Sharing Geographic Information: An Assessment of the Geospatial One-Stop

Michael F. Goodchild,* Pinde Fu,† and Paul Rich‡

*Department of Geography and National Center for Geographic Information and Analysis, University of California, Santa Barbara
†Environmental Systems Research Institute
‡Creekside Center for Earth Observation

Humans have always exchanged geographic information, but the practice has grown exponentially in recent years with the popularization of the Internet and the Web, and with the growth of geographic information technologies. The arguments for sharing include scale economies in production and the desire to avoid duplication. The history of sharing can be viewed in a three-phase conceptual framework, from an early disorganized phase, through one centered on national governments as the primary suppliers of geographic information, to the contemporary somewhat chaotic network of producers and consumers. Recently geolibraries and geoportals have emerged as mechanisms to support searches for geographic information relevant to specific needs. We review the design of the Geospatial One-Stop (GOS), a project sponsored by the U.S. Federal Government to provide a single portal to geographic information, and reflecting the current state of the art. Its design includes a portal to distributed assets, accessible through a simple Web browser, a catalog based on the widely used Federal Geographic Data Committee metadata standard, services to assess and validate potential accessions, directories to available geographic information services, and automated metadata harvesting from registered sites. GOS rep resents a significant technological advance, however, its potential to provide a general marketplace for geographic information beyond government data has not been realized. Its future will be driven in part by technological advances in areas such as searching and automated metadata harvesting, as well as by clearer definition of its domain, either as the geoportal for U.S. data or as a broader geoportal with appropriate international or private partners. Incorporation of informal and heuristic search methods used by humans appears to offer the best direction for improvement in search technologies. *Key Words: digital library, geoportal, metadata, spatial data.*

Humans have always needed to share information about the Earth, whether as bands of hunter-gatherers critically dependent on sharing knowledge about resource locations, as sixteenth-century traders sharing knowledge of sea routes, as nineteenth-century explorers of the polar regions, or as twentieth-century military commanders sharing knowledge of terrain and enemy locations. Accumulating and sharing information about the Earth's surface is central to the discipline of geography, and indeed there would be little purpose in a discipline unable to share knowledge among its practitioners, with the academy in general, and with society at large. Technologies for information sharing are often discussed under the rubric of geographic information science (GIScience), but their significance for the discipline of geography and many of its constituent parts argues for a discussion that is framed by the discipline as a whole.

As humanity's technologies for representation and communication of information have changed, so too have the methods used to share geographic information (GI)—from the stick maps of early Pacific voyagers, to the printed maps of the Renaissance, to the digital technologies of the late twentieth century. In turn, technological change has driven a reconceptualization of the nature of GI, of the purposes of sharing, of the institutional frameworks within which sharing occurs, and of the economic models that make sharing rewarding to both the provider and the user of information. The pace of technological change continues to accelerate, and with it the pace of change in practices of geographic information sharing. The purpose of this article is to reflect on the history of GI sharing and on the current state of the art, and to speculate on future trends.

The current state of the art is represented by the host of Web sites variously known as geospatial data libraries, geolibraries (Mapping Science Committee 1999), geoportals (Maguire and Longley 2005), geospatial archives, and clearinghouses that has emerged over the past decade. A recent Google search for "geoportal" alone resulted in 182,000 hits. All of these technologies and sites offer to facilitate access to GI, by supporting search, evaluation, and downloading, and in some cases licensing or purchase. Such sites have been constructed by national governments, international agencies, local governments, private-sector companies, universities,

Annals of the Association of American Geographers, 97(2), 2007, pp. 250–266
Initial submission, May 2005; revised submission, February 2006; final acceptance, May 2006
Published by Blackwell Publishing, 350 Main Street, Malden, MA 02148, and 9600 Garsington Road, Oxford OX4 2DQ, U.K.

and by individuals. Surveys and analyses have been published at various stages in the development of these technologies (see, e.g., Masser 1999; Groot and McLaughlin 2000; Crompvoets and Bregt 2003; Williamson, Rajabifard, and Feeney 2003; Crompvoets et al. 2004; and for a quantitative analysis of usage patterns see Kline 2004). The European INSPIRE project (inspire.jrc.it) is one of the most ambitious current projects, since it must address differences of language, practice, and culture between all members of the European Union in its efforts to provide a single point of access to GI.

Rather than undertake an updated review, which would itself be a valuable contribution, in this article we use a case study as representative of the current state of the art: the Geospatial One-Stop (GOS) (www.geodata.gov/gos), a portal to GI recently constructed as part of the U.S. President's E-Government Initiative (www.whitehouse.gov/omb/egov). Some geoportals may be more advanced in some respects (and less advanced in others), but GOS is typical of current technology in its objective of being a single point of entry for users in search of GI. The degree to which it achieves this goal, and by extension the degree to which other geoportals achieve similar goals, is a major theme of this article.

Following Goodchild, Egenhofer, Kemp, et al. (1999), we define GI as a collection of tuples of the form $<\mathbf{x},z>$, where $\mathbf{x}$ is a location in space-time and z is a set of properties. In other words, GI connects properties (attributes, characteristics, variables) to locations and times within the geographic domain. We define the domain, consistently with the practice of the discipline of geography, as approximately bounded by locations on or near the Earth's surface (say, from the top of the atmosphere to the lower limits of groundwater circulation), by times near the present (say, from the beginning of the Holocene to 100 years in the future), by spatial resolutions from 10^{-2} to 10^7 m, and by temporal resolutions from 1 to 10^9 sec, at the same time acknowledging that no firm limits on this domain are ever likely to be agreed upon. The terms *geographic* and *geospatial* are essentially synonymous in this context. Defined in this way, GI is a particularly well-defined subset of information in general, and thus amenable to comparatively rigorous theorizing.

Point observations, such as those encountered in weather measurements or spot heights or incidents of crime, are examples of simple tuples. Much GI occurs in aggregate form, however, as when statements are made about entire counties or cities, or terrain is represented by contours, or ownership is defined for an entire land parcel, or when geographers define formal regions and assign uniform characteristics to them. In such cases aggregate information can, in principle, be decomposed into large numbers of point observations or simple tuples, but in practice this is inefficient. We define a *GI object* (GIObject) as one or more tuples or aggregates defined for some practical purpose—in other words a geographic data set or database.[1]

Any GIObject must be created through some process of production, whether by observation, accurate surveying, cartography, aerial photography, printing, digital copying, or any combination of these. Any production process has associated scale economies, in the form of diminishing marginal costs—for the number of copies produced, as in printing, or for the geographic area covered by the GIObject and thus its size, measured in some appropriate units.

We define the *provider* of GI as an individual, agency, Web site, or archive possessing or controlling one or more GIObjects. A *user* is defined as an individual or agency in receipt of one or more GIObjects from a provider. We define the location of the provider as *D*, the location of the user as *U*, and the location or region represented or described by a GIObject as the *subject* location *S*.

In the next section we discuss arguments for and against sharing of GI. We then present a simple conceptual model of the historic phases of GI sharing, culminating in the current concept of spatial data infrastructure (SDI). We discuss the adaptation of the technologies of the Internet and the World Wide Web to GI sharing, with a brief review of developments since the early 1990s. We then describe GOS, including major features of its technical design. This is followed by a discussion of prospects for the future.

Why Share?

There are of course a vast number of reasons why a user at some location *U* might want information about another location *S* that is beyond the reach of the user's senses. Clearly, geographic information is essential to geographic research, and direct observation through field work is central to the discipline. But geographers must often rely on data that are sensed remotely or obtained from secondary sources, or they must refer to past periods that are no longer directly observable. More broadly, humans need GI in order to plan travel, to trade, to conquer and administer, to exploit and manage resources, and to conduct many other forms of human activity. They also have an innate curiosity about the world, a desire to know what it contains, and a tendency to simplify the world when representing it. When *U* and *S* are coincident in space and time, the user can rely to some degree on the human senses to provide knowledge of his or her surroundings. But when they are separated by more than some minimal distance or interval of time then it is necessary to rely on a provider to supply the information by some means, whether in the form of a printed map, a text description, a telephone conversation, or a digital data set. Ultimately, sharing is about the transfer of GI from providers to users.

One fundamental argument for sharing derives from scale economies in production. Because the cost of making a printed map is largely in the high initial costs, there is a clear incentive to collect data and produce maps for generic purposes in order to maximize the number of users; maps are produced for an area that will generate maximum demand, and for a theme that is similarly widely used. The life of a printed map is also maximized by ensuring that the map is valid for the greatest possible length of time, thus encouraging the mapping of themes that are as static as possible, and avoiding maps of dynamic phenomena. The costs of digital reproduction of GIObjects are similarly high initially, and thus any producer will also tend to select areas and themes that are popular and inherently static.

However, GIObjects can also be subject to significant scale diseconomies. Consider a user base that is segmented into distinct *information communities*. We define an information community as a group that shares a common language, a common set of definitions, or a common set of data format standards such that a user within the community can understand a GIObject communicated by a provider within that community. However, sharing across communities is frequently problematic, due to differences of language, culture, terminology, or definitions of terms. An ecologist might have difficulty understanding a land-use classification implemented in a GIObject constructed by a regional planner; a user in the United States might have difficulty understanding a GIObject whose properties are defined in Chinese; and such difficulties even occur in disciplines as diverse as geography. In extreme cases, the use of vague or complex definitions for such properties as soil class in effect partitions the user community into subsets, and even into individuals who are unable to share data if each individual is free to place his or her own distinct interpretation on the meaning of terms. Overcoming such differences—that is, making GIOs *interoperable* across information communities (Goodchild, Egenhofer, Fegeas, et al. 1999)—often requires substantial investment. It may require the development of more-extensive documentation (*metadata*), or the use of software to convert data formats, or the translation of documentation. If these costs accrue to the provider, they create scale diseconomies, discouraging actions that would expand the user base beyond a single information community. If they accrue to the user, then they also act to discourage use, thus limiting the size of the provider's potential user base and the ability to realize scale economies.

In recent years another argument for sharing has become popular: the desire to avoid duplication in GIObject production, and its attendant societal costs (Mapping Science Committee 1993; Masser 1998). Consider a common theme such as elevation, and suppose that one agency has a mandate to produce GIObjects of this theme. If the cost of production of these GIObjects is sufficiently high, it would clearly be undesirable for any other agency or individual to attempt to produce them if they could be obtained more cheaply by sharing. Society as a whole would thus be better off if the first agency shared its elevation GIObjects. In a research context, the cost of field campaigns is very high, so it benefits science in general if data collected from the field can be shared over as many researchers as possible.

There are several nuances to this argument, however. First, suppose that both producers are elements of the same government, such as the U.S. Federal Government. One might suppose in this case that the second agency would try to justify production in order to bolster its own power and budget, that the first agency would try to avoid sharing in order to protect its own "turf," and that duplication would be a clear case of "government waste" facilitated by somewhat dysfunctional bureaucratic behaviors. Pejorative terms such as "stovepipe" and "silo" are often used in such arguments, which have traditionally resonated well with outsiders. In recent years such arguments have been used to push for privatization of GIObject production on the grounds that the private sector cannot afford strategies that are counter to basic scale economies (Mapping Science Committee 2001). Similarly, a researcher may argue for the sole right to use data that he or she collected, at significant expense, at least until some minimal time has elapsed.

Second, the argument is often grounded in naïve assumptions about the nature of geographic themes $\mathbf{z}$. It is easy, for example, to imagine that two agencies both producing GIObjects representing wetlands are duplicating each other's efforts (for an analysis of overlapping wetland mapping activities by four Federal agencies, see FGDC 1996). But the respective definitions of wetlands may be substantially different, reflecting different agency mandates. Scales may also be different, and simple conversions across scales and definitions may be impossible. Even such an apparently well-defined property as elevation is subject to differences of spatial resolution and of the definition of ground surface (bare ground or canopy, with or without structures) that reflect differences of purpose, such that a digital elevation model constructed for air navigation by the Federal Aviation Administration would be substantially different from one constructed by the Federal Emergency Management Agency for modeling floods.

Third, the argument assumes that one representation of a given phenomenon can be sufficient and that another representation provides little added value. In reality, however, it is impossible for any digital representation to provide a perfect replica of a real phenomenon (Goodchild and Gopal 1989). In measurement theory, it is well known that the mean of n independent sample measurements has a standard deviation equal to the standard deviation of each individual measurement reduced by a factor of $\sqrt{n}$. Similarly, the combination or conflation of several independent maps of the same phenomenon can often be expected to provide a better estimate of the phenomenon than can individual maps. According to this argument, two independently produced digital elevation models of the same area will always contain more information than either one, even if their spatial resolutions and levels of accuracy are different. Estimates of elevation made by combining information from digital elevation models, samples of spot heights, and digitized contours are preferable to estimates made from any one of these sources alone. Thus the total benefit of apparently duplicate GIObjects can be greater than the benefit of any one, provided the GIObjects were created independently.

Finally, a blanket opposition to duplication is not warranted, given the value of competition in an open marketplace. Although competition without sufficient resources by any entity can produce incomplete or inadequate GI, multiple versions of the same GIObject will surely lead eventually to innovations and improvements in all versions in an open competitive environment, or to the withdrawal of poorer versions from the market. Clearly society's expectations differ between the public and private sectors: competition is essential to the latter, but is seen as both bureaucratic dysfunction and unnecessary duplication in the former—and competition between the public and private sectors is rarely acceptable.

Models of Sharing

We propose a simple conceptual model of GI sharing consisting of three phases, which we term historic, enlightened, and contemporary.

In the historic phase, which spanned a long period of hand-drawn maps, included the invention of printing, and extended into the nineteenth century, the production and consumption of GI was driven by a range of human needs, from commerce and trade to geographic research, land ownership, and warfare. There was little coordination of production. Maps, charts, texts, and other forms of geographic description were created by individuals, corporations, governments, and publishers. Some were offered for sale; others were made accessible through libraries. Some were closely held, and their great value led in some instances to intrigue and theft as one trading group or nation sought the geographic knowledge of another (Pedley 2005). During this phase geographic knowledge was clearly incomplete, with a large proportion of the Earth's surface being essentially unmapped.

The enlightened phase began in the late eighteenth century, and was led primarily by the imperial powers and their need for GI as a tool for establishing and maintaining dominion (Wood 1992). GI production was extremely expensive, and its high costs could only be borne by national governments who would organize GI production as a central service for public benefit, tied in many instances to national power and prestige. National mapping agencies (NMAs) were established, in many cases as branches of the military, though in the United States isolationism and the need to understand the nation's own largely unexplored landmass led to the establishment of a civilian NMA, the U.S. Geological Survey (USGS). During this phase the idea took hold that mapping could be complete, and by the early twentieth century virtually all of the Earth's surface had been mapped at some level. The International Map of the World project was proposed by Albrecht Penck in 1891 as a collaboration between NMAs to produce the world's first complete large-scale map (at a scale of 1:1,000,000) and was not formally abandoned until the 1980s (Thrower 1996).[2]

We suggest that during this enlightened phase there was belief in the possibility of perfect mapping: that eventually, with enough effort, it would be possible to create a perfect representation of the Earth's surface. NMAs pushed steadily to larger and larger scales, each constituting a more-accurate (and more-expensive) representation. In the United States this culminated in the USGS project to map the forty-eight coterminous states at 1:24,000, a project involving more than 55,000 sheets that was finally completed in 1992. In the U.K., the Ordnance Survey completed (and maintains) coverage at 1:10,000, with much of the nation at 1:2,500. In the last decades of the twentieth century NMAs pushed for more accurate geodetic control, digitized their base mapping, and branched out into new digital products, including in the case of the USGS ortho-imagery at 1:12,000 (the scale of this digital product is defined by its positional accuracy, which is equivalent to that required of a paper map at 1:12,000 by the National Map Accuracy Standards).

Critical social theorists, particularly geographers, have written extensively about the inherent assumptions of this phase and its links to warfare and colonialism (Harley 2001). The notion that a single, powerful central authority could produce an objective representation of the Earth's surface grounded in scientific measurement runs counter to postmodern thinking and marginalizes the less powerful who may have alternative perspectives on geographic reality (Craig, Harris, and Weiner 2002). Kwan (2002) has presented a feminist perspective, and Pickles (1995) has compiled numerous other arguments along parallel lines.

In this phase the dominance of the NMA led to a simple radial model of data sharing, replacing the haphazard network model of the historic phase. Most GI emanated from the NMA, which was responsible for ensuring that it was as widely available and accessible as possible, at the lowest possible cost. Since the NMA's funding was largely derived through appropriation of national funds, there was no need to recover cost from consumers, who could therefore be charged a fee designed to cover no more than the cost of reproduction and distribution.

The contemporary phase began in the late twentieth century, driven by several evidently unsustainable aspects of the enlightened model. First, many governments came under increasing pressure to rein in the growth of expenditures, and began to question the basis on which NMAs were funded. In countries with right-wing governments there was pressure to reduce taxation, to reduce the size of government, and to turn over as many activities to the private sector as possible on the grounds that it was able to perform them more cheaply and efficiently. Many national governments argued that only a small proportion of the population benefited directly from the activities of the NMA, and that the beneficiaries should therefore be made to pay for GI at levels that would recover at least part of the costs of GI production (for a more detailed discussion of these arguments see National Research Council 2004; and for a detailed discussion of the economics of information production and consumption see, e.g., Shapiro and Varian 1999). At the same time the potential market for GI had grown, and more and more private companies found economic success in repackaging and adding value to government GI products that they obtained at minimal cost. Surely, it was argued, they could and should be made to pay part of the cost of government GI production.

Second, the fixed costs of GI production declined dramatically with the widespread adoption of information and communications technology (ICT) in the second half of the twentieth century. Expensive electromechanical photogrammetric stations were replaced by software running on inexpensive, general-purpose desktop computers; the costs of compilation and editing fell dramatically when these operations were moved to computers; and a series of technological innovations were made in surveying (e.g., the Global Positioning System and "total stations"). No longer were large organizations the only ones with sufficient resources to undertake the production of GI; small local governments and even farmers were increasingly able to afford the new, much-lower cost of entry. The radial outflow of GI from the NMA and other national agencies even began to reverse in some instances, as the U.S. Natural Resource Conservation Service began to receive soil data from farmers, and as the U.S. Bureau of the Census began to receive digital updates of street maps from local governments and the private sector. GI production was becoming a distributed enterprise.

Third, several factors undermined the concept of the perfect map (a theme anticipated to some extent by Carroll 1894 and Eco 1994). The enlightened phase had focused on aspects of the Earth's surface that were relatively static, in the interests of producing GI that would find as many uses as possible and be valid for as long as possible, thereby justifying its high costs. But ICT was opening vast new opportunities for acquiring data on movement, change, and other dynamic aspects of geography, and increasing attention was being focused on determining the geographic locations of events, transactions, and other transitory phenomena (Peuquet 2002), such as weather and traffic reports. It was also becoming clear that a steady progression to larger and larger scales (and greater and greater positional accuracy) was technically unachievable because of the inherent problems of partial update of GI (Goodchild 2002), and that instead each new coverage of greater positional accuracy would have to be developed from scratch. Small-scale data were perfectly adequate for many applications, and the difficulties of automating the generalization process (Müller, Lagrange, and Weibel 1995) meant that they could rarely be derived from large-scale data. Moreover, periodic changes to the geodetic datum, tectonic movements, and the errors inherent in measurement instruments all help to ensure that perfect accuracy in GI is in reality unattainable.

The realities of the contemporary phase in the United States were described in the early 1990s by the National Research Council's Mapping Science Committee (1993), as well as by certain prescient USGS employees, and led to the establishment of the National Spatial Data Infrastructure (NSDI) by Executive Order 12906 in 1994. The NSDI model has been widely adopted as the basis for spatial data infrastructures in many other countries, and at the global level. The contemporary phase involves a mixed market, with NMAs still playing large roles as GI providers and developers of standards, but with state and local governments and private businesses taking larger and larger roles in GI creation and distribution. Just as the federally funded space program eventually led to a market for private satellites, so nationally funded NMAs have opened a large and complex GI market. The contemporary phase replaces central GI production and uniform coverage by a vision of partnerships in which a network of producers creates and maintains a patchwork coverage. The scale and currency of each part of the patchwork is determined by local demand and funding; standards are in place to ensure that the various parts of the patchwork interoperate; and special processes are available to address problems that occur at the boundaries between parts of the patchwork, or between overlapping parts.

The NSDI recognizes the value of unifying the functions of production and providership, and of producing data as close as possible to the subject area *S* (for a discussion of the contemporary geography of GI production and dissemination, see Goodchild 1997). A system of geoportals takes advantage of recent developments in digital libraries and computer communications to provide points of entry to a distributed network of providers. In principle, a single geoportal would recreate for the user the old central dissemination system of the enlightened phase, though without its underlying notion of a single, authoritative set of GIObjects. Moreover, a suitably designed geoportal could operationalize the various conditions attached to the dissemination of each GIObject: open access, access restricted to certain users, license, sale, and so forth. A recent report (National Research Council 2004; see also Onsrud et al. 2004) provides detailed discussions of dissemination conditions and the arguments for and against each of them, but such issues are beyond the scope of this article.

In practice, however, there are still substantial problems associated with search for GI over a distributed network (Goodchild and Zhou 2003) unless a single geoportal can be built to access the entire universe of GI. This is the principle behind the GOS, a project that is intended to provide a single point of entry for all U.S. GI. As such it provides a window into the current world of GI sharing, and the use of ICT to operationalize the reality represented by the contemporary phase. In the next section we describe the GOS and its major technical features, against a backdrop of continuing and rapid technological change that renders most designs quickly obsolete. Analysis of the strengths and weaknesses of GOS helps us to understand the current state of data sharing, and to chart a course for the future.

The Geospatial One-Stop

Design

By providing a single point of entry, GOS proposes to solve the primary problem faced by a user of GI in the contemporary world: because there are many providers, some means must be found for searching among them for a GIObject that meets the user's needs. GOS proposes to do this by channeling all searches through a single geoportal containing a single catalog. Of course users of the Internet face the search problem constantly, and normally resolve it by employing one of the standard search tools, such as Google, Yahoo, or Altavista, which successfully act as single search portals. Such tools allow the user to search a catalog maintained at the tool's site, and created by periodically visiting as many Web pages as possible and analyzing them to extract important keywords. Thus a user looking for a recipe for Pisco Sour, a South American cocktail, would quickly find a number of Web pages listed in the search tool's catalog in response to the three words "Pisco Sour recipe" (Google's database provided 92,200 hits in response to this search in January 2006). Any of these pages can then be displayed and examined by the user by exploiting a link between the catalog entries and the respective pages located elsewhere on the Internet.

Unfortunately all of the search tools rely on keywords found in text, and thus are unsuitable for searching for GIObjects, which tend to employ highly specific coding schemes and are thus largely unreadable by the Web-crawling virtual robots that build search catalogs for sites such as Google. Moreover, even a text description of a GIObject would be problematic, since there is no universally agreed-upon vocabulary of GI themes. Placenames are frequently ambiguous as textual descriptions of location, and fail to capture the adjacency and containment relations that pervade GI (e.g., someone searching for a GIObject of the "Sierra Nevada" or the

"Western United States" would not likely be led to a map described as "California," even though such a map might well satisfy the user's need), and the coordinates used by GIObjects to identify locations would be unintelligible to text-based search engines, even if they could be decoded from their GIObject-specific formats.

The catalog of GOS uses the Federal Geographic Data Committee (FGDC) Content Standard for Digital Geospatial Metadata (CSDGM; www.fgdc.gov/metadata/contstan.html), a standard for GIObject description introduced in 1994 as part of the effort to build NSDI. Entries in the catalog are supplied by providers who send catalog entries to GOS in order to *publish* their GIObjects to it, by making them visible to users of GOS. Publishers may use one of a number of accepted formats including CSDGM. Like Google, each entry includes a URL (Universal Resource Locator), a Web address that points GOS to the GIObject itself at the provider's site, though unlike Google GOS requires an explicit action of publication by the provider since it is unable to locate GIObjects automatically. The GOS project is committed to encouraging sharing of GI, and has therefore been designed to make it as easy as possible for providers to publish their GIObjects to it. By implementing various forms of access restriction and mechanisms for electronic commerce, the system is designed to support publication and dissemination of GIObjects by commercial vendors.

The architecture of GOS links users to data providers through a Web-based geoportal (Figure 1). The user of GOS may employ a simple Web browser (*thin client*) to search for GIObjects and to display their contents, or may interface GOS directly with his or her own geographic information system (GIS), an example of a *thick client*. The GOS site includes its catalog, driven by a metadata clearinghouse to which publishers contribute records. It also includes a gazetteer, defined as an index that converts between place-names, the informal means of specifying location, and coordinates, their formal equivalent. In the case of GOS the core gazetteer service is based on place-names from the Geographic Names Information System (GNIS; geonames.usgs.gov), a service maintained by the USGS that includes the place-names recorded on its topographic maps. Additional more-specialized gazetteers can be implemented to recognize such domain-specific place-names as hydrographic unit codes or FIPS codes (Federal Information Processing Standard codes for administrative and census reporting zones). By incorporating a gazetteer GOS supports three ways of specifying the subject area S of a search (Figure 2): by bounding coordinates, by interacting with a digital map presented by GOS, or by specifying a place-name or address. In practice, the vast majority of searches are specified by place-name (Mapping Science Committee 1999).

Depicted at the bottom of Figure 1 are the various servers maintained by providers who have published data to GOS. These can be in any of a wide variety of formats, including those defined by commercial developers of GIS software, or those established by standards promulgated by such agencies as the FGDC, the ISO (the Interna-

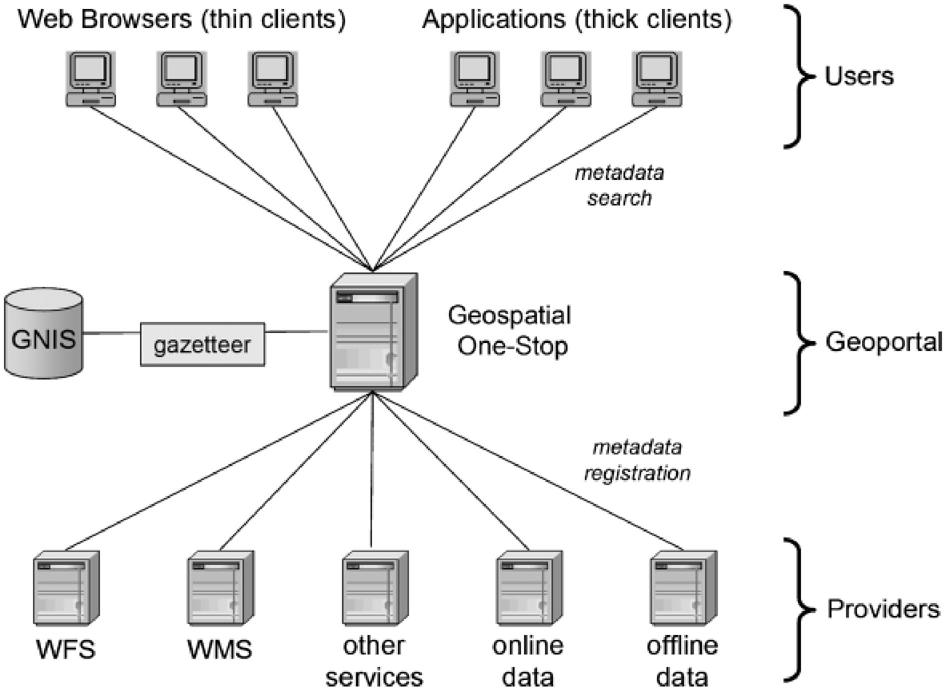
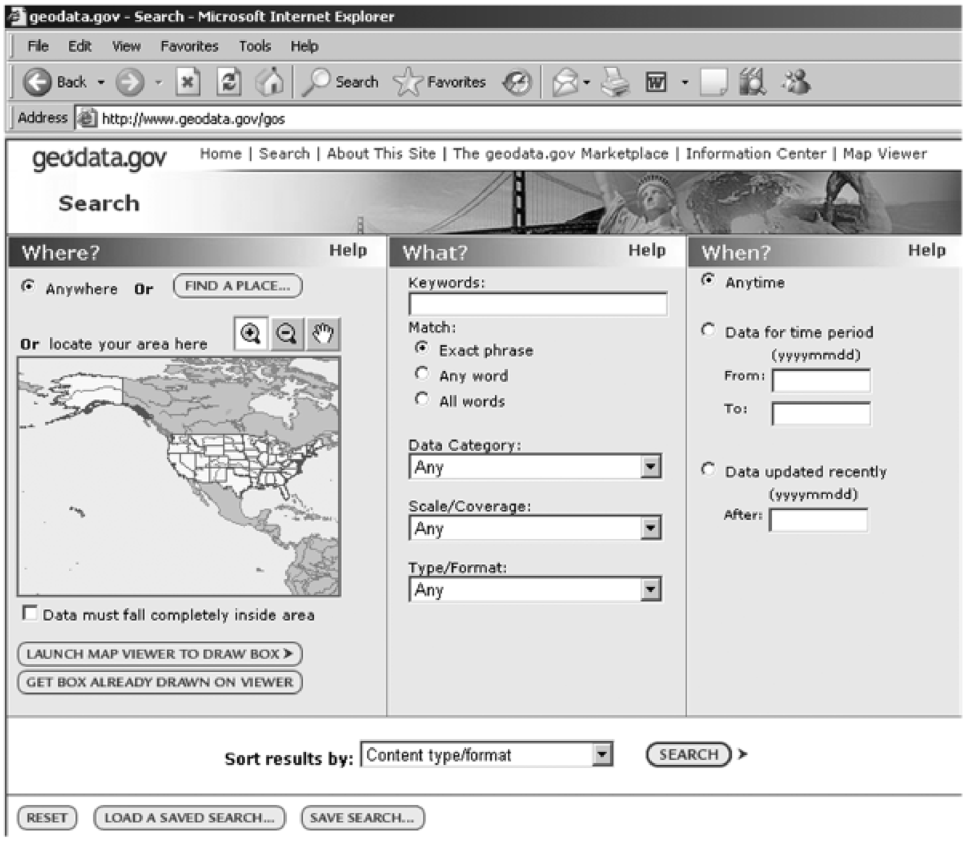


Figure 1. The architecture of GOS.



Address: http://www.geodata.gov/gos

● Anywhere Or

Match: ● Exact phrase

Data Category: Any

Scale/Coverage: Any

Type/Format: Any

● Anytime

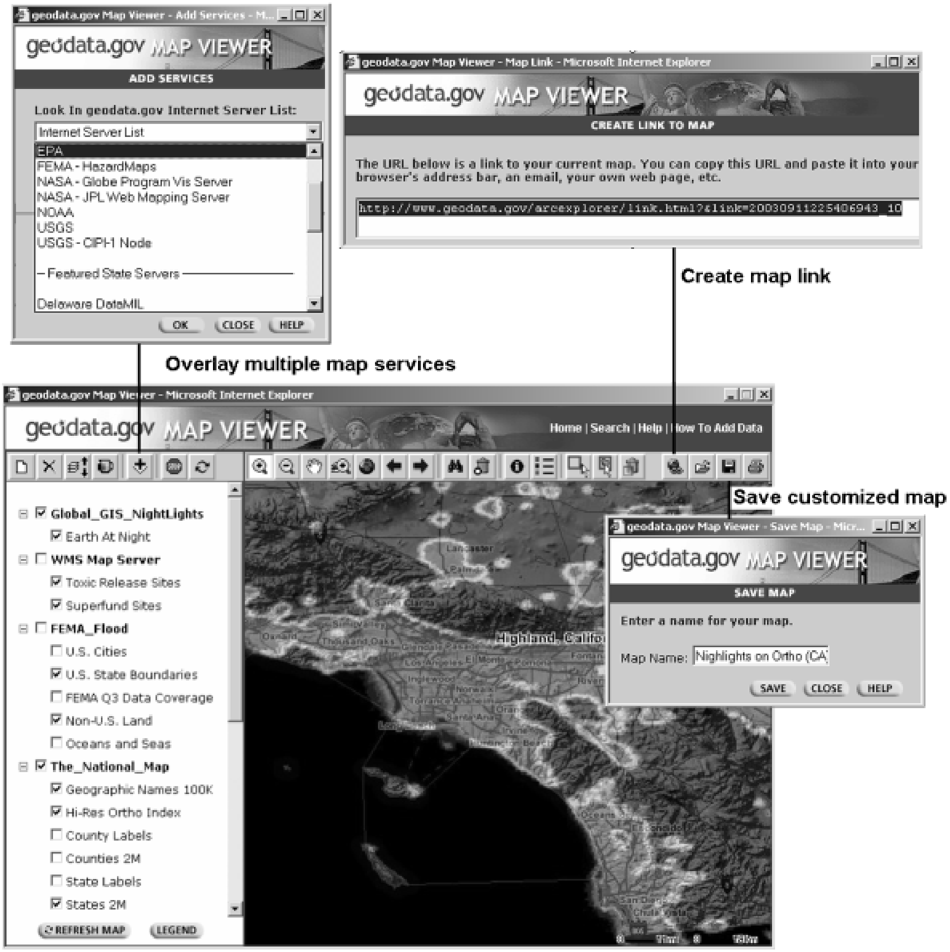
Sort results by: Content type/format

Figure 2. The GOS search screen, allowing the user to specify the properties needed, including geographic location (left), scale, theme (center), and date (right).

tional Standards Organization), or the OGC (the Open Geospatial Consortium; www.opengeospatial.org). GOS is designed both to act as a circulating library, by allowing users to extract data sets for use in their own GISs (thick clients), and as a reading library whose users are able to open and display the contents of GIObjects using thin clients (Figure 3). In the first case GOS acts merely as a package handler, but in the second it must achieve a certain level of interoperability by knowing how to open, access, and display the contents of GIObjects that may be in any of a large number of recognizable formats (see Albrecht 1999 for a historically based review of efforts to develop geospatial information standards). This interoperability is achieved through the use of three OGC specifications: WFS (Web Feature Service) for GIObjects that consist of collections of digitized features; WMS (Web Map Service) for GIObjects that represent compiled maps; and WCS (Web Coverage Service) for GIObjects representing fields, typically in one of the common raster formats.

In addition to supporting search, mapping, and publishing, GOS includes *administrative* services that implement various management functions. Figure 4 shows the overall design of the GOS software and the services it provides. Most significant of these is the function of *gatekeeper*, implemented in the Metadata Validation Service, to provide a degree of quality control over the contents of the GOS catalog. Just as a traditional library employs staff to exert control and direction over its collection of books, journals, and monographs, so GOS involves administrative staff in determining whether or not to accept GIObjects offered to it by their providers. To the user, quality is thus ensured through two overlapping mechanisms: the gatekeeper function, which allows the managers of GOS to determine whether to accept or reject a GIObject; and the brand and authenticity inherited by a GIObject from its providers (e.g., a GIObject published by the USGS inherits the level of quality traditionally associated with that agency). GOS will presumably build its own brand image over time.

The GOS platform provides numerous opportunities to devise and implement other useful services. An update service is available, allowing users to be notified automatically when certain GIObjects of interest are updated by their providers, or when new GIObjects are published that meet user-specified criteria. As a geoportal, GOS is also designed as a mechanism for publishing the existence of geographic information services (GIServices), which are defined as remotely invokable processing services. For example, a commercial vendor of a *geocoding* service for converting street addresses into



The URL below is a link to your current map. You can copy this URL and paste it into your browser's address bar, an email, your own web page, etc.

`http://www.geodata.gov/arcexplorer/link.html?&link=20030911225406943_10`

Enter a name for your map.

Map Name: Nighlights on Ortho (CA

Figure 3. Simple display and GIS functions available with the GOS map viewer, using a thin client (Microsoft's Internet Explorer).

coordinates might publish the existence of the service through GOS. A user discovering the existence of such a service through GOS might send it lists of street addresses, receiving in return lists of corresponding coordinates.

Building the Catalog

Tools such as Google owe their success in part to the fact that Web pages can be loaded into their catalogs without any action on the part of page providers (though an entire industry has emerged dedicated to making pages more likely to be ranked highly in search results). As we saw earlier, it is impossible for GOS to emulate these tools given the nature of GIObjects, and instead it must rely on publication actions by providers, and on the willingness of providers to take such actions. Over the past decade the FGDC has supported several programs designed to encourage publication actions by providers, with mixed success (Mapping Science Committee 2001). GOS currently offers four publication mechanisms, hoping that providers will find at least one of them attractive:

- *An on-line Web form* (Figure 5), designed for less-experienced users who are unfamiliar with metadata and standards such as XML (eXtensible Markup Language, a common format for metadata). The contents of the form can be edited online, validated by automated procedures, and submitted to the catalog.
- *Direct upload of XML-formatted metadata*, the option preferred by experienced users. An on-line Web interface is provided for this purpose.
- *Upload from desktop GIS*, an option designed for GIS professionals. GIS software environments oriented to data management, such as Environmental Systems Research Institute's ArcCatalog, provide suitable tools for sharing metadata with remote servers such as GOS.
- *Harvesting*, an option designed for providers of large collections of GIObjects that are already

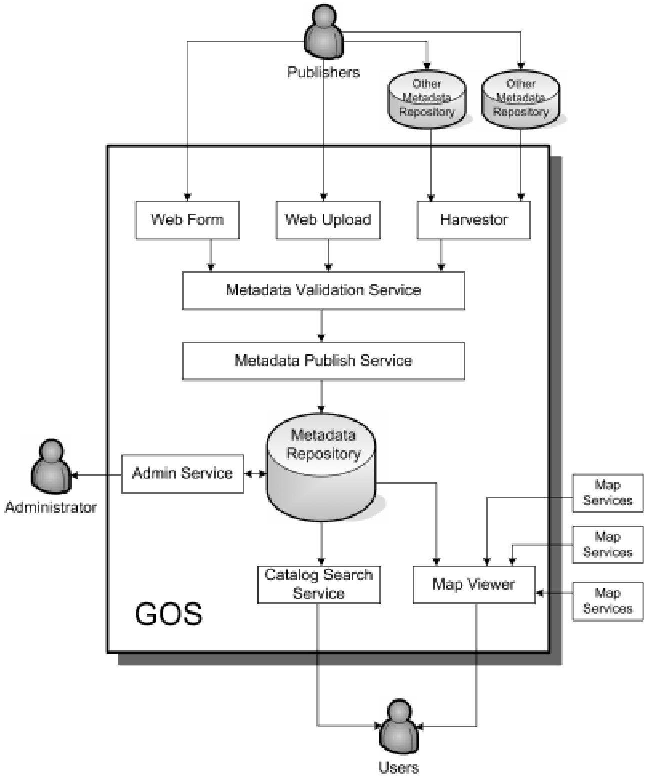
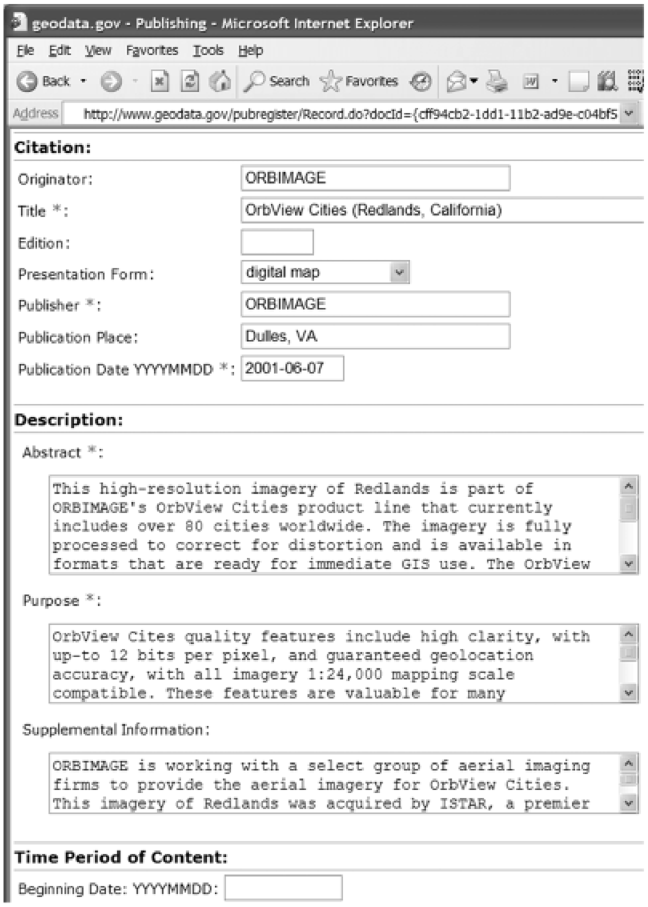


Figure 4. The software architecture of GOS.

documented with metadata. A provider who selects this option first registers his or her site with GOS. The GOS metadata harvester automatically connects to registered sites, retrieving all XML-formatted metadata on the first visit, or updates on subsequent visits. Quality is checked to ensure that each retrieved record is compliant with minimal requirements. GOS can harvest from metadata sites using five types of metadata protocols: Z39.50, ArcIMS, Web-Accessible Folder (WAF), Open Archive Initiative (OAI), and OGC Catalog Service for Web (CSW). Although this harvesting option brings some degree of automation to the publication process, it still requires the provider to take explicit action on an essentially voluntary basis.

Usage

By January 2006 GOS had grown to include records of some 10,400 GIObjects covering parts or all of the United States, whose existence had been published to it by one of some 800 providers and approved by GOS gatekeepers. The providers included government agencies, commercial companies, and academic institutions.



http://www.geodata.gov/pubregister/Record.do?docId={cff94cb2-1dd1-11b2-ad9e-c04bf5

Citation:

Originator: ORBIMAGE
Title *: OrbView Cities (Redlands, California)
Edition:
Presentation Form: digital map
Publisher *: ORBIMAGE
Publication Place: Dulles, VA
Publication Date YYYYMMDD *: 2001-06-07

Description:

Abstract *:

```
This high-resolution imagery of Redlands is part of
ORBIMAGE's OrbView Cities product line that currently
includes over 80 cities worldwide. The imagery is fully
processed to correct for distortion and is available in
formats that are ready for immediate GIS use. The OrbView
```

Purpose *:

```
OrbView Cites quality features include high clarity, with
up-to 12 bits per pixel, and guaranteed geolocation
accuracy, with all imagery 1:24,000 mapping scale
compatible. These features are valuable for many
```

Supplemental Information:

```
ORBIMAGE is working with a select group of aerial imaging
firms to provide the aerial imagery for OrbView Cities.
This imagery of Redlands was acquired by ISTAR, a premier
```

Time Period of Content:

Beginning Date: YYYYMMDD:

Figure 5. The metadata form designed for less-experienced users.

Figure 6 shows the footprints of these GIObjects over the United States and reveals an uneven pattern of coverage, with comparative abundance in the Mid-Atlantic, Rocky Mountain, and West Coast states. GOS coverage is mostly limited to the geographic extent of the United States, though a small number of GIObjects are included because of their relevance to U.S. international relief efforts. Broadly, the pattern reveals the influence of three factors: the availability of data, the level of participation in GOS, and specific activities such as Hurricane Katrina relief efforts, which explain the relatively heavy coverage of New Orleans.

Goodchild and Zhou (2003) have analyzed the much larger collection of footprints of the GIObjects available through the Alexandria Digital Library at the University of California, Santa Barbara (www.alexandria.ucsb.edu), and have found a tendency for the density of coverage of a subject area *S* to decline with distance from the site of the library *D*. As a national geoportal, albeit located in Washington, D.C., one would not expect such a decline with distance within the United States in the case of

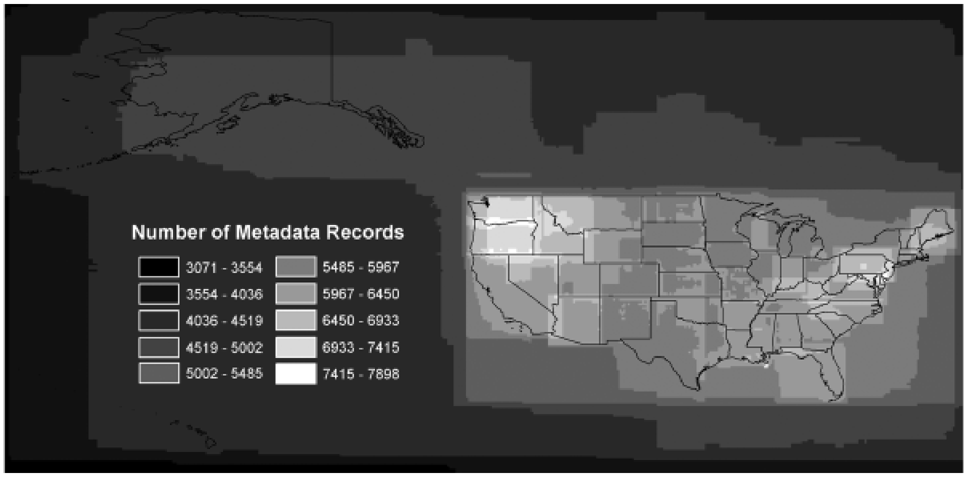


Figure 6. Footprints of the approximately 10,400 entries in the GOS catalog in January 2006. Shading denotes the number of GIObjects overlapping each 0.25 degree by 0.25 degree cell.

GOS; rather, the pattern of footprints reflects the distribution of interests and awareness of the agencies, companies, and institutions that have published to GOS, which in turn tend to reflect the basic distribution of population. We return to this issue in a later section where we discuss the relationship between GOS and other geoportals and geolibraries.

Based on statistics for a typical week (2–8 January 2006), GOS receives on average 71,452 hits per day. An average of 1,678 extended visits were recorded per day, of an average length of 28.5 minutes. During this period, 7,003 users visited, 5,733 of them only once, and 1,270 more than once. The most-queried themes are dominated by the so-called *framework* layers, the types of GIObjects that are used as spatial referencing for many purposes, and correspond to the traditional products of the NMA (Figure 7). Clearly users look to GOS for services they traditionally received from the Federal Government during the enlightened phase, for despite the best efforts of its promoters to include the broad range of GI providers, at present GOS is perceived primarily as an outlet for government data.

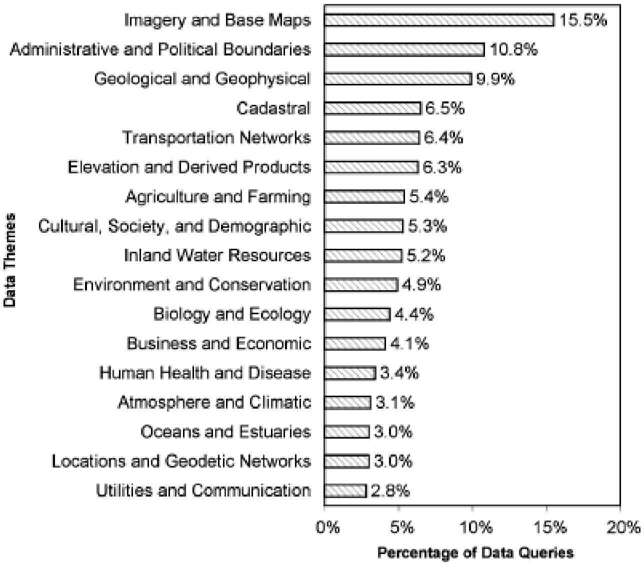


Figure 7. Most-queried GOS themes in January 2005.

Assessment

The most significant technical advances of the GOS project (and other geoportals of which it is a representative example) compared to its predecessors (e.g., the National Geospatial Data Clearinghouse, a central feature of the NSDI mandated in 1994 by Executive Order 12906) lie in its ability to present a single point of access to distributed resources, through a common metadata catalog and pointers to individual GIObjects, and in its harvesting capability, which enables automatic metadata flow from registered sites to the central GOS database. Currently, more than 95 percent of the metadata records in the GOS database were collected via harvesting. Because of legal and privacy concerns, harvesting is currently limited to registered sites.

Harvesting brings new challenges. Because there is not a unique and universal identifier for GIObjects, a metadata record may be harvested more than once if some aspect of it is changed by the provider or if the same GIObject exists in multiple registered sites. To help to avoid this duplication problem, GOS adds to each harvested metadata record a "sourceuri" attribute, in effect providing its own unique identifier. Harvesting also makes the role of the human gatekeeper more challenging, since thousands of records can be harvested within a few hours.

In January 2006, more than 3,900 GIObjects accessible through GOS were *live data*—that is, GIObjects that can be processed directly, without downloading, either through the GOS server's map viewer or through the user's own GIS. Live data can be mapped, projected, or queried without the need to first download a copy, and live data can be combined by the user from multiple remote sources, all accessed through the GOS geoportal.

As its name implies, GOS is intended to provide a single geoportal to all GI, and a rational response to the need to find GIObjects in the distributed world of the contemporary phase. As such, it has the potential to provide one of the key elements of the NSDI, and an effective advance on its predecessors. In principle, GI includes all information that satisfies the formal requirement that it be decomposable into tuples of the form $<\mathbf{x},z>$. This would include all georeferenced imagery collected from Earth-observing satellites, which alone amounts to several terabytes per day, and is available through NASA's EOSDIS (the Earth Observing System Data and Information System) and its DAACs (Distributed Active Archive Centers), as well as other sources. In practice, however, the NSDI and such related projects as GOS tend to be oriented toward other types of data, particularly digitized maps and other GI products of the agencies that participate strongly in the FGDC.

Such subtleties are part of the environment in which users must operate, and are well known to experienced GI professionals. Over the past decade, large archives, clearinghouses, and geoportals have been organized along many lines: by institutions that house large library collections, and have moved to providing access to them via the Internet (e.g., UCSB's Alexandria Digital Library); by organizations representing various domains of interest, based on geographic regions (e.g., the Antarctic Data Directory System of the Joint Committee on Antarctic Data Management, www.jcadm.scar.org/adds.html) or fields of scientific research (e.g., the Global Change Master Directory, gcmd.gsfc.nasa.gov); or by different levels of government (e.g., the Minnesota Geographic Data Clearinghouse, www.lmic.state.mn.us/chouse), which "serves as a 'One Stop' source for geospatial data"). In recent years, most states have instituted state-level GI clearinghouses. There are also instances of directories being maintained by commercial companies (e.g., ESRI's Geography Network, www.geographynetwork.com, the prototype technology for GOS), and of sites that include substantial amounts of GI even though their domain is much broader.

From the perspective of the potential user, therefore, there is always more than one "One Stop" source for virtually any GI requirement. An experienced user may be able to navigate through this complex landscape, but, as the potential sources proliferate and as their catalogs grow, navigation is becoming more rather than less difficult. In an ideal world one source would come to dominate by harvesting catalogs from all other sources, but this would require almost unlimited resources for gatekeeping, or its abandonment, and would require the collaboration of every provider. From the perspective of a provider, only limited time is available to identify and publish to other sites, and this time must come at the expense of other activities. In the world of conventional search tools it may be that one tool will come to dominate all others, and Google is already dominant in some areas, but there is an essential difference in the GI world, where action is required by the provider, and where gatekeeping in the interests of quality control is an established component of many sites. Moreover, many GIObjects are global, and many others produced by U.S. agencies extend outside the borders of the United States, into Canada and Mexico at least, so if one source is to dominate it will have to be a global source, not a source such as GOS whose catalog is clearly focused on one country.

We conclude, therefore, that though GOS represents a significant step forward, its goal to be a one-stop source for GI is unattainable in the present context. Several solutions are available, however, that would move GOS closer to its goals. First, progress could be made in further automating the harvesting process by developing digital agents that would identify and copy catalogs of GIObject metadata, without intervention by providers. As noted earlier, this is currently impeded by the constraints under which a government-sponsored geoportal must operate. It is difficult to imagine harvesting by GOS providing a complete solution to this problem, given the wide variety of metadata and catalog formats, but even partial automation would greatly improve the effectiveness of GOS. Technologies such as Webinator (www.thunderstone.com) allow specialized agents to be developed that will look for particular characteristics of relevant spatial data archives, clearinghouses, and digital libraries. Some accommodation would be necessary in the gatekeeping process, perhaps by making decisions at the level of collections rather than individual GIObjects, or by making appropriate disclaimers.

Second, sources could be organized along clearly demarcated divisions of discipline, region, or level of government, such that clear criteria were available to users for selecting the most appropriate one. In essence, this proposes the construction of *collection-level* metadata (CLM), consisting of formal descriptions of the contents of each source's catalog. Some progress has been made in developing specifications for CLM (Hill et al. 1999), but in practice the problem is at least as difficult as the traditional *object-level* metadata problem.[3] Moreover, this solution would require a massive initial phase of trading, as providers swapped GIObjects to reflect the agreed divisions.

Finally, a dominant GOS would necessarily be a global GOS, which would conflict with its origins as an effort by the U.S. Federal Government to improve the mechanisms it uses to disseminate and share data about the United States. In this environment, it seems that the most appropriate solution would be for GOS to establish its own domain unilaterally, and to be consistent in building its catalog to cover that domain. As it is currently most closely associated with the U.S. Federal Government, and particularly with the agencies actively participating in the FGDC, it could make sense for it to define its domain as delimited by the borders of the United States (with overlaps into bordering countries where necessary), and by the products of those participating agencies. Alternatively, GOS could expand to a truly global scale in collaboration with appropriate international partners. Since the most obvious element of CLM is the name of the site, it would be appropriate for that name to reflect the chosen domain as accurately as possible.

An important question for any system such as GOS is whether it *scales*, whether its services will be provided as effectively when its catalog is 10, 100, or 1,000 times as large as the current catalog. Users expect the response time to queries to be constant, and are likely to be impatient if response time grows with the size of the collection. One of the techniques used by search tools such as Google to avoid problems as the number of catalogued pages grows is to pay careful attention to how hits are ranked, since users rarely examine more than the first few hits to queries that may generate thousands. Currently, GOS search results do provide the "sort by relevance" choice, but relevance is calculated by the presence of words in a subset of the metadata, including title, abstract, purpose, keywords, amd so forth. Ranking of GOS results is a complex issue, given the overlapping nature of many search characteristics, especially the all-important geographic coverage, theme, and scale. How, for example, should the system evaluate a GIObject with a scale of 1:100,000 when the user requested 1:24,000, or a GIObject with a footprint of all of the Western United States when the user requested California? The GOS catalog is currently approaching a size that will force these issues, and they will clearly become more and more important in the future. To provide rapid access to GIObjects of great topical importance, GOS already allows the user to bypass a full metadata search through the use of *data channels*, which have been implemented for such topics as the Indian Ocean disaster, or Hurricane Jean.

Prospects

Technical Issues

Geoportals in general, and the GOS geoportal in particular, provide a mechanism to connect GI providers and users, and as such are expected to perform an important role in the emerging GI marketplace, as well as in the practice of academic research, particularly by geographers. Already GOS has proved to be important for locating and accessing government data, in part due to strong promotion by the government itself, and because of the active participation of various key government departments, including the USGS and other members of the FGDC. Other portals of similar design are playing increasingly important roles as mechanisms for the distribution of research data, notably the products of remote sensing.

It is still too early to assess how important GOS will prove to be to private GI providers. Though many of the necessary provisions are in place, to date GOS has not been widely promoted and widely recognized as a marketplace for GI commerce. Many private GI providers cater to highly specialized niche markets, and may choose other means to connect with their customers.

GOS can best be viewed as a step in the evolution of GI sharing; there are already plans for major changes, and its eventual significance will depend on a complex set of economic, political, and historic factors that have not yet played out. As a step in the evolution of GI sharing, the GOS geoportal serves as a first approximation to a Web-based national exchange and marketplace for GI that at a minimum will prompt the design of better ways to share both governmental and private data, and may yet prove to be the primary connection between highly decentralized U.S. communities of GI providers and users.

The need for geoportals as a mechanism for data sharing has developed against the backdrop of rapidly changing technology, and a growing focus on cyberinfrastructure, or a shared computing environment. At the same time GI capabilities of individual organizations, initially built to meet project needs, have focused increasingly on enterprise design, with investment in a common infrastructure of value to a broad set of projects and institutional functions. Geoportals contribute to the goal of building common infrastructure to share GI more effectively, providing a needed link between diverse providers and users. Yet despite a generation of technological developments, each of which originally offered to simplify the process of searching for GI, to increase sharing, and to reduce duplication, the landscape of GI archives, clearinghouses, and digital libraries is as complex as ever. GOS reflects the cutting edge of geoportal technology, and a significant improvement over its predecessors. We argue nevertheless that it cannot deliver on its promise to provide a true one-stop source of all GI. Instead, it will best succeed if it focuses its attention on a clearly defined domain and becomes known as the dominant source for data in that domain.

Significant developments in automated harvesting of GI metadata are likely in the next few years. But it is already clear that they can only succeed in the context of sites like GOS at the expense of a weakening of the gatekeeper role, and thus of the site's assurance of quality, and its replacement by a strategy of *caveat emptor*. The same dilemma faces the traditional search tools, of course, forcing users to resort to heuristics to determine the truth (e.g., a search of Google in January 2006 returned 9,490,000 hits for the misspelled term "accomodate" versus 140,000,000 for the correct spelling, so a heuristic based on relative frequency would give the correct spelling). In the specific case of GOS, if content were limited to Federal sources perhaps the gatekeeper role could be abandoned in favor of relying on the source agency's explicit or implied quality assurance. But this approach clearly will not work for other non-Federal sources.

GOS relies on the FGDC's CSDGM standard to define its data catalog, and to support search, assessment, and downloading. Yet the FGDC standard and other standards that have been based on it are limited in several respects. Much of the research on spatial data quality and uncertainty of the past two decades has yet to influence the standard, which adopted the five dimensions of data quality devised in the 1980s. The standard describes only the properties of single, discrete data sets, and needs to be adapted to the complex object-oriented designs of contemporary spatial databases. The ability of pairs of data sets to interoperate would also be a useful addition to metadata. Schuurman (2005) has discussed other currently missing dimensions of geospatial metadata.

One of the traditional library's most valuable functions is its support of browsing. By organizing books on similar topics close together in the library, it is possible for a user to scan numerous sources without explicitly requesting them through a formal catalog. Web sites achieve similar ends by devices such as "people who read . . . also read . . ." based on previous patterns of site use or purchase. In essence, a formal Boolean search of a catalog followed by a listing of hits fails to emulate many of the functions that users find valuable in traditional information-organization environments, whether they be designed to serve specialized research communities or the general public. We have already commented on the importance of ranking, particularly as the catalog grows. Shneiderman and Plaisant (2005) have written about the importance of browse, and of letting users move rapidly between a holistic view of a collection and a more focused view (e.g., compare Figure 6 with Figure 2).

Cognitive Issues

The GOS user interface is designed for users who are familiar with the basic concepts of GI, such as scale, and with conventional practices concerning themes or layers; in other words, users are assumed to be *spatially aware professionals* (Longley et al. 2005). This is a reasonable approach if one assumes that having identified and downloaded suitable data, the user will then have to navigate through the user interface of a GIS to obtain useful results. Increasingly, however, GI is being accessed by users who lack this level of professional familiarity, and expect the kind of simplicity found in the interfaces of Google Earth or MapQuest. Such interfaces avoid any reference to technical concepts such as scale, instead allowing the user to manipulate sliders and other simple devices. Moreover the technical means now exist to access GOS data through such interfaces.

More generally, the problem of search for GI is something humans have addressed for centuries, with the temporary exception of the enlightened phase. They have adopted heuristic procedures, including conversation with other experts, to help them to work through a complex world to identify suitable sources. Conversations are often iterative, moving from poorly expressed general needs to more-exact specifications of requirements as the user's needs and the availability of information slowly converge. Against this informal but in the long run successful strategy, the simple Boolean models of search adopted by such sites as GOS seem simplistic. In the future, improvements may have to come from a greater effort to model the search process on human experience, with all its fuzziness and informality. Techniques for visualizing large information spaces (Fabrikant and Buttenfield 2001; Shneiderman and Plaisant 2005) will be increasingly useful as the collection grows and as demands are placed on GOS by users with less professional expertise.

Security and Privacy Issues

In a previous section we presented several arguments for and against the sharing of GI. Underlying those arguments was the implicit, altruistic assumption that all

GI is ethically neutral, and that sharing should be driven purely by issues of cost and utility. Despite the broad sweep of its stated objectives, GOS is at heart a government-sponsored project that is intended to improve access to the data assets of government.

We argued that in practice the domain of GI is much smaller than a rigorous definition would suggest. Smith (1992) has presented compelling evidence of another aspect of this conventional narrowing: the apparent unwillingness to acknowledge the military and intelligence value of GI. From a sharing perspective, issues of national security add another dimension to the debate that rapidly became more important, at least in the United States, following the events of 11 September 2001. A report of the Rand Corporation (Baker et al. 2004) has argued that only a very small fraction of the GI that is publicly available via the Internet can in any sense be regarded as compromising national security; nevertheless the issue keeps surfacing, as it did more recently in the case of Google Earth's coverage of India (*New York Times* 2005), a country that has traditionally taken a very restrictive attitude to GI sharing.

The kinds of open sharing of GI that underlie GOS and comparable geoportals also raise questions of individual privacy. Curry (1998) has enumerated the threats to privacy represented by high-resolution GI and the record-linking capabilities of GIS. Government agencies such as the Bureau of the Census tend to take a very cautious attitude to the release of data that may compromise privacy, but the issue is clearly much more serious for geoportals that harvest metadata from nongovernment as well as government sources without effective gatekeeping.

Conclusions

Much of the recent literature on sharing geographic data has assumed that sharing is always desirable, but there are valid arguments both for and against that are significant both for GIScience and for the geographic enterprise in general. Strong scale economies have characterized traditional production methods, but may be diminishing as technology advances. Achieving shared understanding across information communities can incur substantial costs in the form of metadata investments, standards, and translators. Bureaucratic turf, the interests of researchers who collect primary data, and the different needs of users can also lead to scale diseconomies. Finally, in the private sector sharing runs counter to the competitive traditions of an open market.

We proposed a three-phase model for the history of geographic data sharing: an initial historic phase of networked dissemination, an enlightened phase of central production and radial dissemination, and a contemporary phase that emerged as the central-production model became unsustainable. Three factors appear to have contributed to that emergence: the attractiveness of a user-pays model to government agencies, the decline of fixed and marginal costs of GI production due to the widespread availability of new digital technologies, and a growing realization that the concept of perfect GI was unattainable.

A networked model, in which the GI necessary for some task or project might be found anywhere in a vast and distributed system of producers, warehouses, libraries, and portals, places the onus on the user to develop suitable strategies of search, and to evaluate fitness for use. GOS represents merely the latest in a series of steadily more sophisticated technologies that have been designed to address this set of issues. The nature of GI ensures that generic, all-purpose solutions to search over distributed resources, such as Google, are of very limited value. Instead, the GIScience community has devoted much of its effort over the past decade to developing special-purpose methods for search, for overcoming problems of interoperability, and for evaluating fitness for use.

Several features of GOS represent substantial advances over its predecessors: the ability to search through a single portal for GI that may be located anywhere, the ability to publish the existence of GI to the portal, the existence of a gatekeeper to help to ensure quality, and the ability to harvest catalog records from willing sources. At the same time, we have argued that GOS must ultimately fail in its mission to become a *single* source of GI. We have argued, starting from a rigorous definition of GI, that the domain currently covered by GOS is only a fraction of the full domain. Moreover, other communities may define their own domains in ways that only partially overlap our definition of GI—based on geographic, disciplinary, corporate, or administrative criteria. In our view there will always be more than one "One-Stop," and users will always be faced with the problem of knowing where to look.

We proposed several directions for further development that might lead to a next-generation GOS. Digital agents could be developed that would automatically identify and catalog GIObjects. A concerted approach to the problem of CLM would simplify the user's task of knowing where to look, and we proposed several alternatives. Finally, a systematic approach to ranking will be needed as the GOS collection grows.

In the longer term, it seems that the kinds of Boolean search strategies supported by resources such as GOS must ultimately yield to a more human-centric approach that reflects the heuristics that users have developed over the centuries for coping with an essentially vague task. Human needs will rarely be defined precisely, and searches will rarely be simple matters of hit or miss. As the GOS collection expands, carefully designed strategies for ranking GIObjects will be needed, and users will need to be able to see the collection from various perspectives that go well beyond the simple unsorted list. Such issues present a major research opportunity for GIScience.

Finally, we note that the GI sharing strategy represented by GOS is based on a fixed concept of data granularity, the GIObject. This emulates the traditional role of the library, which ended when the fixed granule—the book or journal issue—was in the hands of the user (earlier we called this a package-handling role). In principle the eventual use of a GIObject must be to provide answers to specific questions, whether through simple query or visualization, or through more advanced analysis, and such answers are traditionally provided by the user manipulating his or her own GIS. GOS already allows users to crop GIObjects through its live data services, based on OGC standards. In the long term we envisage a GOS that will offer a much larger range of GIServices (Goodchild 2004) that require simultaneous access to multiple or partial GIObjects, thus moving beyond the traditional role of the library and its fixed granularity to a truly distributed geographic information system. Such services will require the existence of tools to address the differences of meaning that commonly arise between disparate GIObjects. The efforts of the FGDC have succeeded in overcoming many of these across the U.S. Federal Government, but they remain prevalent in many other domains. GOS has recently implemented a thesaurus that allows searches to be generalized over terms with similar meaning, but more-formal ontological tools will clearly be needed if major progress is to be made.

Acknowledgments

This work was supported, in part, by the National Science Foundation (award BCS 0417131 to Goodchild), by Environmental Systems Research Institute (ESRI), the Department of Energy National Energy Technology Laboratory (DOE NETL), and the Los Alamos National Laboratory Earth and Environmental Science Division (LANL EES).

Notes

1. The term *object* is used in many ways in the information-technology and GIScience literatures. We use it here in the sense in which it is normally used in the digital library literature, to refer to a single cataloged item. We do not distinguish objects further, such as between data files and databases, and this use of the term has no relationship to object orientation or to the distinction made by some GIS vendors between objects and features.
2. Arguably leaving the title of "first complete large-scale map" to the Digital Chart of the World, also at 1:1,000,000.
3. The term *object-level metadata* is defined in the digital library literature to refer to metadata regarding a single cataloged item. Metadata as normally understood in the GI literature are object-level metadata.

References

Albrecht, J. 1999. Geospatial information standards: A comparative study of approaches to the standardization of geospatial information. *Computers and Geosciences* 25: 9–24.

Baker, J. C., B. Lachman, D. Frelinger, K. M. O'Connell, A. C. Hou, M. S. Tseng, D. T. Orletsky, and C. Yost. 2004. *Mapping the risks: Assessing the homeland security implications of publicly available geospatial information*. Santa Monica, CA: Rand Corporation. Available at www.rand.org/pubs/monographs/MG142/(last accessed 18 January 2006).

Carroll, L. 1894. *Sylvie and Bruno concluded*. New York: Macmillan.

Craig, W. J., T. M. Harris, and D. Weiner. 2002. *Community participation and geographic information systems*. New York: Taylor & Francis.

Crompvoets, J., and A. Bregt. 2003. World status of national spatial data clearinghouses. *URISA Journal* 15:43–50.

Crompvoets, J., A. Rajabifard, A. Bregt, and I. Williamson. 2004. Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science* 18 (7): 665–89.

Curry, M. R. 1998. *Digital places: Living with geographic information technologies*. New York: Routledge.

Eco, U. 1994. *How to travel with a salmon and other essays*, trans. William Weaver. New York: Harcourt Brace.

Fabrikant, S. I., and B. P. Buttenfield. 2001. Formalizing semantic spaces for information access. *Annals of the Association of American Geographers* 91:263–80.

Federal Geographic Data Committee. 1996. *Coordination and integration of wetland data for status and trends and inventory estimates*. Technical Report 2. Washington, DC: Department of the Interior.

Goodchild, M. F. 1997. Towards a geography of geographic information in a digital world. *Computers, Environment and Urban Systems* 21 (6): 377–91.

———. 2002. Measurement-based GIS. In *Spatial data quality*, ed. W. Shi, P. F. Fisher, and M. F. Goodchild, 5–17. New York: Taylor & Francis.

———. 2004. The Alexandria Digital Library: Review, assessment, and prospects. *D-Lib Magazine* 10(5). Available at www.dlib.org/dlib/may04/goodchild/05goodchild.html (last accessed 18 January 2006).

Goodchild, M. F., M. J. Egenhofer, R. Fegeas, and C. A. Kottman, eds. 1999. *Interoperating geographic information systems*. Boston: Kluwer Academic.

Goodchild, M. F., M. J. Egenhofer, K. K. Kemp, D. M. Mark, and E. Sheppard. 1999. Introduction to the Varenius project. *International Journal of Geographical Information Science* 13 (8): 731–45.

Goodchild, M. F., and S. Gopal. 1989. *Accuracy of spatial databases*. New York: Taylor & Francis.

Goodchild, M. F., and J. Zhou. 2003. Finding geographic information: Collection-level metadata. *GeoInformatica* 7 (2): 95–112.

Groot, R., and J. McLaughlin, eds. 2000. *Geospatial data infrastructure: Concepts, cases and good practice*. Oxford, U.K.: Oxford University Press.

Harley, J. B. 2001. *The new nature of maps: Essays in the history of cartography*, ed. P. Laxton. Baltimore: Johns Hopkins University Press.

Hill, L. L., G. Janée, R. Dolin, J. Frew, and M. Larsgaard. 1999. *Collection metadata solutions for digital library applications*. Santa Barbara, CA: Alexandria Digital Library, University of California.

Kline, K. D. 2004. A quantitative analysis of the users of global environmental data sets. Unpublished PhD diss., Department of Geography, University of California, Santa Barbara.

Kwan, M.-P. 2002. Feminist visualization: Re-envisioning GIS as a method in feminist geographic research. *Annals of the Association of American Geographers* 92 (4): 645–61.

Longley, P. A., M. F. Goodchild, D. J. Maguire, and D. W. Rhind. 2005. *Geographic information systems and science*, 2nd ed. New York: Wiley.

Maguire, D. J., and P. A. Longley. 2005. The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems* 29 (1): 3–14.

Mapping Science Committee, National Research Council. 1993. *Toward a coordinated spatial data infrastructure for the nation*. Washington, DC: National Academy Press.

———. 1999. *Distributed geolibraries: Spatial information resources*. Washington, DC: National Academy Press.

———. 2001. *National spatial data infrastructure partnership programs: Rethinking the focus*. Washington, DC: National Academy Press.

Masser, I. 1998. *Governments and geographic information*. New York: Taylor & Francis.

———. 1999. All shapes and sizes: The first generation of national spatial data infrastructures. *International Journal of Geographical Information Science* 13:67–84.

Müller, J.-C., J.-P. Lagrange, and R. Weibel, eds. 1995. *GIS and generalization: Methodology and practice*. London: Taylor & Francis.

National Research Council. 2004. *Licensing geographic data and services*. Washington, DC: National Academies Press. Available at books.nap.edu/catalog/11079.html (last accessed 18 January 2006).

New York Times. 2005 Governments tremble at Google's bird's-eye view. *New York Times*, 20 December.

Onsrud, H. J., G. Camara, J. Campbell, and N. S. Chakravarthy. 2004. Public commons of geographic data: Research and development challenges. In *Geographic information science*, ed. M. J. Egenhofer, C. Freksa, and H. J. Miller, 223–38. Lecture Notes in Computer Science no. 3234. Berlin: Springer-Verlag.

Pedley, M. S. 2005. *The commerce of cartography: Making and marketing maps in eighteenth-century France and England*. Chicago: University of Chicago Press.

Peuquet, D. J. 2002. *Representations of space and time*. New York: Guilford.

Pickles, J., ed. 1995. *Ground truth: The social implications of geographic information systems*. New York: Guilford.

Schuurman, N. 2005. Database ethnographies: Creating spatial metadata for informed decision-making. Paper presented to the Department of Geography Colloquium, University of California, Santa Barbara, 17 February.

Shapiro, C., and H. R. Varian. 1999. *Information rules: A strategic guide to the network economy.* Boston: Harvard Business School Press.

Shneiderman, B., and C. Plaisant. 2005. *Designing the user interface: Strategies for effective human-computer interaction*. Boston: Pearson/Addison Wesley.

Smith, N. 1992. Real wars, theory wars. *Progress in Human Geography* 16:257–71.

Thrower, N. J. W. 1996. *Maps and civilization: Cartography in culture and society.* Chicago: University of Chicago Press.

Williamson, I., A. Rajabifard, and M. E. Feeney. 2003. *Developing spatial data infrastructures: From concept to reality.* London: Taylor & Francis.

Wood, D., with J. Fels. 1992. *The power of maps*. New York: Guilford.

Correspondence: Department of Geography and National Center for Geographic Information and Analysis, University of California, Santa Barbara, CA 93106-4060, e-mail: good@geog.ucsb.edu (Goodchild); Environmental Systems Research Institute, 380 New York St., Redlands, CA 92373, e-mail: pfu@esri.com (Fu); Creekside Center for Earth Observation, 27 Bishop Lane, Menlo Park, CA, e-mail: paul.creekside@yahoo.com (Rich).