Technical Baseline for Accessing a Virtual Global SDI

Douglas D. Nebert
U.S. Federal Geographic Data Committee
Global Spatial Data Infrastructure Secretariat
12201 Sunrise Valley Drive, Mail Stop 590
Reston, Virginia 20192 USA
Phone: +1 703 648 4151
Fax: +1 703 648 5755
Email: ddnebert@fgdc.gov

Abstract

The construction of a cohesive Global Spatial Data Infrastructure (GSDI) as a community depends on the adoption of common vocabularies, practices, standards, and technical specifications, and operational components by participating organizations to facilitate worldwide access to online geospatial information. This paper describes current technical methodologies being implemented in the Information Technology (IT) and geospatial disciplines that are supportive of traditional community building and are essential to the construction of a globally consistent architecture in which geospatial information and services can flourish. These include the consideration of enterprise architecture as a unifying design technique, creation of an online provider and services directory, identification of adopted standards and specifications within an architecture, the publication of spatial data structures, syntax, and semantics to enable their re-use, and the construction of integrative data and mapping gateways and desktop applications that exploit online data.

Background

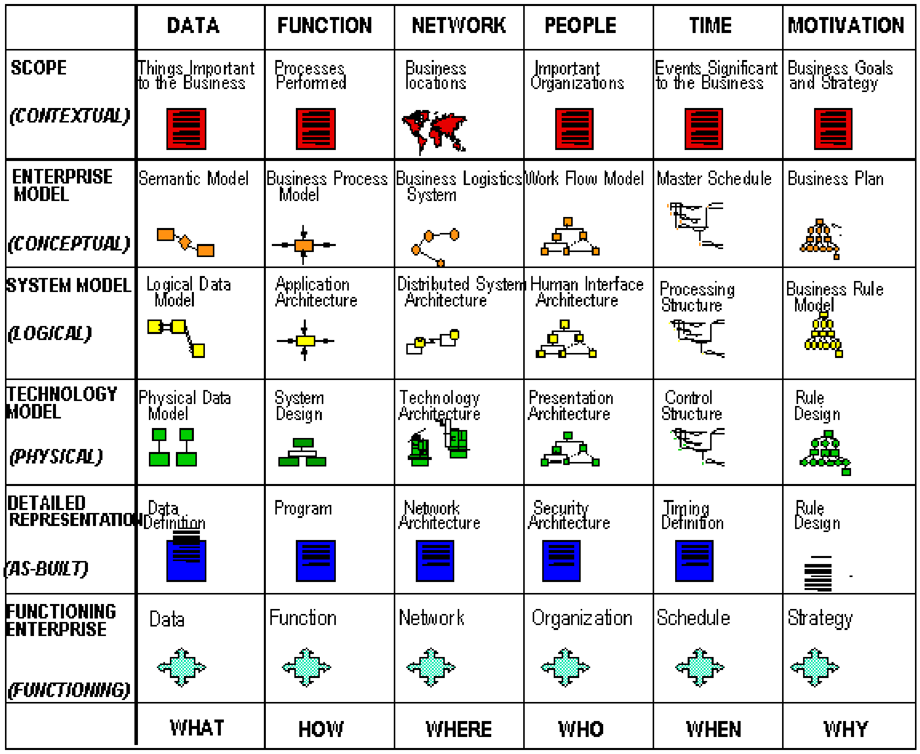
Spatial data and services and applications that access them are typically constructed with a specific problem set in mind. Such single-purpose solutions may work well for solving individual problems, but the systems are often not quickly adaptable to other problem sets. The development of parallel and independent applications and databases is not a unique phenomenon to the geospatial discipline and occurs in all types of organizations. In the 1980s, industrial process re-engineering was implemented in corporations to evaluate and streamline industrial processes within a business and with suppliers and distributors that interact with the business, resulting in improved productivity and competitiveness. Within the scope of a business, the supportive and responsive IT functions are included in the re-evaluation and re-design. The principles of business process re-engineering apply equally well to both the industrial economy and the growing service economy. Information service-oriented organizations can directly visualize their information flows and processes required to meet their recurring mission objectives.

The definition of an “Enterprise Architecture” within a business or governmental agency now requires an evaluation of all functions across the enterprise and the data requirements behind them in order to simplify or expedite repetitive processes. An enterprise architecture is “the set of descriptive representations (i.e. models) that are relevant for describing an enterprise such that is can be produced to management’s requirements (quality) and maintained over the period of its useful life (change).”[1] The scope of the problem is compounded when one recognizes the interaction between organizations – just as exists between traditional industrial suppliers, consumers, and distributors – requires a definition of the ‘enterprise’ at a much higher and inclusive level.

Interagency coordination bodies are emerging within government to identify common lines of business and reusable services in the context of interagency exchange of digital information known as electronic government (e-government) to satisfy the information needs of citizens, commercial, and other governmental clients. The development of an online Spatial Data Infrastructure (SDI) at a local, regional, or national level will need to identify who the participants are in the “national enterprise,” the remits or mission requirements of these participating organizations, the functions they must perform to meet their remits, what information content these functions require, and finally what relevant business process flows exist or should exist within and between the participants. Rather than creating maps or services on data for single purposes, such an approach would provide justification for implementing a community SDI by identifying opportunities for collaboration and a means to quantify costs, benefits, and savings across the broader community in meeting mission requirements.

The construction of an enterprise architecture for an agency or community such as a national SDI will provide a framework for interaction. Figure 1 depicts the “Zachman Framework” that includes the basic constructs of an enterprise architecture. At a minimum, filling in these boxes will help an organization or community to identify the various parts and their interaction. Although enterprise architecture reference models provide illustrative or informative support for systems design, some software is now available to convert diagrams via their underlying notation to software to assist in implementation. What must come from these diagrams is the commitment to establish certain services and adopted practices, several of which are described below, that will help define a tangible presence for a given national or global SDI community.

[1] Information Resource Management Glossary, http://www.cio.gov.bc.ca/other/daf/irm_glossary.htm#E



	DATA	FUNCTION	NETWORK	PEOPLE	TIME	MOTIVATION
SCOPE (CONTEXTUAL)	Things Important to the Business	Processes Performed	Business locations	Important Organizations	Events Significant to the Business	Business Goals and Strategy
ENTERPRISE MODEL (CONCEPTUAL)	Semantic Model	Business Process Model	Business Logistics System	Work Flow Model	Master Schedule	Business Plan
SYSTEM MODEL (LOGICAL)	Logical Data Model	Application Architecture	Distributed System Architecture	Human Interface Architecture	Processing Structure	Business Rule Model
TECHNOLOGY MODEL (PHYSICAL)	Physical Data Model	System Design	Technology Architecture	Presentation Architecture	Control Structure	Rule Design
DETAILED REPRESENTATION (AS-BUILT)	Data Definition	Program	Network Architecture	Security Architecture	Timing Definition	Rule Design
FUNCTIONING ENTERPRISE (FUNCTIONING)	Data	Function	Network	Organization	Schedule	Strategy
	WHAT	HOW	WHERE	WHO	WHEN	WHY

Figure 1. Zachman Framework, modified from
http://apps.adcom.uci.edu/EnterpriseArch/Zachman/Resources/ZachmanTutorial.ppt

Universal Provider Registry

A key element of any collective endeavor is the identification of the participants and their roles and contributions to a community. Closely matching Zachman's concept of "Scope," such a declaration defines the bounds of the community and facilitates the discovery and interaction of the participants. In the Web Services community, the notion of a service or business registry has become popular as an implementation of human and software brokers to find and connect to services operated by organizations.

Two industry models exist for such a service registry, both hosted within the OASIS XML consortium. The dominant 'business' and 'services' registration system is the UDDI, which stands for the Universal Description, Discovery, and Integration of Web Services[2]. It is sponsored by over 100 IT companies and culminates in a public, replicated Universal Business Registry (UBR) of organizations, services, and operations (interfaces) that are searchable and navigable. Another business registry has been developed through the electronic business XML (ebXML) group, whose product is known as the ebRIM or Electronic Business Registry Information Model. The ebRIM is typically applied within enterprise environments and allows for the registrtation and association of many types of information objects and service types. However, the ebRIM is not manifest in a public or global repository of services such as UDDI. Work continues in OASIS to define

[2] Universal Description, Discovery and Integration of Web Services, http://www.uddi.org

relationships and a possible future merger of the ebXML and UDDI activities, but in the meantime extensive support for programming tools (APIs) exists to access the UDDI, and to a lesser extent, to access ebRIM resources.

The use of UDDI would provide NSDI publishers with a place to register and access high-level provider and service information. Rather than contain metadata for individual data layers or resources in the NSDI, the public UDDI instead stores information about participating Providers, the specific services that they provide (e.g. Web Map Service, Catalog, Data download service), and connection information that can be read by humans and software, as shown in Figure 2. To paraphrase the information model of UDDI depicted in the figure, businesses (called ***providers*** in the Microsoft implementation of UDDI) may operate one or more ***services*** to their clients. These services may be human oriented services such as ordering or customer service, or may be named groups of software operations also described as a service. Each ***service*** may bundle one or more operations and provide computer-readable connection information known as ***bindings***. Where the bindings – or software connection information – are described as instances of registered ***types*** of software interfaces (e.g. OGC Web Map Service Version 1.1.1) client or browser software that recognizes such a ***type*** could easily connect to it. Providers, Services, and Bindings can also be associated with various ***categories*** to assist browse and search.

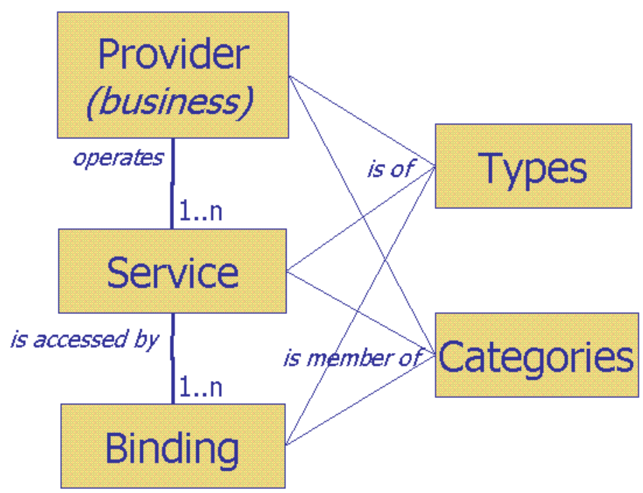


Figure 2. Basic Model of UDDI, Version 3

The use of a standards-based public registry to store organizational and service-level descriptions for all SDI geoservices (catalog/Z39.50, web directories, map and feature services, gazetteers, applications, etc.) is suggested as a practice that organizational, national, regional or international SDI gateways and applications can draw from and that publishers in all countries could publish to. Entries in a universal provider registry would be categorized to identify which networks or SDI affiliations a given provider or service is associated with. This would allow one to identify precisely what catalogs or map services were participating in a given national SDI and permit clients to discover and access services spanning national SDI boundaries.

All registered GSDI Clearinghouse Nodes (Metadata Servers) have been uploaded to the public UDDI or Universal Business Registry, operated and replicated by Microsoft, IBM, and SAP as a public information resource on the Web. The entries for each metadata server include the 'business' or organizational information and service/operations information required to connect to a given geospatial Z39.50 service anywhere in the world. Through a well-described use of UDDI, other nations and organizations would be able to use the same facility to register their services. Thus portals in other countries or portals with an interest in fully international content could apply the same techniques for a variety of applications. From the GSDI perspective, it would be highly desirable for all providers (organizations) and their services and bindings to be entered into the public UDDI/UBR following yet-to-be-agreed categorization schemes and type models for services and data that support published standards and specifications. Search for, visualization of, and provisional access to relevant geospatial data and services across international boundaries would be greatly facilitated by such an implementation agreement.

Identification of Relevant Standards

The population of a registry of national or global SDI will rely on the publication and adoption of selected standards and specifications in order to promote interoperability and ease-of-use. The identification of adopted standards is a second fundamental capability that also helps to define membership within a community. Standards and specifications are not explicitly called out in the Zachman Framework but one would expect the various models, definitions, and architectures to reference appropriate standards and specifications and how they interact.

The standardization process often specifies focused functionality that is within the scope of the standard or specification and should reference related documents or dependencies. The full scope of interaction of possible standards and specifications and the roles that they play within an SDI again will require that some type of architecture be constructed. The U.S. FGDC has published a Geospatial Interoperability Reference Model (GIRM)[3] as a living document to advise practitioners on relevant software interfaces and formats to be aware of when implementing compatible systems. The Technology Advisory Panel of the GeoConnections program in Canada has also published a document of endorsed standards and specifications employed by participants in the Canadian Geospatial Data Infrastructure[4]. The European INSPIRE initiative and the Australia Spatial Data Infrastructure are developing similar guidance to participants on what public standards and specifications are required to interact in their national and regional contexts. Fortunately, due to the common adoption of ISO standards and OGC and W3C specifications, there is already some level of service interoperability between different national SDIs.

---

[3] Geospatial Interoperability Reference Model (GIRM), Version 1.1, http://gai.fgdc.gov/girm/

[4] CGDI TAP, http://cgdi.gc.ca/CGDI.cfm/fuseaction/technology.keyDocs/pgm_id/12/gcs.cfm

Service Specifications

Over the past several years there has been progress in defining and implementing Web Services by the World Wide Web Consortium (W3C) and in publishing interface implementation specifications by the OpenGIS Consortium (OGC). Both of these consortia represent trade associations of implementing members – companies and individuals, commercial, open-source, and public sector – that offer experts to document common software interfaces. In both cases, W3C Recommendations and OGC Specifications are typically backed by evidence of implementation or are documentation of the methodology behind proven implementations. This means that working implementations are typically available for inspection or use prior to their formal adoption, shortening the time-to-deployment and improving the rigor of review and testing of the candidate specification before it is published.

Several W3C specifications and OGC Recommendations are relevant to the construction of services within a national Internet-based SDI. Selected services are described in Table 1; for more detail on selected services and their interaction consult the GSDI Cookbook[5] and nationally endorsed specifications listed above.

Table 1. List of selected adopted OGC and W3C technologies

Org	Specification name	Current Version	Description
OGC	Web Map Server (WMS)	1.1.1	Provides four protocols (GetCapabilities, GetMap, GetFeatureInfo and DescribeLayer) in support of the creation and display of registered and superimposed map-like views of information that come simultaneously from multiple sources that are both remote and heterogeneous over HTTP.
OGC	Web Coverage Server (WCS)	1.0	Extends the Web Map Server (WMS) interface to allow access to geospatial "coverages" that represent surfaces of values or properties of geographic locations, rather than WMS generated maps (pictures).
OGC	Web Feature Server (WFS)	1.0	The purpose of the Web Feature Server Interface Specification (WFS) is to describe data access operations on OpenGIS® Simple Features (feature instances) such that servers and clients can "communicate" at the feature level.
OGC	Catalog Service (CS)	1.1.1	Defines common interfaces over CORBA and Z39.50 to perform discovery, browse and query operations against distributed and potentially heterogeneous catalogs of metadata. This permits the discovery of services or information content based on field and full-text search.

[5] GSDI Cookbook, Version 1.1, http://www.gsdi.org/pubs/cookbook/cookbook0515.pdf

W3C	HyperText Transfer Protocol (HTTP)	1.1	Supports the definition of GET and POST operations for handling requests and responses between Web clients and Web servers. This standard is ubiquitous but essential to Web Services.
W3C	Simple Object Access Protocol (SOAP)	1.2	Defines the structure and exchange of XML-based messages between peers in a decentralized, distributed environment. SOAP is being used to interact with certain Web Services such as UDDI.
W3C	Web Services Description Language (WSDL)	2.0	An XML language for describing Web services. This is a structured representation of the connection information required to bind to a Web Service.

Ultimately, communities must adopt and implement standards for data layer content in concert with access or analysis services in order to readily use geospatial data in multiple applications. This may require the development of architectural plans that go beyond the scope of a given standard, or list of standards, and declare their interaction of data and services in specific ways for common public scenarios. This orchestration is a key benefit to developing an enterprise architecture within and across organizational boundaries.

Standards related to information content

Data content guidelines are being developed by professional communities worldwide that describe information structure and intended semantic content for selected data themes. The form that these data guidelines take varies with each organization but often include a data dictionary, a data model, guidance on the threshold for collection of content, and suggested structures for the encoding and exchange of the information. ISO Technical Committee 211 has several draft International Standards that are useful to consider in the development of national or community data content specifications. Through adherence to common ISO standards, the exchange and interpretation of data content by and between countries for cross-border applications is greatly facilitated. Coupled with specifications from W3C and OGC, such information content can even be automatically detected and accessed by compatible software.

A synopsis of relevant ISO TC 211 Standards that should be considered in organizing and standardizing information content is shown in Table 2.

Table 2. List of selected ISO TC 211 work items

ISO Org	Standard name	Current Version	Description
TC211	Rules for Application Schema	DIS 19109	Defines means to organize and describe a specific packaging of information to be used by an external software application. Each SDI data theme should be described following such rules.
TC211	Methodology for Feature Cataloguing	DIS 19110	Provides guidance on the descriptive elements that should be included in the construction of a data dictionary for one or more data themes. Although

			a physical (database) model of the content is not provided, data fields are given and are worth including in SDI standards.
TC211	Spatial Referencing by Geographical Identifiers	CD 19112	This standard provides an abstract information model for the storage and navigation of features with geographic Ids. This standard should be consulted in the construction of national place-name gazetteer services for the lookup of places by identifier, or named locations by coordinate.
TC211	Metadata	IS 19115	Defines the schema required for describing geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.
TC211	Web Map Server Interface	CD 19128	ISO harmonized publication of OGC WMS Specification (see Table 1)
TC211	Geography Markup Language	DIS 19136	ISO harmonized publication of OGC Geography Markup Language (GML 3.0) specification. GML is an XML encoding in compliance with ISO 19118 for the transport and storage of geographic information modelled according to the ISO 19100 series and includes spatial and non-spatial properties of geographic features.
TC211	Metadata Implementation Specification	DTS 19139	This draft technical specification supplements ISO 19115 by expressing ISO metadata using XML as its encoding mechanism.

While ISO provides a context and structured methodology for describing specific information content, e.g. data themes, it is generally accepted that national and international communities will create profiles of standards that describe and express the information content of a theme in conformance with these ISO standards or their national profiles. Public definitions of thematic data content are a common feature of national SDIs and of international efforts such as Global Map, coordinated by the International Steering Committee for Global Map.

Semantic Registries

Communities, such as countries, professional and linguistic groups are often defined by the common set of adopted vocabulary that they use in a consistent way. This will be true within different areas of geographic applications, but also in the scooping of a local or national SDI. With the ability for any organization, country, or region to define their own data content standards independently – even when using ISO standards and OGC specifications – the actual content and structure of data for a given theme will vary. This requires a mechanism to facilitate the interpretation or conversion of data to match a common or foreign model. Models for a single theme at the local, regional, corporate, national, and multi-national level may include different definitions of the features being mapped, the relationship between the features, they types of geometry being used, and the meanings of attributes and their values. The publication of these models improves the understanding and adoption of the information designs.

A semantic registry is a structured resource much like a data dictionary in which the meaning and syntax of all pieces of information used within a broad enterprise can be stored. Two notable semantic registries have been established in a global context for specific domains of application. The Basic Semantic Register (BSR) is a work item of ISO TC 154 that includes semantic descriptions of resources commonly used in Electronic Data Interchange (EDI, UN EDIFACT)[6]. The goal is to provide a navigation capability for developers and users of software systems to query, adopt, and apply common fields, concepts, and schemas. (usbr.org). The Environmental Data Registry (EDR)[7] operated by the U.S. Environmental Protection Agency on behalf of the broader environmental reporting community is a comprehensive, authoritative reference for information about the definition, source, and uses of environmental data. The EDR promotes the efficient sharing of environmental information among EPA, states, tribes, and other information trading partners by defining the structure and meaning of information elements used in data exchange. Neither the BSR nor the EDR contain real 'data;' they contain the descriptive information of the concepts and their instantiation that can be navigated in ways that go well beyond traditional data set metadata.

Unless a common data model and encoding for all geospatial data is adopted for every theme of data around the world, the syntactic and semantic interoperability of similar data sets will remain a challenge in the absence of an operational semantic registry at the national and/or global level for geospatial information. Using such a system, individual organizations could publish digital geospatial information, the data model or schema used to interpret its syntax (structure), and the semantic model used to describe its explicit content. Wherever possible, the elements of the semantic model would be associated with equivalent semantic elements in national or professional information schemas that would help in the automated interpretation or transformation of both the data structure and its semantic content. The application of semantic registries and their automated processing is being pursued by the W3C in the development of a "Web Ontology Language (OWL)."[8] OWL is a semantic markup language for publishing and sharing ontologies on the World Wide Web that builds upon a variant of XML that expressly describes relationships between objects, known as the Resource Description Framework (RDF).

Until semantic registries are in place, and the World Wide Web widely implements the infrastructure supportive of a "Semantic Web" to help understand the meaning of information, the use of standard metadata to describe data and services suggested practice to publish as much semantic information as possible. This will allow programmers, providers, and end-users of data to read about the information content – hopefully down

---

[6] Basic Semantic Register (BSR), ISO/TC 154 "*Processes, data elements and documents in commerce, industry and administration*" http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/TC154WG1/PUBLIC/WEB/ENGLISH/content.htm

[7] Environmental Data Registry (EDR), http://www.epa.gov/edr/

[8] Web Ontology Language Reference, W3C Proposed Recommendation, http://www.w3.org/TR/owl-ref/

to the feature or data element level – and do intelligent interpretation of the data or permit its manual integration with other data.

Access through Gateways and Desktop Applications

Communities are also defined by their gathering places – the locations that individuals and groups can congregate to exchange information. In the World Wide Web, the notion of a 'portal' has been promoted as a unifying interface that provides access to linked but diffuse community resources. Typical community portals include facilities for individuals and members to join or subscribe, to identify news and events through a calendar facility, and to exchange messages on lists or in chat rooms, and, most importantly, to provide access to information resources of specific interest to the community.

The metaphor of a community Web portal is perfectly suited to the interests and dynamics of local, national, or professional SDI groups as a locus for information exchange. Various portals or information resource gateways have been established with one or more of the aforementioned capabilities for the geospatial communities. In the remote sensing community, gateways that permit browse and search of deep inventories of satellite imagery have been created allowing an analyst to search for data across many collections in many countries[9]. In commercial product communities, data and service discovery portals like the "Geography Network"[10] have been designed to broker easy access to spatial data for GIS software. In many countries, spatial data catalogs and portals have been established to search distributed geospatial data through metadata descriptions. Links to request, download, display, or purchase these data are included in the metadata to apply the data more rapidly to a problem, though the issues of data format, availability, and semantics as described in prior sections of this paper remain as obstacles to 'ease of use.'

In order to access a set of resources across the globe that collectively instantiate a virtual GSDI, a prototype data gateway was set up in 2000 by the GSDI Secretariat based on a search gateway developed in the US NSDI[11]. This gateway provides access to search a distributed set of over 250 collections of metadata in over 40 countries using a common search protocol (ISO 23950:1995, ANSI Z39.50) but delivering metadata in several different metadata standard formats (ISO 19115, FGDC Content Standard for Digital Geospatial Metadata, and ANZLIC). The gateway is very "data centric" in that the primary resource being catalogued is a geospatial data set, layer, or theme. Associated with its description are links to the one or many ways by which it can be accessed. In other words, its access methods are characteristics of the data set itself. Where map services are identified in the metadata, they show up as links that launch a new web browser window and immediately display the data in map form.

A new search gateway or portal was commissioned by the U.S. government in 2003 to help organize diverse geospatial data and services exposed using ISO standards and OGC

[9] Earth Observing System Data Gateway, http://edcimswww.cr.usgs.gov/pub/imswelcome/
[10] Geography Network, http://www.geographynetwork.com
[11] SDI Search Gateway, http://clearinghouse1.fgdc.gov/servlet/FGDCServlet

specifications[12]. An instance of this portal was recently contributed to the GSDI Association by the U.S. government to provide a more robust set of services to register, search, display, and analyze spatial data exercising many standards-based interfaces. This new gateway to GSDI resources, both online web services and data, provides an opportunity for individual countries to register and expose their metadata catalogs, their data order or access systems, and their online mapping capabilities using emerging international standards. While not replacing the community portals, gateways, and websites operated by agencies, professional associations, universities, or companies, the new GSDI Association Gateway allows us to begin to visualize a virtual global network of compatible geographic information services and work towards their integration over the Web.

Most portals and gateways to geospatial data are intentionally general-purpose, providing search and evaluation support for data for a broad community of users but not necessarily fulfilling end-user analytical requirements. To solve real-world analytical problems, geospatial data must be linked to geoprocessing functions as would be found in desktop or organizational GIS software. For maximum benefit to the end-users, an SDI must support these desktop GIS users directly such that they can find and bind to geospatial data from within their applications without need to visit a portal using a Web browser.

The functions and resources behind an SDI portal or gateway – the catalogs of data, services, schemas, semantics resources – need to be accessible to clients using web browsers *and* to clients using desktop GIS. It may not be conventional for a portal to expose interfaces that can be accessed by software other than web browsers, but the exposure of these services to applications has wide applicability. The Discovery Portal operated by GeoConnections Canada has seen a rapid increase in the number of clients using portal client components through published Application Programming Interfaces (API) despite a constant or declining number of traditional Web browser clients[13]. In the U.S., a government-wide procurement document is being developed to simplify the purchase of geospatial software components with standards-based interfaces that could be used in construction of agency and community portals and geoprocessing services[14]. With such a procurement vehicle, organizations including other federal agencies, state and local governments could more easily adapt their existing websites and applications to integrate with the SDI resources listed above.

Summary

The development of SDIs and the GSDI requires the consideration of many architectural components that go well beyond the foundational data catalogs and metadata. An SDI is a type of community that can be defined by its members (providers, experts, and practitioners), their capabilities (services, information resources), a shared vocabulary (semantics), shared conventions (endorsed standards and specifications), and the communal facilities that promote interaction (portals, catalogues). The development of an

---

[12] OGC Geospatial One-Stop Portal Prototype, http://gospi.saic.com/gospi/portal/
[13] Discovery Portal Web API, http://geodiscover.cgdi.ca/about/en/6.4.html
[14] Contract for Interoperable Geospatial Components, http://www.fgdc.gov/geoportal/

Enterprise Architecture to encompass the organizational processes (functions and data in support of mission requirements) is a useful methodology to build rationale and opportunities for interaction between organizations to apply geospatial data and services. An operational portal for the GSDI provides a composite of national and regional SDI services and data that are one expression of the virtual GSDI.