Handbook of Research on Innovations in Database Technologies and Applications:
Current and Future Trends

Viviana E. Ferraggine
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Jorge H. Doorn
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Laura C. Rivero
Universidad Nacional del Centro de la Provincia de Buenos Aires, Argentina

Volume II

INFORMATION SCIENCE REFERENCE
Hershey · New York

Director of Editorial Content: Kristin Klinger
Managing Editor: Jamie Snavely
Assistant Managing Editor: Carole Coulson
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com/reference

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanbookstore.com

Library of Congress Cataloging-in-Publication Data

Handbook of research on innovations in database technologies and applications: current and future trends / Viviana E. Ferraggine, Jorge H. Doorn, and Laura C. Rivero, editors.

p. cm.

Includes bibliographical references and indexes.

Summary: "This book provides a wide compendium of references to topics in the field of the databases systems and applications"--Provided by publisher.

ISBN 978-1-60566-242-8 (hardcover) -- ISBN 978-1-60566-243-5 (ebook) 1. Database management. 2. Database design--Economic aspects. 3. Technological innovations. I. Ferraggine, Viviana E., 1961- II. Doorn, Jorge H., 1946- III. Rivero, Laura C., 1956-

QA76.9.D3H347326 2009

005.74--dc22

2008050170

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter XXXVI
Spatial Data Integration Over the Web

Laura Diaz
Universitat Jaume I, Spain

Carlos Granell
Universitat Jaume I, Spain

Michael Gould
Universitat Jaume I, Spain

INTRODUCTION

Spatial data are increasingly becoming available on the Internet in applications such as routing portals that involve map-based and satellite imagery backgrounds, allowing a large audience to access and share the rich databases that are currently used by the specialized geographic community. These spatial data are heterogeneous, being available in various formats, and stored in disparate formats (flat files, relational or object-oriented databases, etc.). Some data are structured according to well-established data modeling techniques such as the relational or object-oriented data models; other data, such as data maintained in various information systems, spreadsheets, or Internet repositories, are in proprietary formats, semi-structured, or unstructured. In practice, this situation of multiple models and schemas combined with the difficulty for establishment agreements for data representation in the application domains becomes spatial data in special regarding other types of scientific data, making the interoperability problem a nontrivial task (Lemmens, Wytzisk, de By, Granell, Gould & van Oosterom, 2006). In addition to the scale of data integration, the complex and heterogeneous query processing and domain-specific computational capabilities supported by these sources make spatial data integration a real challenge (Boulcema, Essid & Lacroix, 2002; Devogele, Parent & Spaccapietra, 1998; Goodchild, Egenhofer, Fegeas & Kottman, 1999).

Historically, due to specialized characteristics and the nature of spatial data, geographic information systems (GISs) were managed separately from existing database systems. As first steps to spatial data integration in the mid-1990s, advances in database technology enabled accommodating spatial data in relational databases, allowing organizations to take the first steps toward enterprise GIS and the elimination of organizational "spatial data islands" (ESRI, 2003). Some examples are the appearance of Oracle Spatial (www.oracle.com), PostgreSQL with the PostGIS extension (www.postgresql.org), and MySQL (www.mysql.com). The early work for spatial data integration in database systems focused on sharing simple spatial features in a relational database. Then, standard data manipulation languages such as SQL (Structured Query Language) began to adopt common spatial functionalities to embed, for example, spatial selections and topological queries in SQL statements. The arrival of the first relational models capable of storing both spatial and attribute data led to spatial databases (Rigaux, Scholl & Voisard, 2001), which provided methods for spatial data modeling, algorithms, access methods, and query processing extending traditional database systems.

The success factor of Web services technology has permitted promoting service integration and interoperability among heterogeneous distributed information sources. The GIS approach to service-oriented architecture (SOA) is represented by the Spatial Data Infrastructure (SDI) paradigm, which offers the possibility to access distributed, heterogeneous spatial data through a set of policies, common rules, and standards that facilitate interconnecting spatial information users in an interoperable way (Granell, Gould, Manso & Bernabé, 2007b).

Several approaches and techniques exist for data integration over the Web (Fileto, 2001); however, some of the most representative are the following: *Gateways* as middleware that allows an application running in one DBMS (Data Base Management System) to access data maintained by another DBMS; and *Data Warehouses*, which are separated databases built specifically for decision support, useful when decisions depend on heavy analysis of large amounts of data collected from a variety of possibly heterogeneous data sources. In the GIS domain, most data integration approaches have included one or both of the more common techniques: *wrapper* (Roth & Schwarz, 1997*)* and *mediator* (Wiederhold, 1992). Wrappers provide interfaces for accessing concrete data sources, retrieving the results, and translating them into a common scheme. Mediators are software components in the middleware in charge of specifying such a common scheme that provides an integrated interface to a set of wrappers; indeed, underlying data sources. Data integration approaches use mediators to handle client queries, submit them directly to the wrappers, and integrate the results before delivering the response to client applications. GIS systems manage data integration mostly by using wrappers and mediators implementing standard interfaces specified by Open Geospatial Consortium (OGC), which provides open standards and specifications for supporting standardized information exchange, sharing, accessing, and processing geospatial data.

The demand for interoperability has boosted the development of standards and tools to facilitate data transformation and integration. Furthermore, this chapter focuses on interface standards as key to spatial data syntactical integration over the Web. Nevertheless, there are still many challenges to be met, especially those concerned with data semantics and harmonization of interoperating systems.

BACKGROUND: SPATIAL DATA INFRASTRUCTURE

Many standards have been proposed for exchanging geographical data among systems (Albrecht, 1999), although they are not enough to enable interoperability because data conversion among these formats often produces information loss. The adoption of a common geographical data model or at least a framework to unify heterogeneous models constitutes one ingredient to achieve GIS interoperability in the wide sense. A first attempt in this way has been featured by the GML standard offering access and retrieval of spatial data in a standard exchange format.

Geography Markup Language (GML) (Cox, Daisay, Lake, Portele & Whiteside, 2002) describes an encoding specification for spatial data in XML that enables the storage, transport, exchange, processing, and transformation of geographic information. GML provides a variety of object types for describing geography, including features, coordinate reference systems, geometry, topology, time, units of measure, and generalized values. Current database implementations (with spatial extensions) permit storing directly GML-based data, thus enabling interacting with other spatial and nonspatial data sources using an international standard. However, implementers also may decide to convert from some other storage format on demand and use GML only for schema and data transport. In this case, spatial data integration over the Web implies geospatial services, which normally exchange GML data, following the SDI architecture.

Spatial Data Infrastructure (SDI) comprises a set of policy and standards activities promoting creation of a geospatial information infrastructure to assist diverse user communities to collect, share, access, and exploit georeferenced information resources. As depicted in Figure 1, traditional SDI vision focused on data is shifting to a service-based vision (Bernard, Craglia, Gould & Kuhn, 2005) in which geospatial services in the middleware are increasingly used to discover and access geospatial data stored on data repositories, transform it into useful information to users, and then deliver it to them.

Although user applications can access directly to spatial data sources as in traditional database systems (see rightmost arrow in Figure 1), they normally use a set of standard geospatial interface services to discover and access spatial data sources in SDI in order to achieve the spatial data integration. Such geospatial service interfaces, mainly published by OGC, act typically as wrappers to add the level of abstraction needed for the syntactic integration of spatial data. The next section introduces some interesting examples of OGC specifications for data integration in the SDI context.

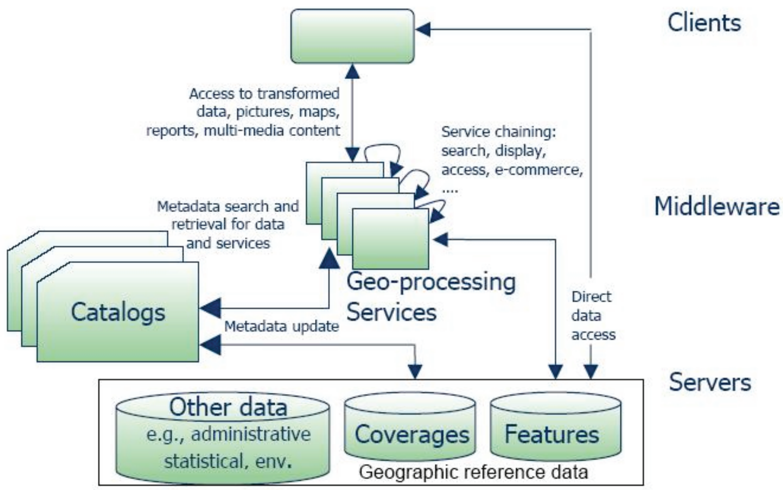
SPATIAL DATA INTEGRATION

Integrating spatial data over the Web implies several components and services linked via The Internet following the SDI architecture. First, this section describes some key service interfaces to spatial data integration. Next, Figure 2 summarizes how such service interfaces are put together to offer spatial access, edition, and processing over heterogeneous and remote data sources.

GEOSPATIAL SERVICE INTERFACES

The OGC Simple Features Access (SFA) is a standard that consists of two parts. The first one describes the common architecture for simple feature geometry (http://www.opengeospatial.org/standards/sfa). A feature has both spatial (geometry valued) and nonspatial attributes. This first part then defines the simple feature geometry object model that is distributed computing platform neutral, which means a common, platform-independent interface to handle geometry

Figure 1. High-level SDI architecture, taken from Smits (2002)



objects in a standard format. All geometry objects, such as *Point*, *Curve*, *Surface*, and *Geometry Collection,* extend the basic functionality of the *Geometry* base class. Each geometric object is also associated with a concrete spatial reference system, which specifies the coordinate space in which the geometric object is present. The second part of the OGC SFA standard (http://www.opengeospatial.org/standards/sfs) is to define standard SQL schema that supports storage, retrieval, query, and update of a geometry object or feature collection, terms already defined in part 1 of this standard.

The OGC Web Feature Service (WFS) (Vretanos, 2005) defines interfaces for data access and manipulation operations on geographic features using HTTP as transport protocol. Via these interfaces, a Web user or service can combine, use, and manage spatial data—the feature information behind a map image—from various data sources by invoking the following WFS operations on geographic features and elements:

- Create a new feature instance
- Delete a feature instance
- Update a feature instance
- Lock a feature instance
- Get or query features based on spatial and nonspatial constraints

The OGC Filter Encoding Implementation Specification (Filter) (http://www.opengeospatial.org/standards/filter) defines an XML encoding for filter expressions to be used in conjunction with other specifications, as the case of WFS services. A filter expression constrains property values to create a subset of a group of objects (features). The goal is to select a determined subset of objects and operate on just such subset by, for example, rendering them in a different color to save them to another format or edit them. An XML encoded filter behaves normally as an SQL *where* structure, because any set of filters could be transformed into an SQL *where* clause for a SQL *select* statement to fetch spatial and nonspatial data stored in any database.

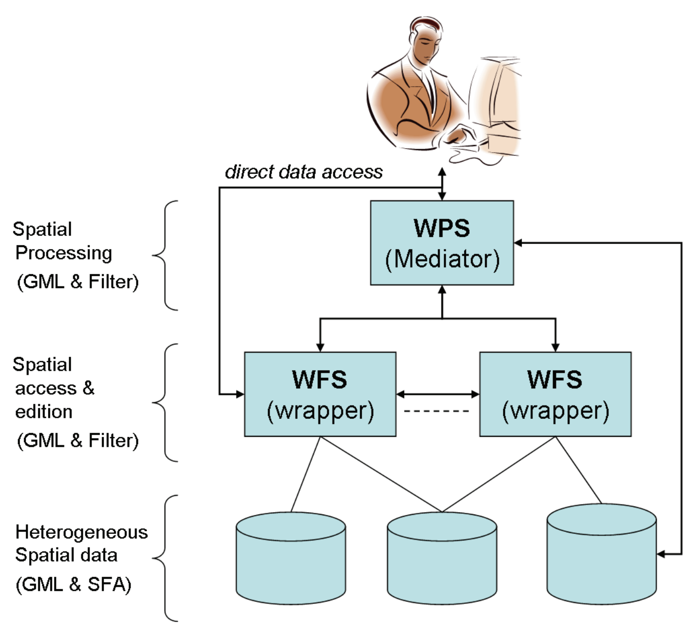
The OGC Web Processing Service (WPS) (Schut, 2007) provides access to calculations or models that operate on spatially referenced data. The data required by the service can be available locally or delivered across a network using data exchange standards such as GML or Geolinked Data Access Service (GDAS). The calculation can be as simple as subtracting one set of spatially referenced numbers from another (e.g., determining the difference in influenza cases between two seasons) or as complicated as a global climate change model. While most OGC specifications and standards are devoted for spatial data abstraction, access, and integration, the WPS specification is about spatial data processing over heterogeneous data sources. The main steps in this process are to identify the spatially referenced data required by the calculation, initiate the calculation, and manage the output from the calculation so it can be accessed by the client. The WPS is targeted at both vector and raster databased processing.

Spatial Access and Edition

As depicted in Figure 2, wrapping spatial databases with OGC WFS Services means that we provide one more level of abstraction to data access and retrieval, since knowing the database-specific model and settings are no longer needed. Indeed, the WFS layer acts as a wrapper to manipulate the underlying data sources and/or spatial databases; it receives requests from remote users, application, or mediators; executes them against the corresponding data sources; and delivers the results to remote users, applications, or mediators.

In most cases, the exchange format of a WFS server is GML. WFS has a rich query interface by using the OGC Filter encoding specification, which describes an XML encoding of the OGC Common Catalogue Query Language (CQL) as a system-neutral representation of a query predicate. As WFS services return GML data, such XML-based representations facilitate data edition by using the numerous XML tools available today.

Figure 2. Components for spatial data integration over the Web



GML data can be easily validated, parsed, edited, and transformed into whatever target language or persistent object.

A WFS service supports *insert*, *update*, *delete*, *query*, and *discovery* operations that are applied to one or more geographic features. A WFS service delivers GML representations of geospatial features in response to queries from HTTP clients. Clients access geographic feature data through WFS by submitting a request for just those features needed for an application.

Spatial Processing

More complex geospatial services have to be specified in order to distribute over the Internet all the functionalities (computation, analysis, etc.) common in our desktop GIS and local data. The first steps toward advanced geoprocessing services online are outlined by the recently published OGC Web Processing Service (WPS) discussion paper (Schut, 2007), which provides interface specifications to enable geospatial Web services to support a limited range of geoprocessing tasks by creating accessible libraries of geoprocessing algorithms under the appearance of geospatial Web service chains (Granell, Gould *Esbrí, 2007a).

WPS services allow users not only to access and visualize distributed data through services but also to realize complex spatial operations (e.g., spatial analysis, shortest path, buffer, etc.). In addition, Figure 2 shows how WPS services may play the role of mediators that provide an integrated interface to various WFS services (wrappers). Client queries containing GML data and optionally filter expressions are handled by WPS services, which submit them directly to the WFS services. WPS services are also in charge of integrating the results (GML) before delivering the response to client applications to perform complex spatial processing. Users or applications can access WFS directly via WFS clients to access and edit the underlying data sources and perform the operations given by the WFS interface.

CHALLENGES AND FUTURE TRENDS

Table 1 lists some challenges that are being solved or will be solved in the future in order to improve spatial data integration over the Web. One of the key areas is that surrounding semantics and spatial data harmonization.

In the semantic context, new approaches and prototypes are emerging (Lemmens et al., 2006; Lutz, 2007; Tanasescu et al., 2007; Tomai & Prastacos, 2006) to improve semantic aspects of spatial data integration by mediating legacy spatial data sources to high-level spatial ontologies through Semantic Web Services (McIlraith, Son & Zeng, 2001). The semantic integration of heterogeneous data requires specific domain knowledge exposed as ontologies that rely on shared concepts, terms, and structuring constructs found in source data. They involve metadata enrichment to support semantic matching among data items from distinct datasets (Stoimenov & Djordjevic-Kajan, 2006). Future developments on spatial data integration with Semantic Web Services will include an increase in the complexity of the integration ontologies in order to allow the unified representation, access, and edition of heterogeneous spatial data sources as envisioning through the *Geospatial* Semantic Web vision (Egenhofer, 2002). This vision requires the development of multiple spatial and terminological ontologies, each with a formal semantics. This will lead to a new framework for geospatial information based on the semantics of spatial and terminological ontologies in which geospatial queries and results are processed semantically. The goal of the Geospatial Semantic Web is to implement components for discovery, querying, and consumption of geospatial content in a distributed architecture, which are based on formal semantic specifications.

Interesting examples of spatial data harmonization are the ongoing projects and current best practice recommendation efforts (e.g., by the INSPIRE Drafting Team on Data Specifica-

Table 1. Challenges and future trends in spatial data integration

- Semantics and data harmonization
- Security: access control, authentication
- Transactional integrity

tions) devoted to the GMES-INSPIRE vision of seamless interoperability. The GMES (Global Monitoring for Environment and Security) initiative focuses on six application fields: land cover, water resources, ocean/marine applications, atmosphere, risk management, and security. The INSPIRE (Infrastructure for Spatial Information in Europe) infrastructure will then facilitate implementation of services to achieve GMES goals of harmonizing space-based, in situ, and traditional cartographic data sources. In consequence, the GMES program recently funded data harmonization projects such as the RISE project (www.eurogeographics.org/eng/03_RISE.asp) that aim at developing tools and methodologies to enable and facilitate cross-disciplinary interoperability. From this viewpoint, data harmonization pursues the goal of producing geospatial data implementation specifications consistent with international standards encompassing the development of application schema and data product specifications to enable sustainable and interoperable functioning of INSPIRE and GMES.

Other important issues listed in Table 1 are related to security and transactional aspects. Security is used here in the sense of ensuring that communications and transactions among services, both those acting as mediators and as wrappers, are conducted in a protected environment and that messages are reliably delivered to the correct destinations. Other challenges to achieve full interoperability of GIS and related spatial data integration are more related to social and policy restrictions; when accessing distributed data, security and integrity issues will normally arise.

CONCLUSION

GML permits representation of spatial and non-spatial data with an XML-based format, providing spatial services with common data models for spatial data access and interchange. An environment to support data integration must be based on this common data model, which may also be used for data transfer and exchange.

SDIs provide the infrastructure in which spatial wrappers and mediators play a facilitating role. WFS services wrap the data sources, abstracting data from its machine representation, and become accessible to diverse users in a uniform way. As far as spatial processing, WPS services provide an interface to allow not only data access but also data processing and analysis in an interoperable fashion. Adoption of OGC interfaces and standards makes possible spatial data integration in a distributed environment when semantic differences are not too great. Otherwise, future research on semantic interoperability is needed to reach generally acceptable levels of ad hoc spatial data integration.

REFERENCES

Albrecht, J. (1999). Geospatial information standards: A comparative study of approaches in the standardization of geospatial information. *Computers & Geosciences, 25*(1), 9–24.

Bernard, L., Craglia, M., Gould, M., & Kuhn, W. (2005). *Towards an SDI research agenda*. Proceedings of the EC & GIS Workshop, Sardinia.

Boulcema, O., Essid, E., & Lacroix, Z. (2002). *A WFS-based mediation system for GIS interoperability*. Proceedings of the 10th ACM International

Symposium on Advances in Geographic Information Systems, 23–28.

Cox, S., Daisay, P., Lake, R., Portele, C., & Whiteside, A. (Eds.). (2002). OpenGIS geography markup language (GML), Version 3.0. Open Geospatial Consortium. Retrieved from http://www.opengeospatial.org/standards/gml

Devogele, T., Parent, C., & Spaccapietra, S. (1998). On spatial database integration. *Intl. Journal of Geographical Information Science, 12*(4), 335–352.

Egenhofer, M. (2002). *Toward the semantic geospatial Web*. Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems, 1–4.

ESRI. (2003). *Spatial data standards and GIS interoperability*. Retrieved from www.isotc211.org/Outreach/Newsletter/Newsletter_04_2004/Appendix_4.pdf

Fileto, R. (2001). *Issues on interoperability and integration of heterogeneous geographical data*. Proceedings of the 3rd Brazilian Symposium on GeoInformatics.

Goodchild, M., Egenhofer, M., Fegeas, R., & Kottman, C. (1999). Interoperating geographic information systems. Norwell: Kluwer Academic Publishers.

Granell, C., Gould, M., & Esbrí, M.A. (2007a). Geospatial Web service chaining. In H. Karimi (Ed.), *Handbook of research on geoinformatics*. Hershey: Information Science Reference.

Granell, C., Gould, M., Manso, M.A., & Bernabé, M.A. (2007b). Spatial data infrastructures. In H. Karimi (Ed.), *Handbook of research on geoinformatics*. Hershey: Information Science Reference.

Lemmens, L, Wytzisk, A., de By, R., Granell, C., Gould, M., & van Oosterom, P. (2006). Integrating semantic and syntactic descriptions to chain geographic services. *IEEE Internet Computing, 10*(5), 42–52.

Lutz, M. (2007). Ontology-based descriptions for semantic discovery and composition of geoprocessing services. *Geoinformatica, 11*(1), 1–36.

McIlraith, S.A., Son, T.C., & Zeng, H. (2001). Semantic Web services. *IEEE Intelligent Systems, 16*(2), 46–53.

Rigaux, P., Scholl, M., & Voisard, A. (2001). *Spatial databases—With applications to GIS*. San Francisco: Morgan Kaufmann.

Roth, M.T., & Schwarz, P. (1997). *Don't scrap it, wrap it! A wrapper architecture for legacy data sources*. Proceedings of the International Conference on Very Large DataBases, 266–275.

Schut, P. (Ed.). (2007). OpenGIS Web processing service specification, version 1.0.0. Open Geospatial Consortium. Retrieved from http://portal.opengeospatial.org/files/?artifact_id=21988&version=1

Smits, P. (Ed.). (2002). INSPIRE architecture and standards position paper. Architecture and Standards Working Group. Retrieved from http://inspire.jrc.it/documents/inspire_ast_pp_v4_3_en.pdf

Stoimenov, L., & Djordjevic-Kajan, S. (2006). *Discovering mappings between ontologies in semantic integration process*. Proceedings of the AGILE Conference on Geographic Information Science, Visegrad, Hungary, 213–219.

Tanasescu, V., et al. (2007). Geospatial data integration with semantic Web services: The eMerges approach. In A. Scharl, & K. Tochtermann (Eds.), *The geospatial Web*. **London:** Springer.

Tomai, E., & Prastacos, P. (2006). *A framework for intensional and extensional integration of geographic ontologies*. Proceedings of the AGILE Conference on Geographic Information Science, Visegrad, Hungary, 220–227.

Vretanos, P.A. (2005). Web feature service implementation specification, Version 1.1.0. Open Geospatial Consortium. Retrieved from http://www.opengeospatial.org/standards/wfs

Wiederhold, G. (1992). Mediators in the architecture of future information systems. *IEEE Computer, 25*(3), 38–49.

KEY TERMS

Feature: The fundamental unit of geospatial information. For example, depending on the application, a feature could be any part of the landscape, whether natural (such as a stream or ridge) or artificial (such as a road or power line). A feature object then corresponds to a real-world or abstract entity. Attributes of this feature object describe measurable or describable phenomena about this entity. Feature object instances derive their semantics and valid use or analysis from the corresponding real-world entities' meaning.

GML: Geography Markup Language is an XML grammar defined by OGC to express geographical features. To help users and developers structure and facilitate the creation of GML-based application, GML provides *GML profiles* that are XML schemas that extend the very GML specification in a modular fashion. A GML profile is a GML subset for a concrete context or application but without the need for the full GML grammar, thus simplifying the adoption of GML and facilitating its rapid usage. Some common examples of GML profiles that have been published are *Point Profile* for applications with point geometric data and *GML Simple Features profile*, supporting vector feature requests and responses as the case of WFS.

Mediator: A negotiator who acts as a link between parties, the neutral who carries out the dispute resolution process called mediation.

OGC: Open Geospatial Consortium (http://www.opengeospatial.org), a membership body of 300-plus organizations from the commercial, government, and academic sectors that creates consensus interface specifications in an effort to maximize interoperability among software detailing with geographic data.

SDI: Spatial Data Infrastructure. Many government administrations have initiated coordinated actions to facilitate the discovery and sharing of spatial data, creating the institutional basis for SDI creation. The Global Spatial Data Infrastructure (GSDI) association (http://www.gsdi.org) defines SDI as a coordinated series of agreements on technology standards, institutional arrangements, and policies that enable the discovery and facilitate the availability of and access to spatial data. The SDI, once agreed upon and implemented, serves to connect Geographic Information Systems (GIS) and other spatial data users to a myriad of spatial data sources, the majority of which are held by public sector agencies.

Service: Functionality provided by a service provider through interfaces (paraphrased from ISO 19119—Geographic Information Services).

WFS: The OpenGIS Web Feature Service specification allows a client to retrieve and update geospatial data encoded in Geography Markup Language (GML) from multiple WFS. The specification defines interfaces for data access and manipulation operations on geographic features using HTTP as the distributed computing platform. Via these interfaces, a Web user or service can combine, use, and manage geodata, the feature information behind a map image, from various sources.

Wrapper: A package that changes the interface to an existing package without substantially increasing its functionality.