The Spatial Web

An Open GIS Consortium (OGC) White Paper

1 Introduction - Establishing the “Spatial Web”

Speed-of-light communication and faster-than-a-horse transportation are fairly recent capabilities that have allowed us to overcome many geographic barriers. But digital technologies that help us think about an expanded spatial world, such as GIS, remote sensing, GPS and location-based services, are much more recent than those technologies which expanded our ability to act in the world, and they are used little. On the whole, our accommodation to our modern, expanded realm of action is limited by a spatial awareness that hasn’t advanced much beyond the range of our seeing and hearing and our memories of places we have been.

Digital spatial[1] data has historically been isolated in “stovepipes” or “islands of automation,” resulting in expensive duplication of data production and difficulty in sharing information. It has been difficult to 1) integrate spatial information from different geoprocessing systems and 2) integrate spatial information into non-spatial information systems because:

1. Different types of geoprocessing systems (vector GIS, raster GIS, imaging systems, CAD and facilities management systems, transportation systems and navigation systems) produce very different types of data.
2. Different vendors’ geoprocessing systems use internal data formats and produce data in formats that are different and in most cases proprietary.
3. Different vendors’ systems use proprietary software libraries with proprietary interfaces that restrict opportunities for network inter-process communication between systems.
4. Different data producers, even if using the same type of system from the same vendor, do not, without coordination, name spatial features in the same way.
5. Different data producers, if they provide metadata to help data seekers evaluate the producers’ data, do not, without coordination, structure their metadata in standard schemas that enable effective automated metadata searches.

1, 2, and 3 above are the causes of technical non-interoperability. 4 and 5 are the causes of semantic non-interoperability.

The Web offers an unprecedented opportunity to overcome technical non-interoperability because it is an almost universal platform for distributed computing, with a Web services architecture that is designed for integration of diverse information systems. The Web offers an unprecedented opportunity to overcome semantic non-interoperability because it provides unique facilities for

---

[1] We could use the term "geographic" data, but not all geospatial data is graphic, thus our use of the words "spatial," "geospatial" and "geodata."

semantic processing of structured text. By enabling interoperability, the Web greatly increases access to spatial data and processing resources, and thus greatly increases the value of those resources. This will have wide-ranging consequences.

Today a manager working to reduce disaster risk needs hours or days to bring together data from many sources describing floodplain, roads, earthquake data, dwellings, sensors and sensor networks etc. before running what-if scenarios. As the Web begins to present such information in real time, the decision process will proceed much more rapidly and fluidly. As laborious spatial data production and data logistics cease to dominate the work time of people who need spatial information, those people will begin to engage in more creative, intuitive thinking about real world phenomena. People without spatial data expertise will have equal access, greatly expanding the value of spatial data and spatial processing resources.

Spatial content, services, and applications are inherently different from – and more than – video, sound, text, and image content and applications. In fact, much if not most video, sound, text, and image data refers in some way to a location. The spatial component often matters. Location information integrated into our digital information environment adds a previously missing conceptual dimension to our formulation of problems and questions. This added dimension will benefit people in research, urban planning, environmental management, transportation planning and every other geoprocessing application domain, as well as consumers and citizens. "Spatial" becomes an essential element of workflow. It will touch most Web-mediated human activities.

We believe it is extraordinarily important to weave into the Web a robust and integral spatial/temporal strand, which we call the "Spatial Web." The Spatial Web is in its infancy. In this paper we review policy issues, impacts on institutions and programs, and technical issues that will influence its development in critical ways.

2 Spatial Web Policy Issues

1. **Participation in and cooperation among standards groups:** The Spatial Web grows through the diverse commercial and non-commercial activities of countless people who use the Web to publish, find, and process information about places, objects, people, phenomena and events in "Earth space." Governments, industries, corporations, NGOs, and academia all have a stake in giving the Spatial Web a sound foundation. Only by learning about and participating in standards activities relevant to their work can they ensure that specifications for interfaces, encodings, schemas etc. will support their goals and be "right" for their applications. Data coordination groups, OGC, ISO TC/211, FGDC and organizations that specify standards for the larger Web and Internet environment all welcome the interest of their stakeholders. Standards organizations are increasingly open to "harmonization" with other standards groups to help them further their common goals. Through this global network of consensus-based standards processes, we can work to ensure that standards advancements will support and guide a constructive unfolding of the Spatial Web.

2. **Openness of the Spatial Web**. We must ensure that the Spatial Web, like the Internet itself, remains a community resource that is not controlled or excessively characterized by any commercial interest or national or regional government. The Spatial web must be built and maintained as one of humanity's critical infrastructure elements and cultural resources, an international public resource that people can use for activities of all kinds: personal, commercial, cultural, social and political.

3. **Flexibility and adaptability of the Spatial Web**: The Internet was originally designed as a peer-to-peer network where anyone can connect to anyone, a main reason for its success. HTTP and the hypertext concept is dependent on a peer-to-peer model, where anything may link to anything. This creates a democratic web, where there is no single point of control, no middle man in control of the network. Consensus standards processes can ensure both lively commerce and protection of the public interest. Some spatial content, services and applications will be for sale, sometimes on a transaction basis, while others will be free. Some will be delivered via centrally managed systems, proprietary client-server application architectures, and distributed services, and others via decentralized peer-to-peer architectures. Some will be protected for use

by emergency planners, health officials, law enforcement and other organizations in the public interest. This range of possibilities – enabled by an expanding consensus standards infrastructure – is in the spirit of the Web and the National Spatial Data Infrastructure, too.

4. **Certification of content**. Big organizations and ordinary individuals, as providers or users, may want to certify and be able to depend on the certification of spatial content. That is, how can one guarantee or be sure of the origins and lineage of data? Has it been modified in some way? Does it accurately reflect what is "one the ground"? Is metadata (data about the data) objective information about data, or are some elements subjective interpretations? Finding ways to handle certification and trust will be an important task for builders of the Spatial Web.

3 The Spatial Web's Impact on Institutions and Programs

Below we summarize what the Spatial Web means to a number of spatial data programs, initiatives, institutions, markets, and application domains.

- **Global, National, Regional and Local Spatial Data Infrastructures (SDI)**: The Spatial Web is being populated with connected nodes through the efforts of many organizations: the international Global Spatial Data Infrastructure Steering Committee, multi-nation geographic data organizations such as NATO's DGIWG, universities, the world's national mapping agencies and national spatial data coordination groups such as Geoconnections Canada (working to advance the Canadian Geospatial Data Infrastructure) and the US Federal Geographic Data Committee, the Permanent Committee on Spatial Data Infrastructure for the Americas, and formal and informal data coordination teams in states, provinces, counties and cities around the world. Spatial data infrastructure refers to the full spectrum of institutional, educational, technical, and legal activities involved in making spatial data of high quality readily accessible to people for their various purposes. These activities began before the Web became the dominant platform for distributed processing and data access, and they all contribute to the growth of the Spatial Web.

- **Sustainable Development community**: As demonstrated at the 2002 World Summit on Sustainable Development, dozens of data sources, many of them hosted in Africa by African GIS and Earth imaging practitioners, are now searchable and accessible through a portal with a catalog that implements the OpenGIS Catalog Interface Implementation Specification. The catalog enables users to drill through spatial data from all over the world, in all formats, to track down the information needed about a specific area. The portal is at http://edcw2ks15.cr.usgs.gov/servlet/UNEPServlet?srp=2.

- **Geospatial One-Stop**: This U.S. Office of Management and Budget initiative aims to provide a geographic component for use in all Internet-based E-Government activities across local, state, tribal and Federal government. Geospatial One-Stop provides standards, models and coordination for Framework data, a government geospatial data catalog, and an online access point to this catalog and the data it indexes. Thus it is positioned to both benefit from and contribute to the growth of the Spatial Web. OGC's Geospatial One-Stop Transportation Pilot will test the encoding of transportation data and use of related Web-based services to effectively deal with the many differences that exist in geospatial data maintained by multiple government jurisdictions, providers and users.

- **Homeland Security, Critical Infrastructure Protection and Disaster management**. The data resources and data access provided by Geospatial One-Stop will be critically important in all of these areas. OGC's CIPI-1 pilot program will help U.S. and Canadian agencies develop additional capabilities in the areas of Web-based online geoprocessing and GeoFusion Services, an infrastructure for spatial referencing of varied types of data that include spatial information. The pilot will also demonstrate how public and restricted access data sources can be used together in Homeland Security and disaster scenarios.

- **INSPIRE (Infrastructure for Spatial Information in Europe)**: This European Commission initiative aims at making available relevant, harmonized and quality geographic information for the purpose of formulation, implementation, monitoring and evaluation of Community policy-making, beginning with environmental data. The European Spatial Data Infrastructure (ESDI)

Action Plan (December, 2001, http://www.ec-gis.org/inspire/reports/ESDI_Action_plan1aa13.pdf) offers the following principles, which are essential principles for the Spatial Web:

1. Data should be collected once and maintained at the level where this can be done most effectively.
2. It should be possible to combine seamlessly spatial information from different sources across Europe and share it between many users and applications.
3. It should be possible for information collected at one level to be shared between all the different levels, detailed for detailed investigations, general for strategic purpose.
4. Geographic information needed for good governance at all levels should be abundant under conditions that do not refrain its extensive use.
5. It should be easy to discover which geographic information is available, fits the needs for a particular use and under which conditions it can be acquired and used.
6. Geographic data should become easy to understand an interpret because it can be visualized within the appropriate context selected in a user friendly way.

- **Global change research**: Thousands of agencies, academic institutions and companies engaged in research produce and use spatial information, and thus the Spatial Web is developing as important infrastructure to support their work. Though researchers typically search for spatial data that they might use and reference, the data they produce is often idiosyncratic and unaccompanied by machine-searchable metadata. Discovery of prior research data and the ability to repeat results that involve spatial data are major areas of concern in global change research. The Columbia University Earth Institute and OGC are engaged together in the GISD initiative described above and in activities that specifically address these two problems as they relate to global change and sustainable development research.
- **Utilities**: Like the physical assets of governments, the physical assets of utilities – pipelines, utility lines, communications antennas – are widely distributed geographically. A utility's information about its physical assets and customers typically resides in multiple different systems, particularly in this era of telecom and energy company mergers and acquisitions. Web services whose open interfaces enable integration of workflows that depend on diverse geoprocessing systems will be essential to the survival of companies struggling to compete in the coming decade of change in telecommunications, energy, and water supply.
- **Location Services**: Mobile device service providers have begun to offer location services based on the device's location. The technical architectures are not usually based on standards at this point, but the limitations of a proprietary location services infrastructure will act to limit service opportunities as the Spatial Web advances. The ability, through standards-based applications, to render a location and apply this information in various applications that run on any device, through any transmission architecture is key for rapid advancement of location services in support of the Spatial Web. "Voice over IP" and Web services that are harmonized with wireless physical transport and locating methods will open up a cornucopia of services to indirectly and directly serve the needs of cell phone users and users of other mobile devices.

4 Spatial Web Technical Issues

Below we outline how spatial content, services and applications are being integrated into the Web.

4.1 Web Services – Web-based Distributed Processing

Web Services are Web accessible applications and application components that exchange data, share tasks, and automate processes over the Internet. Because they are based on simple and non-proprietary standards, Web Services make it possible for computer programs to communicate directly with one another and exchange data regardless of location, processing platforms, operating systems, or languages. Web Services can thus lower the costs of software integration and data-sharing. The Web Services standards infrastructure greatly extends users' access to processing resources, including geoprocessing resources. The same infrastructure that supports desktop computing can support mobile computing.

The computing industry's Web Services architecture assumes that there are three roles – Service ***Providers***, Service ***Requestors***, and Service ***Brokers*** – that perform three essential kinds of operations – ***publish, find and bind***. That is, Service Providers publish machine-readable information – service metadata – about their service capabilities (much as Web sites currently publish metadata about their data offerings). Service Requestors send out requests that announce what kind of service is requested. Service Brokers (relying on service catalogs and service registries) function something like today's search engines, receiving service requests and "binding" a service to a service request. After the service available from the Service Provider has been "bound" to the Service Requestor, the service executes. Services can be chained to create more complex applications. In the client-server model, the Service Requestor is a client, the Service Provider is a server, and the Service Broker is middleware.

4.1.1 OGC Web Services (OWS)

Spatial enablement of the Web is important for users of the web, as explained in the introduction to this paper. It is also true that spatial processing (or "geoprocessing") is a processing domain that critically needs the Web. Geoprocessing encompasses a complex and diverse set of operations that are expensive to maintain in standalone, full-featured systems. The remedy is Web Services, which are uniquely suited to provide users with integrated, automated and selective use of geoprocessing functions, such as converting data from two or more servers to the same coordinate reference system. Also, encoding both geodata and geodata metadata (and geoservices metadata) in the Web's eXtensible Markup Language (XML) provides for precise searches of geospatial data sets (see Section **Error! Reference source not found.**) and geoprocessing resources, as well as a client platform (the Web browser) for processing and presenting geodata and integrating it seamlessly with other information.

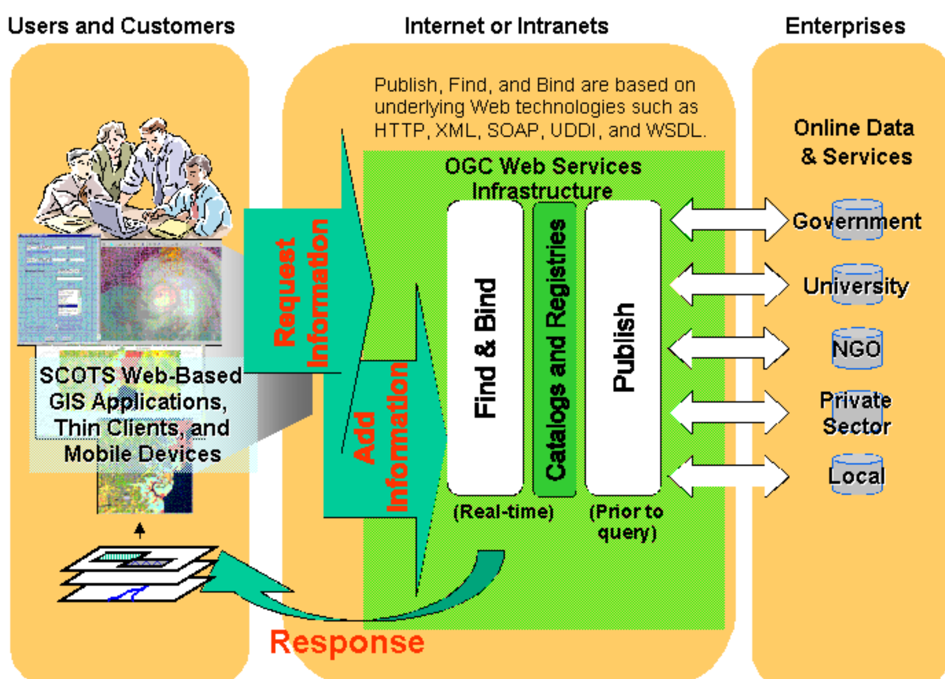

Figure 1. Conceptual Architecture for OGC Web Services

In the OGC Web Services (OWS) vision (see Figure 1), any Internet device potentially has access to tens of thousands of spatial data resources and spatial processing resources. The spatial processing resources might be simple or complex:

- A component resource might provide a single function, such as transforming a coordinate system to enable an overly.
- A full-featured server might provide a complicated decision support capability, for example.
- Component software modules might be gathered "on the fly" to perform a complex application function whose components are actually be hosted on different servers, despite the apparent complexity of passing parameters and instructions among those modules.

Many types of services must work together to enable useful applications of "publish, find and bind." The basic OGC Web Services architecture must accommodate Data Services, Registry Services, Presentation Services, Processing Services, Encodings, and Application Clients. Each service is a collection of operations, accessible through an interface, which allows a user to evoke a behavior of value to the user. When provided with certain parameters through such interfaces, the service calculates a result, which is "served" to an application client. In the OGC Web Services Initiative, the members of OGC are building the interfaces for spatial data and services as well a defining the metadata information to make sure the architecture works in a distributed geoprocessing environment.

4.1.2 Current issues in Web Services development

1. **An open foundation for Web Services**. The World Wide Web Consortium and major computing technology providers like Sun, IBM, Microsoft and others support the idea of a common Web Services architecture, and consensus is being reached on many of the basic elements of underlying Web technologies such as HTTP, XML, SOAP, UDDI, and WSDL. There are many layers. The lower the level (TCP-IP, for example, is low in the "stack") the more easily big competing IT companies reach agreement. The higher the level, the more likely it is that these companies will choose to compete instead of cooperate. It is important for users and providers of spatial technologies to understand where their interests lie and to work for standards that define a Spatial Web that is, from their points of view, fully functional, fertile for high level proprietary offerings, and not constrained by proprietary foundations.
2. **Maintaining a peer-to-peer architecture option**. The Web was designed as a peer-to-peer system, but some Web experts feel that the Web has come to be dominated by centrally managed enterprise systems and proprietary client-server application architectures. Decentralized peer-to-peer architectures for Web Services will depend on someone, perhaps user consortia, putting key application services online for public use.
3. **Payment, security and privacy**. Main application areas for spatial Web Services include e-business and e-government. Many spatial data commerce schemes will require effective and widespread metering and payment mechanisms. Both domains will require effective security and privacy mechanisms. It is likely that such mechanisms developed for the Web in general will apply easily to the Spatial Web, but the Spatial Web's growth may be slowed if these do not become available and widely used relatively soon.

4.1.3 Issues in OGC Web Services development

1. **Dependencies of one pending specification on another**. The OGC Interoperability Program provides a consensus process to develop, test, demonstrate, and promote the use of interfaces, protocols and schemas that enable interoperable geoprocessing. The Interoperability Program organizes testbeds, pilot projects, planning studies, insertion projects and feasibility studies, and publishes resulting technical documents, training materials, test suites, and reference implementations on the OGCNetwork (www.ogcnetwork.org). Its various initiatives also produce numerous Interoperability Program Reports that pass through the OGC Specification Program,

usually resulting in approved OpenGIS Specifications. The OGC Reference Model is designed to help manage the complexity and dynamism of these efforts and the complications of building on an equally dynamic and complex platform of general purpose Web Services specifications. Often there are dependencies among specifications. These issues are discussed and resolved in the OGC Review Board.

2. **Sponsor support for particular capabilities**. OGC Interoperability Initiatives are usually funded jointly by several sponsors, usually government agencies or large companies, who set the purpose, scope and timetable of an initiative. In planning new initiatives, OGC staff and members introduce discussion of capabilities that have not yet been addressed and are not being addressed by current initiatives.

3. **Need for a dependable, available, no cost catalog service**. The global community, or we might say the Spatial Web, needs a catalog service hosted by a reliable and unbiased third party organization(s) to publish, find and bind a key set of fundamental spatial web services.

4.2 The Semantic Web and Geospatial Semantics

4.2.1 Definition of Semantic Web and its key elements

"The Semantic Web," to quote the W3C web site, "is a mesh of information linked up in such a way as to be easily processable by machines, on a global scale. You can think of it as being an efficient way of representing data on the World Wide Web, or as a globally linked database." It is a set of rules for publishing text data so that it can be processed very flexibly, in many ways for many purposes.

Some critics believe that the Semantic Web will develop very slowly, and they point to the slow acceptance of Resource Description Framework (RDF), one of the technologies put forward along with eXtensible Markup Language (XML) as a means for implementing the Semantic Web. Proponents reply that enterprises have <u>not</u> been slow to adopt XML and RDF, even if adoption in the public Web has been slow. They also point out that distributors and users of spatial data have been working with metadata schemas for years, which gives them a head start, and there is rapid market acceptance of Geography Markup Language, the industry's standard for encoding spatial data in XML. Thus, whether or not the Semantic Web gets off to a slow start, the geospatial industry is likely to be leading other industries in practical applications of semantic technologies.

4.2.1.1 eXtensible Markup Language (XML) and XML Schemas

XML is a system of encoding data in text. It can be read and understood by developers and data managers alike, but its main feature is that it can be "understood" and processed by software. XML is thus easily transformed, and so it is easy to integrate and combine XML-based data from many disparate sources. XML will play an important role in the Spatial Web because it provides a platform for the Geography Markup Language (GML), a standard XML encoding for spatial data, and because XML-encoded metadata for spatial data and spatial services provides a basis for fine-grained data and service catalog searches.

XML is a language for writing markup, or data encoding languages and grammars. HTML, for example, (Hypertext Markup Language) has been written in XML. HTML describes the content (the actual text) and the layout (for instance, which text gets bolded) of a Web page, but it only allows users to point to complete Web pages and to single points in those pages. XML goes much further. It provides a mechanism for linking multiple resources into a complex association. It separates presentation and content, enabling the same data to be presented in different styles.

XML is a family of technologies and standards for data modeling as well as simple semantic expression. These include RDF schema definition, XSD (XML Schema) and XLink (eXtensible Linking Language). XML tools and components can be leveraged in the deployment of any domain specific technology that builds on XML. The rich set of XML standards is giving rise to robust applications in a variety of industries, including finance, chemistry, Internet business-to-business, document publishing, multimedia, telecommunications, graphics, and E-government.