The Importance of Going "Open"

An Open Geospatial Consortium (OGC) White Paper

"Open" means your software works with mine,
independent of vendor, like Web browsers and Web servers.

1   Introduction

Standardization is the reason for the success of the Internet, the World Wide Web, e-Commerce, and the emerging wireless revolution. The reason is simple: our world is going through a communications revolution on top of a computing revolution. Communication means "transmitting or exchanging through a <u>common system</u> of symbols, signs or behavior." Standardization means "agreeing on a <u>common system</u>."

Most of today's communications technology standards (HTML, FTP, XML, TCP/IP[1] etc.) are created by standards setting organizations (consortiums or task forces) that seek to develop effective standards that meet everyone's needs without favoring any single company or organization. Organizations like the OGC, the World Wide Web Consortium (W3C), the Internet Engineering Task Force (IETF), and others are open organizations in the sense that any organization can participate, the topics of debate are largely public, decisions are democratic (usually by consensus), and specifications are free and readily available. Each of these organizations is different, all have their internal politics (much is at stake!) and none is perfect. But they are, overall, transparent and democratic international organizations whose visibility, transparency and broad social and economic importance attract careful scrutiny. An "open" process is necessary to arrive at an "open" standard. The openness that OGC promotes is part of this general progress.

In this paper we look at how peoples' needs for geographic information are served by OGC's "opening" of geospatial information technology, and we show the crucial role that users play. We conclude that buyers of geoprocessing software, data and services should review their requirements and draft "open architectures" that lead to purchase of solutions that implement the appropriate OpenGIS Specifications.

2   GIS User Needs and the Need for Openness

The OGC sets GIS standards[2], addressing user needs that can only be addressed by cooperation among GIS vendors. In describing these needs, we begin with the most general statement and then proceed to more detailed statements:

Overall, users want to maximize the value of past and future investments in geoprocessing systems and data.

---

[1] Hypertext Markup Language (HTML), File Transfer Protocol (FTP), eXtensible Markup Language (XML), and Transmission Control Protocol/Internet Protocol (TCP/IP).

[2] More accurately, the OGC sets standards for "geoprocessing," which includes capabilities now found in geographic information systems (GIS) and digital systems for Earth imaging, web mapping, location based services, surveying and mapping, CAD-based facilities management, webs of geolocated sensors, navigation, cartography, automated mapping etc. The "standards" are consensus-derived specifications for open interfaces, protocols, schemas etc. that enable different vendors' systems to exchange data and instructions, and that enable full integration of these capabilities into all kinds of information systems.

That general need points to the following three more specific classes of user needs:

1. The need to share and reuse data in order to decrease costs (avoid redundant data collection), get more or better information, and increase the value of data holdings.

2. The need to choose the best tool for the job, and the related need of reducing technology and procurement risk (i.e., the need to avoid being locked in to one vendor).

3. The need for more people with less training to benefit from using geospatial data in more applications: That is, the need to leverage investments in software and data.

Those three classes of user needs point to the following still more specific needs:

1. **The need for organizations to have access to each other's spatial information without copying and converting whole data sets**. This includes:
   a. The need for passing data and instructions between different vendor's systems.
   b. The need to integrate various data models, formats and coordinate systems.
   c. The need to visually integrate map displays (symbology) from different data servers.
   d. The need to find and evaluate data and services held in other locations.

2. **The need to have the pieces of a solution work together.** This includes:
   a. The need to add or replace a capability in a current vendor solution, with minimal integration costs, and have it work seamlessly.
   b. The need to understand the interoperability requirements of application domains and define architecture profiles and application design strategies for each.
   c. The need to integrate geoprocessing Web services with mainstream Web services, and to develop "loosely coupled systems" using network-resident services.

   3. **The need to base geoprocessing on the World Wide Web open architecture**, which includes common best practices, "reusable" data and Web Services-based components.

   Once geoprocessing systems work together and work with other systems on the open network, new opportunities/needs arise that require a standards foundation:
   a. The need to organize geographic data stored in text and on video, audio, and other media.
   b. The need to access and process on-line sensor data from multiple sources.
   c. The need for Location Based Services portable across devices, networks, and providers.
   d. The need for "semantic translation" from one data model to another.
   e. The need to take advantage of grid computing for geoprocessing applications.

The general solution to these needs: geoprocessing systems and components that interoperate across open interfaces in the context of global (or in some cases local) distributed computing platforms (usually the Web.) Technical participants in OGC translate the user needs above into technology requirements that are formalized in open standards – OGC's OpenGIS Specifications. When users plan information systems by designing "open architectures" based on open standards, and when they buy products that conform to these open standards, they get the interoperability they need.

3   The Definitions of "Openness" and Open GIS

3.1   Open Standard

What is an "open standard"? OGC defines an open standard as one that:

1. **Is created in an open, international, participatory industry process**, as described above. The standard is thus **non-proprietary**, that is, owned in common. It will continue to be revised in that open process, in which any company, agency or organization can participate.

2. **Has free rights of distribution**: An "open" license shall not restrict any party from selling or giving away the specification as part of a software distribution. The "open" license shall not require a royalty or other fee.

3. **Has open specification access:** An "open" environment must include free, public, and open access to all interface specifications. Developers are allowed to distribute the specifications.

4. **Does not discriminate against persons or groups:** "Open" specification licenses must not discriminate against any person or group of persons.

5. **Ensures that the specification and the license must be technology neutral:** No provision of the license may be predicated on any individual technology or style of interface.

By this definition, a **de facto standard** established by one company or an exclusive group of companies or by a government is not an open standard, even if it is published and available for use by anyone at no charge. The Spatial Web, like the Web, needs open standards, as defined above.

3.2   Frameworks, Open Architectures, and Reference Architectures

The Open Group Architecture Framework (TOGAF)[3] defines an **architectural framework** as a tool for assisting in the production of organization-specific architectures. An architectural framework consists of a *technical reference model*, a method for architecture development and a list of component standards, specifications, products and their interrelationships which can be used to build up architectures. In other words, a "framework" is a broad, high level, conceptual model for a technology system that provides interoperability among diverse systems.

A **Technical Reference Model** (TRM) is defined in the TOGAF as a structure that allows the components of an information system to be described in a consistent manner. In OGC, our TRM is ISO RM-ODP[4]. Our architectural framework is the *OGC Reference Model*[5].

An **interoperability platform** generally refers to the actual interfaces, based on a framework, that are available to support interoperability. An "architecture" is a more specific kind of model, usually for either a specific enterprise information system or a specific vendor's set of products, that lays out in detail how specific types of processing modules will use specific interfaces to enable specific information flows involving specific kinds of data. An "open architecture" is thus an architecture that specifies certain open interfaces and open data models (see Section 3.8). That is, it is a nonproprietary architecture. A **reference architecture** is an open source (see page 3.9) architecture intended to provide architecture developers with a template and implementation guidance.

3.3   OpenGIS®, Open GIS®, and open GIS™

OGC registered the trademark "Open GIS" and OpenGIS" in countries around the world to assert the importance of open standards in geoprocessing and to protect its standards with a legal brand. A software vendor whose software implements interfaces based on OGC's standards can claim that a product "implements" particular OpenGIS Specifications. If the product has passed a conformance

---

[3] The Open Group Architecture Framework can be seen at www.opengroup.org/architecture/togaf/ .

[4] The ISO Reference Model for Open Distributed Processing (ISO RM-ODP) can be seen at http://www.enterprise-architecture.info/Architecture_Standards.htm .

[5] *OGC Reference Model*, OGC document no. OGC 03-040, September 16, 2003. The ORM documents a framework of interoperability for geospatial processing ranging from tightly coupled, real time systems on a single CPU to the "Spatial Web" -- the open environment that enables barrier-free communication of geographic information among users of the World Wide Web. The ORM is a living document and will be updated periodically as OGC membership continues to advance geoprocessing interoperability. To download the ORM free of charge and to subscribe to future updates, visit http://portal.opengeospatial.org/files/?artifact_id=3836.

test for a particular OpenGIS Specification, the vendor can claim that its product conforms to that version of a specification and it can use OGC's trademarks to assure buyers of the veracity of those claims. The phrase "open GIS" (with a small "o") is also a trademark of OGC, with the same meaning as "Open GIS," though "open GIS" is not a registered trademark.

Organizations that commit to "OpenGIS" commit to deploying and using distributed systems based on the *OGC Reference Model* (see Section 3.2).

3.4   Interoperability

Software **interoperability** describes the ability of locally managed and heterogeneous systems to exchange data and instructions in real time to provide *services*[6]. Interoperable systems are generally *distributed* (i.e., at different places on the network), though in OGC's case, interoperability also applies to different types of systems or similar systems from different vendors communicating while running on the same computer. The interoperability challenge, successfully met by means of consensus reached in inclusive consensus processes, is to balance the users' need for compatibility with the autonomy and heterogeneity of the interoperating systems.[7]

It is important to remember that proprietary algorithms typically run unseen in the "black box" component whose public face is the open interface. Some server components will outperform others and/or offer capabilities not offered by others, though they may all communicate with clients through a common interface. In an interoperable environment, competition among vendors is based on such differences in capabilities and performance, and is not based on which format the user's data is stored in, or which software provides the display function.

Interoperability also refers to interoperability across *time* (evolution of systems over time with **backward and forward compatibility**). When users participate in standard setting, backward and forward compatibility have a high priority.

Recall that the Internet Protocols (IP) were introduced as *inter-net* protocols, for inter-networking, that is, moving data between different networks. At the time there were many different networks, whose names we no longer remember. Inter-networking gave way to the use ONLY of the IP protocols. The result was the Internet and then the World Wide Web, which provided a platform, an interoperability platform, supporting an extraordinary proliferation of services and applications. This is the appropriate model for a geoprocessing interoperability platform.

3.5   Open Platforms

The meaning of *open platform* depends on the context. In general, the term *platform* used to denote any specific hardware and operating system combination, such as the Windows/Intel platform or the Solaris/SPARC platform. It is now used more generally to describe an application programming interface (API) or set of APIs that provide access to computing power, database, GIS or other services hidden "underneath" those APIs. The acronym "API" is giving way to "interface" in programmer-speak. By the definition of "open" in this paper, no single vendor provides an open platform unless all the exposed interfaces are open interfaces, as described below. An open platform needs to be like the IT industry's Web Services platform, which is still, as of September, 2003, largely unencumbered by proprietary restrictions and is the product of a consensus process.

3.6   Open Systems and Open Interfaces

The definition of **open systems** has changed over time, but today open systems are usually considered to be systems that interoperate through **open interfaces**[8]. An **interface** is simply a

---

[6] A service here is an activity, such as data access or coordinate transformation, performed by a server component on behalf of a client component.

[7] Finkelstein <http://www.sistm.unsw.edu.au/people/FDABOUS/proposal/node139.html#Finkelstein00>

[8] Another paradigm for interoperability is based on *brokers*, such as OMG's Common Object Request **Broker** Architecture. Such a "broker" converts one product's interface into another product's interface "on the fly."

common boundary, a means to make a connection between two software components. An interface on the client presents an ordered set of parameters (with specific names and data types) and instructions (with specific names and functions) to an interface on the server that is structured to read and respond to just such a set of parameters and instructions. Thus an interface enables one processing component to exchange data and instructions with another processing component.

Some interfaces satisfy part but not all of the "openness" definition above. The information technology world has been steadily evolving toward greater openness, so many older systems still in use interoperate in what now appear to be limited ways. Such systems from a variety of geoprocessing software companies use interfaces that the companies have published for coding by integrators and application developers. Reaching that situation was progress, considering that at one time, few proprietary interfaces were published. From today's perspective, however, there are reasons not to depend on such **published, but proprietary interfaces**:

- In the old paradigm, a client system needs a separate interface for each vendor's system. The biggest advantage of open interfaces is "build one, access many." With truly open systems, solution providers no longer need to build custom interfaces. Users are no longer isolated in technology **stovepipes[9]** and no longer **captive to ("locked in to") single vendor solutions[10]**. For example, in the early days of Web mapping, you needed the map server and the map viewer (or client) to be from the same vendor. Now (or as soon as all map servers implement the OpenGIS Web Map Server Specification) all spatial clients and servers are able to request and provide map images regardless of what company wrote the software. (Other OpenGIS Specifications, of course, go far beyond mere exchange and overlay of map images.)

- From time to time, **vendors change or enhance their interfaces, forcing client systems to change and forcing users to upgrade**, perhaps without notice or opportunity for input. In contrast, the consensus process in a consortium gives users, integrators and developers both notice and opportunity for input, increasing continuity. This collective approach may seem like it would be a burden to vendors, who want to move quickly and independently. But vendors recognize that we have moved from the frontier age of software to an age of communication, in which many interdependent stakeholders share data and work together. Open standards impose a few constraints on developers, but they open huge opportunities, as demonstrated by the explosion of innovation and business opportunity that has resulted from the Web.

- Integrators and application developers would probably spend more time learning how to use the proprietary interfaces than they will spend learning how to use OGC's interfaces. (Consensus interfaces need to be well documented.) And if they need to integrate two or more such systems into their solutions, the learning time and development time are compounded. One reason **open systems result in greater innovation** is that they remove this burden from development budgets, freeing resources for innovation. One bad result of the old paradigm has been that integrators tend to learn and then use one system exclusively simply because the cost of mastering more than one is too high, which further limits the choices available to the user.

3.7   Open Formats and GML

In the not so distant past, it was important to know whether your data was in a particular vendor's (published or unpublished) format, such as SHAPE or DXF, or in an open government format such as TIGER or VPF, or in an exchange format such as SDTS or SAIF. Now, format is not a major issue when vendors' systems communicate through open interfaces. People sometimes want to archive whole data sets in the format native to the software they are using, or in an exchange format, but

---

[9] "Stovepipe" is a metaphor commonly used to describe systems that are integrated "from top to bottom" but isolated laterally, i.e., from other systems. A stovepipe system might be a system from a single vendor or it might be a system built by an integrator, but it is not an open system.

[10] "Captive to a vendor" means a buyer must buy from a particular vendor. This might be because all the buyer's potential data sharing partners use that vendor's software, or because the buyer's legacy systems are from that vendor, or because the buyer's institution mandates that purchases shall be only from that vendor.

bulk conversion of data files from one format to another is becoming less and less necessary. The new world of "open" enables "transparent" conversion of small amounts of data "on the fly" when the data are needed. This avoids the enormous investment in converting extra data that may never be used afterwards, and also provides access to up-to-date data.

The interesting twist here is that the Web provides justification for something like a universal open format: Virtually all Web browsers now include software to process text encoded in the eXtensible Mark-up Language (XML). XML can be described as a language for creating self-describing data files, that is, data files whose headers explain how to interpret the data that comes after the header. This has turned out to be a very powerful concept. Scores of industries and professional domains have seized on the opportunity to develop XML schemas (schemas are essentially formats) to capture the specific kinds of information that need to be shared within those industries and domains by organizations whose legacy systems are very different from each other's.

Similarly, the members of the OGC developed the Geography Markup Language, which is well on its way to becoming the standard XML encoding for geospatial information.[11] XML-encoded geospatial metadata, (parts of which conform with GML) are a keystone element of the OGC Web Services architecture that makes possible detailed, complex, automated searches for spatial data and spatial services on the Web. Also, GML separates content from presentation, so the way in which data is presented (on desktop systems and PDAs, for example) is entirely under program control and can thus be tailored on the fly to suit display device capabilities. Very importantly, one of the major breakthroughs with GML is that, when used with XML tools, GML makes it possible to resolve many of the difficulties associated with incompatible data models (see section 3.8).

It is not difficult to create profiles (application-specific variations) of GML, and this is what most data developers will do. The Ordnance Survey of Great Britain and the US Census Bureau (in its TIGER data) have committed to GML. But everyone in the geoprocessing industry should be aware that it is also easy to create new XML schemas for geographic information that are not profiles of GML, and herein lies a risk of a new Tower of Babel rising in the world of geographic information.

3.8   Standard Data Models

Efforts are underway in many countries to develop standard geospatial metadata schemas and standard data models. However, a nationwide standard that meets both local and national needs is very difficult to achieve, and the cost of attaining consistent data content seems to many (particularly those at the local level) to make this an impractical goal. These standard data models will, however, have an important role as "Rosetta stones" that enable each user to map their data to a common model so software can go from one local model to the national model and thence to the user's own local model that is different from the first. One-to-one mapping of data models is unworkable when there are thousands of models to map between. GML enables a one-to-many solution.

One-to-many mapping of data models is made possible by XML tools. The XML tools (prototyped in OGC's GOS-TP and CIPI-2 pilot projects) map GML-encoded data from a local model to the national model and vice versa. The data thus becomes "as useful as possible" to the data sharing partner who uses a different model. Certain elements of one model do not map to the other, but the XML tools make these inconsistencies plain in all their details, so that it is easy for data managers to focus on the critical schema elements that don't map. This makes both data sharing and data coordination much easier. It makes it easier for people at the local level to accommodate national standards in an affordable and practical way, and it makes it easier for people at the national level to work with local data that hasn't been converted in all its details to the national standard.

Another benefit of the GML approach is that this technology makes content standards easier for software vendors and integrators to support. Currently, content standards are expensive to support,

---

[11] GML has been compared to the Spatial Data Transfer Standard (SDTS), but because GML is implemented in XML, and because XML is a robust and integral part of the (World Wide, not US) Web, GML is in a separate category from SDTS.

and smaller companies that do not support them are at a disadvantage. The new approach thus enhances competition, increasing the choices available to users in the market.

3.9 Open Source

It is important not to confuse "open source" with "open standards." They are entirely different. The special licenses that govern use and sale of open source software exist not to ensure profits to the software's owner, but to ensure that the software's source code remains in the public domain (free to all), though companies are allowed to sell products that include the source code. Open source software is usually developed not by single company but by a distributed, informal team of developers. Open source software developers use OpenGIS Specifications for the same reasons commercial developers use them: to make their products interoperate with others.

3.10 Open Portals

Geospatial portals can be thought of as the "hubs" or "geoinformation resource supermarkets" in the Spatial Web. A portal is a Web site that gives visitors organized access, typically through catalog services (services not too different from those provided by search engines), to data and processing resources on the Web, and perhaps also to people, organizations and publications. A portal offers an organized collection of links to many other sites. A portal thus can be used to aggregate content. And by attracting a large number of visitors who share a common interest, a portal also aggregates content seekers for the benefit of content providers and advertisers and potentially for the benefit of that community of content seekers.

Users of geospatial portals that are based on OpenGIS Specifications for software interfaces and GML encodings can immediately access – pan, zoom, compose, save and print – views of digital geospatial content held on diverse Web-connected servers. Multiple maps from multiple servers can be overlaid and "flipped through." Data providers register their data for access via the portal. Applications (and other portals) can integrate portal resources into information offerings and work flows.

Software comprising a portal can be from one or several vendors. The key criteria for an open portal is that it be fully interoperable with all the other spatial resources on the Web that are equipped with interfaces, encodings etc. that implement OpenGIS Specifications.

3.11 Open Procurements

Open interfaces make it possible to specify in a procurement a type of software component, rather than specifying one particular vendor's software. The goal is to be able to build incrementally with "best of breed" components, and to be able to "swap out" and "swap in" software components. For example, the procurement language might be, "Application shall implement a geocoding service that is accessible via the OpenGIS Location Service Geocoder Interface Specification." This offers geoprocessing software buyers unprecedented savings and flexibility. With respect to a geospatial portal or other Web-based geospatial solution, whether or not the solution uses components from multiple vendors, all of its connections to outside resources and users must be through open interfaces. If not, the implementation remains a closed system, a stovepipe, an island of automation that prevents present and future inter-institutional interoperability.

Open standards make open procurements possible. The open procurement specifies functional requirements and interoperability requirements. Products may be benchmarked in an interoperability pilot. Vendors that meet contractual requirements and demonstrate functional requirements and interoperability are thus qualified to provide components. This multi-vendor procurement process has been used successfully by numerous government organizations over the years in other technology domains, and now organizations can use this process for geoprocessing software purchases.

Up until now, the user's choices have been the competing views from competing vendors who sell "soup to nuts," full-featured, expensive monolithic systems. Components interoperating through OGC Web Services now allow users to construct a solution based on *their* special view of the world, not a view that is the view of a particular vendor.

3.12 Integration and Interoperability

Integration in our industry means making spatial data accessible from multiple technologies and software vendors and making spatial data and spatial functionality available to other IT systems such as customer response management, logistics, location-based services for wireless devices, etc. The benefit is that users can thus access, combine, and disseminate geospatial information from distributed and varied information sources. Integration streamlines workflow and reduces costs of information production, maintenance and dissemination.

Integration is far more efficient, with significant immediate and downstream cost savings, if the integration can be accomplished with open standard interfaces instead of proprietary and/or custom interfaces. Enterprise systems integrated using open interfaces can enjoy the "network effects" that result from the same interfaces being used in the world outside the enterprise. OGC's goal is to create a single, vendor-neutral infrastructure for integration that works everywhere, across all platforms, technologies and types of devices.

4 Conclusion – What Users Must Do

Just as the Web has changed the way we access information and information processing resources, OGC's standards are changing the way we access spatial information and spatial information processing resources.

Standards setting depends on technology providers implementing the standards in products, <u>and it depends on technology users buying those products</u>. The OGC members cannot do it alone:

- Vendors in OGC who have committed significant resources to developing the OpenGIS Specifications, with input from users, did so as an act of faith. They have implemented many of these specifications in products, but **they depend on their customers and potential customers to understand the standards and ask for them.** Public and private sector organizations owe it to themselves, their customers, shareholders, stakeholders, data sharing partners and constituencies to use the OGC Reference Architecture (ORM) as the model for their next purchases. The ORM makes it easy to see which OpenGIS Specifications are relevant to their needs.

- User organizations in OGC that have committed resources to developing the OpenGIS Specifications by providing requirements in testbeds and pilot projects need other user organizations to help "move the ball forward." **Particularly with respect to governments' needs to protect citizens and property (with special attention to critical infrastructure) in time of disaster, widespread user acceptance of open standards is critical.** Similarly, industries that depend heavily on geospatial information are motivated to reduce their costs of system integration and data integration, and they can do this best if all their internal geoprocessing systems and the geoprocessing systems of data sharing partners are interoperable. In areas like location-based services and sensor webs, the opening of whole new markets depends on product strategies based on open standards.

Thus it is incumbent upon buyers of geoprocessing software, data and services to carefully review their requirements and then draft interoperability architecture documents that lead to purchase of solutions that implement the appropriate OpenGIS Specifications. This can be done piecemeal, one upgrade or add-on at a time, or, if it is time for the organization to put a whole new solution in place, it can be done comprehensively, all at once. OGC and OGC's members can help by examining use cases and explaining where open interfaces can be specified into the architecture on which procurements will be based.

Much is at stake, and much will be set in motion when a large number of people each take a small step in the direction of openness.