover and integrate data from multiple locations and perform special operations on the data, such as make and symbolize a map from external data, and deliver it to the user.

In these models one sees many different interactions between data and processes, called interfaces, where the contents and structure of a request and a response could be described. Some examples are given below:

- data access – specific spatial features are requested through a query and a specific format of the response is expected in return
- coordinate transformation – data in one coordinate reference system are converted into another coordinate reference system
- symbolization – vector features from a data access service are symbolized based on rules selected by the user or application and displayed
- web mapping – a request for a specific data source, symbol style, geographic extent, and format is made and a bitmapped map picture is returned
- navigation – a starting location, reference to a linear network or surface, and an ending location are given and a vector path and/or navigation instructions are returned

Figure 2 shows a modified view of the stack model with respect to some of the known service interface types we see in a Web services architecture for geospatial information based on the experiences in the Internet and the OGC Testbed activities.

Data sources represent repositories of information that have native (usually proprietary) interfaces on them. By themselves they represent stores of information that could be used in other contexts. Spatial data represents raster, vector, point, and surface information and the attributes and relationships inherent between them. The metadata is tied to the data or service instances being described. *Registered information* is a special category of supportive information that can be used by a community, including namespaces, valid values, schemas, keys, identifiers, and service type descriptions. Perhaps this is a type of metadata but it largely represents type collections or a standard vocabulary/dictionary that is to be used for reference by software and humans.

Access services, in the OGC context, represent basic public (standard) interfaces to the data sources. These supplement or replace native interfaces to information, representing access building blocks to all types of data – not just geospatial. Validation, registration, lookup, and authentication services are targeted at registered information types of data. Catalog services are targeted at metadata. Data access services are targeted at spatial data in this case. It could be argued that these access services are all specializations of an access service in some way.

Integrative services are where the integration of multiple information sources takes place or a program works upon information provided by an Access service. In a Web services/application server environment, these services may take place on one or more servers. In a robust client environment they may take place on the client. The location is less important than the functions performed. Typically, data access interfaces are used to feed these services and either data service interfaces or user interface components are exposed. This lets services be used or chained.

The ***user interfaces*** tier represents the client software ultimately residing with the client. It may be local to or remote from the integrative services. This tier is shown to remind developers of the important role played by application servers or service interfaces in supporting both software applications and light browser clients.

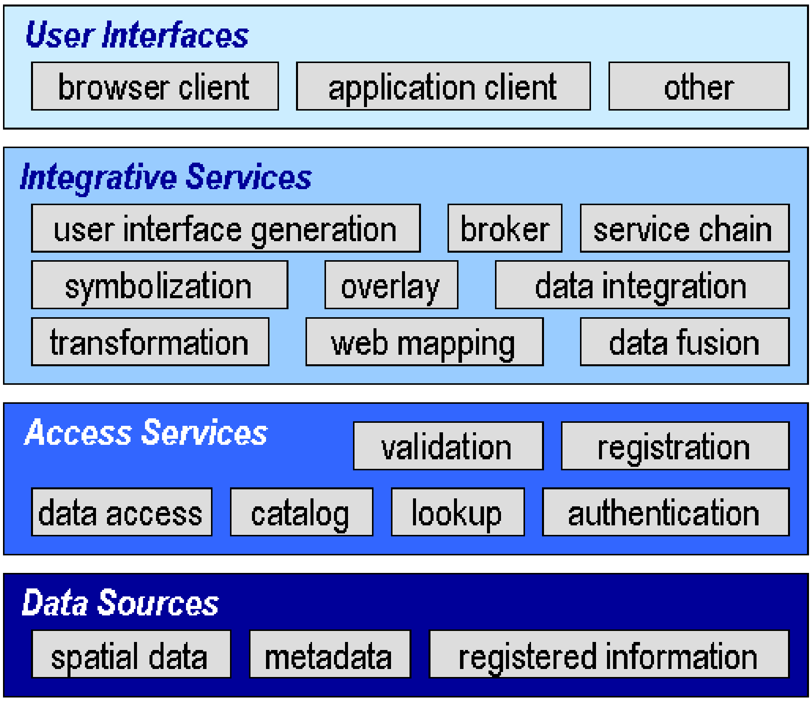

Figure 2. A "Stack" model of geospatial information and services

This stack approach to dissecting spatial services has another useful application – that of helping to define the scope of what sponsoring organizations define as the supported "infrastructure" as opposed to a larger enterprise. According to the Webster dictionary [WEBSTER], infrastructure is either "the underlying foundation or basic framework (as of a system or organization)" or the system of resources required for an activity. In either case, the role is a supporting one to broader activities and applications. One could argue, based on the evidence of the intent and implementations to-date that most SDIs are focused on providing the foundation on which applications can be reliably written and deployed. In these cases, the infrastructure supports, but does not include such applications. Reviewing Figure 2, one could interpret that the implementation of most of the top two tiers is not the responsibility of SDI organizations, but of their constituency or user base; the applications ride on the infrastructure.

Growing an interoperable network

Between each of components in these tiers exist native and standards-based interfaces allowing, in an ideal environment, the construction of linked services in a directed way. One can begin to see how the outputs from one request (e.g. data access) could conceivably become the input to another process (e.g. coordinate transformation) and even to another (e.g. symbolization/web mapping). This ability to chain services together is of great interest to software developers wishing to create new applications for decision support or alternatives analysis quickly.

The discovery and invocation of these many types of software operations will require a discovery and description facility beyond that of a pure data catalog. The Services Architecture Special Interest Group (SIG) and the Web Mapping SIG of the OGC recognize the need for a service registry to unambiguously discover and execute geospatial services that reside either on the Web or in an Intranet environment. A services catalog will need to make searchable metadata on the style of the software interface, data associations and input/output types, the hosting organization, and other descriptive information sufficient enough to evaluate its fitness for use and support its invocation over the Web.

Within the vendor community on the Web there has been recent activity to develop and deploy an environment where Web services can be located and invoked. This initiative is known the Universal Description, Discovery, and Integration (UDDI) specification. According to their documentation, “UDDI is the building block that will enable businesses to quickly, easily and dynamically find and transact business with one another using their preferred applications.” [UDDI] With the industrial backing of Microsoft, IBM, Ariba and over 120 implementing companies, UDDI promises to be a framework that the SDI community could specialize and exploit for services discovery – as a complement to, but not a replacement of, data discovery services. The OGC will be formalizing its services architecture over the coming year through sponsored test bed and pilot project activities between member sponsoring organizations, integrators, and vendors.

The previous figures show how classes of services might interact with each other within an SDI but do not illustrate an implemented environment. Figure 3 uses an informal modeling notation to describe the potential interactions between the information objects and functions present in the U.S. NSDI Clearinghouse. In this diagram, read relationships between objects using the directionality implied by the arrows – “Metadata are derived from Spatial Data.” White text in boxes indicates a planned or prototyped functionality whereas black text indicates existing functionality. Dashed lines indicate a conditional or planned interaction. Application Programming Interfaces (API) and user interface (UI) symbols indicate potential or actual opportunities for software interface specifications and user interaction, respectively. Not all relationship types are shown, but fundamental ones are present, thereby illustrating the flow of information regarding data (and metadata) and the basic, planned interaction of specific services that permit discovery of, mapping of, and access to geospatial data.

Complex interactions such as service chaining or the construction of special directed solutions for decision support or ad hoc geospatial analysis are not shown in the figure. The use of data catalogs and service catalogs that are aware of each other will make possible the construction of a more elaborate and responsive “data workbench” for geospatial analysis fed by currently available services and real-time data feeds.

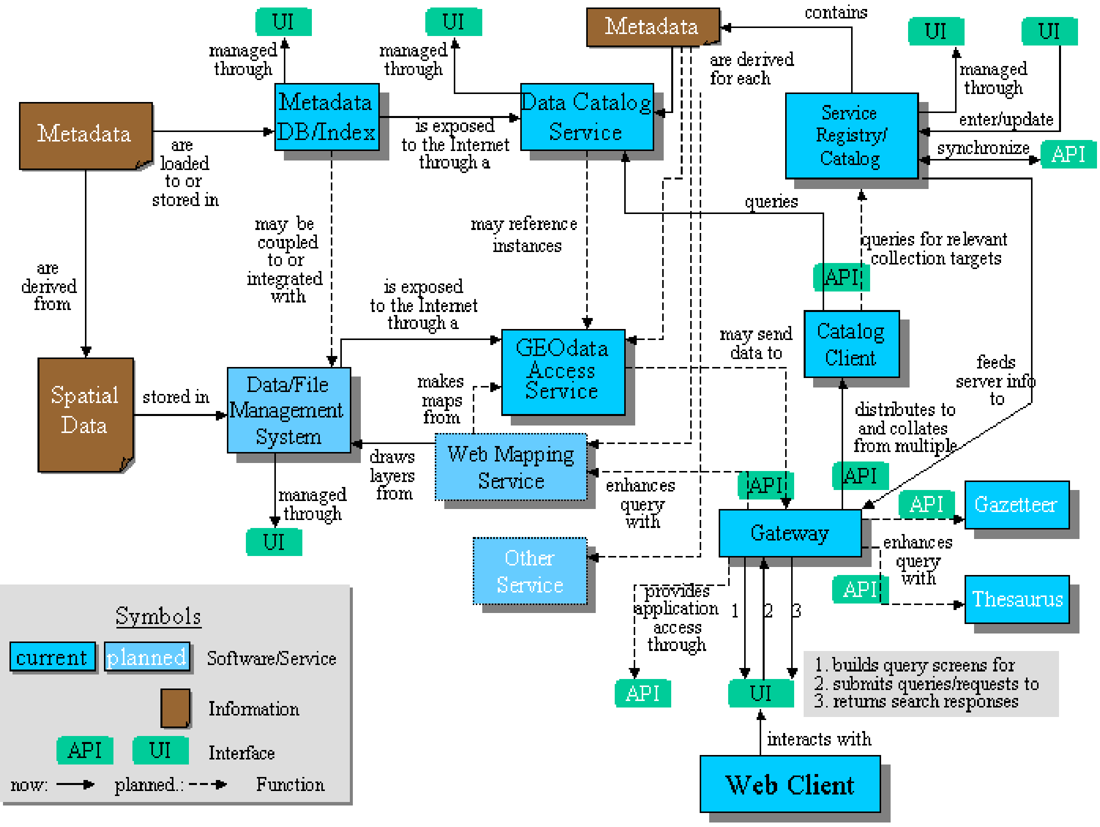

Figure 3 – Object Interaction Diagram of the U.S. NSDI Clearinghouse

Issues

The majority of the problems in constructing a coherent and coordinated GSDI are not technical, but derive from differences in meaning. By their nature, SDIs unite many different application domains, disciplines, users, and providers.

Data produced for one purpose may (or more likely may not) be useful or suitable to use in another domain. The desire to define Framework or fundamental data content models is noble and obvious and yet the organizational requirements for mission-critical data elements – especially within a given domain – may cause disagreement on what should constitute an adequate or minimal representation for community use. Within the SDI architecture, schemas or models of specific packages of information may be developed and registered and discovered by participants, once they agree on the content. The majority of the work will be procedural and organizational: developing consensus on data content and schemas and in providing the data in these packages.

Broad information communities need shared information (semantic) registries as a dictionary of common vocabularies, operations, valid values, and data structure definitions. The construction of a semantic registry that operates as an elaborate electronic dictionary is technically feasible, and simple hierarchical data element-style dictionary specifications do exist, such as ISO 11179. Although the topics of semantic repositories, ontology, and reasoning are the subject of extensive academic research, some interesting enabling systems are being deployed on the Web to help store and expose these networks of meaning. [OIL] A semantic research activity within the OGC is seeking sponsorship this year to undertake work on implementing semantic repositories for trial information communities in connection with other OGC services.

Neither a data-centric approach nor a service-centric approach to information discovery adequately supports the needs of a growing GSDI community. The current catalog models of implemented SDIs primarily inventory data and not invokable services. Since the OGC web mapping test bed it is recognized that a service registry is also needed. Now there are many types of services available that need to be discovered. The interaction of data and services is still the subject of active modeling and debate, however, with a general recognition that both data and service instances will need to be described in special ways and, where dependencies exist, reference each other. In other words, a data set will be able to be associated with every service instance, and vice versa. Advanced service discovery and invocation models must be built with participation of the (geospatial) application community and the Internet business community (e.g. UDDI) with an understanding of how to support general and specific semantic information requirements.

Conclusions

The development of a coherent strategy for global, regional, and national spatial data infrastructures must include an understanding of the types of data, services, and their interaction in a number of environments. By examining how the pieces fit together, the creation of a specific organizational SDI can be done more quickly by example. Where standards-based software interfaces and protocols are adopted, a compatible trans-organizational infrastructure can be built. SDIs will cautiously move beyond data discovery and access as the technologies and capabilities to implement them improve.

The GSDI should more aggressively promote its role as an advocate for standard practices and specifications. There is abundant work to be done to reconcile the technical issues and the policies and practices that, together, will make GSDI a reality in practice. National SDIs will be testing and implementing geomatics standards of the ISO Technical Committee (TC) 211, the interface specifications and draft web services architecture of the OGC, the UDDI services registry, and protocols and recommendations promoted by the World Wide Web Consortium (W3C). This implementation experience will provide a fertile ground for the continued development of the implementation guide (Cookbook) and its selection and recommendation of an infrastructure implemented from proven practices.

Because multiple disciplines will meet to discover, share, and use spatial information and services, the GSDI should consider the infrastructure needed to promote semantic interoperability. By its nature, the GSDI will represent at best many languages (and character sets), and even within one language encounter the need to manage and expose the varied meaning of data, operations, and their characteristics. The deployment of phased support for multi-disciplinary and multi-lingual semantics should be the target of near-term research, from both a technical and policy point of view. We should be able to solve problems with geospatial information in the next few years by:

- searching for data anywhere on the globe,

- finding specific types of information (whose native language may not your own), and
- processing it via chained Web-based services with other data in user-defined ways.

The ability to conduct such work should be a measure of performance of the future GSDI environment as a supportive infrastructure.

References

[ANZLIC] ANZLIC, 1996, Spatial Data Infrastructure for Australia and New Zealand, http://www.anzlic.org.au/asdi/anzdiscu.htm

[CGDI] GeoConnections, 2000, The Five Policy Thrusts of GeoConnections, http://www.geoconnections.org/english/policy/index.html

[FGDC] "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," Executive Order 12906, April 1994, in: Federal Register, Volume 59, Number 71, pp. 17671-17674. http://www.fgdc.gov/publications/documents/geninfo/execord.html

[GSDI] GSDI 2000, GSDI Brochure, http://www.gsdi.org/pubs/press/brochure.pdf

[NGDF] National Geospatial Data Framework, 1998, "What is NGDF?" http://www.ngdf.org.uk/WhatisNGDF/overview.ppt

[NSIF] National Spatial Information Framework, 1999, Vision and Mission Statement, http://www.nsif.org.za/vision.html

[OGC] OpenGIS Consortium , 2001, OGC Web Services Initiative Request for Technology, http://ip.opengis.org/ows/index.html

[OIL] Ontology Inference Layer, 2000, http://www.ontoknowledge.org/oil/

[UDDI] UDDI, 2000, http://www.uddi.org

[WEBSTER] http://www.webster.com/cgi-bin/dictionary?va=infrastructure