
# The GeoWeb

## LEARNING OBJECTIVES

After studying this chapter you will be familiar with:

- How the parts of GI systems can be distributed instead of centralized.
- Geoportals, and the standards and protocols that allow remotely stored data to be discovered and accessed.
- The technologies that support real-time acquisition and distribution of geographic information.
- The service-oriented architectures and mashups that combine interoperable GI services from different Web sites.
- The capabilities of mobile devices, including mobile phones and wearable computers.
- The concepts of augmented and virtual reality.

Traditionally, the only practical way to solve a problem using a geographic information (GI) system was to assemble all the necessary parts in one place, on the user's desktop or on a convenient server. But today all the parts—the data and the software—can be accessed remotely. The GeoWeb is a vision for the future, already substantially realized, in which all these parts are able to operate together, in effect turning the Internet into a massive GI system that is accessible anywhere, at any time, from any device. The GeoWeb vision also includes real-time feeds of data from sensors and the ability to integrate these with other data. The Cloud provides a versatile, powerful, economical, and largely transparent host for online GI systems and services. This chapter introduces the basic concepts of the GeoWeb, describes current capabilities, and looks to a future in which GI systems and services are increasingly mobile and available everywhere.

## 10.1 Introduction

Early computers were extremely expensive, forcing organizations like universities to provide computing services centrally, from a single site, and to require users to come to the computing center to access its services. As the cost of computers fell, from millions of dollars in the 1960s, to hundreds of thousands in the late 1970s, to less than a thousand, it became possible for departments, small groups, and finally individuals to own computers and to install them on their desktops. In the past five years desktop and mobile phone technologies have converged, allowing many advanced services to be offered through a smartphone powered by a central processor with the power and capacity of the super-computers of 20 years ago. Computers are being embedded in familiar devices such as car and entertainment systems, suggesting a future in which computing will be everywhere and to an increasing extent invisible to the user. To get there, however, will require extensive research, both on the technical issues and on the social implications. Box 10.1 describes the work of Yu Liu, a leading Chinese researcher in GI science, whose recent research illustrates the potential of the GeoWeb for gaining an improved understanding of human geography.

In Figure 1.12 we identified the six component parts of a GI system as its hardware, software, data, users, procedures, and network. This chapter describes how the network, to which almost all

## Biographical Box 10.1

### Yu Liu, leading GeoWeb researcher.

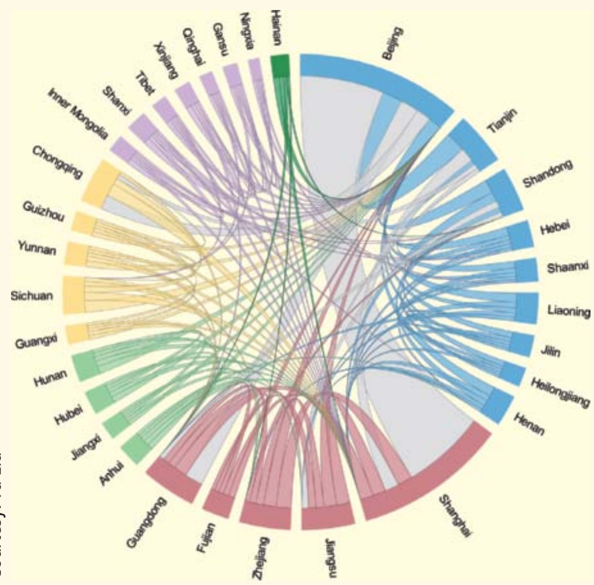
Courtesy: Yu Liu

**Figure 10.1** Yu Liu, Peking University.

Yu Liu (Figure 10.1) is professor of GI science and director of the Geosoft Lab at the Institute of Remote Sensing and Geographical Information Systems, Peking University. His research interests span a wide spectrum of GI science topics, including digital gazetteers, geographic information retrieval, spatial relations, and uncertainty. Each of these topics can be applied toward extracting geographical information from textual documents, especially Web pages. In a recent study, he developed a method that uses the number of toponym cooccurrences retrieved from Web pages to conduct spatial analyses, demonstrating that Web pages can serve as a low-cost data source for geographical applications (Figure 10.2).

In the coming Big Data era, Yu also recognizes the high potential of geospatial Big Data (e.g., volunteered geographical information) for both GI systems and GI science. The broad spread of information and communication technologies has provided a wide range of individual-level and spatiotemporally tagged data sources (Figure 10.3). It thus becomes possible to investigate human mobility patterns and social networks from social media data, mobile phone data, taxi trajectory data, and so on. Recently the work of his group has focused on data mining and knowledge discovery from big data. He believes that such research will provide an innovative perspective on traditional geographical topics, including spatial interaction and distance-decay effects. Revealing the underlying geographical impacts of observed patterns leads to a better understanding of our socioeconomic environments.

Yu received a BS in geography (1994) and a PhD in software engineering (2003), both from Peking University. He led a team to develop a GI system platform 10 years ago, and although the project ultimately failed, this experience provided him with special insight into GI science research. Yu also coauthored a GIS textbook that is well known in China and participated in translating the second edition of this book to Chinese.



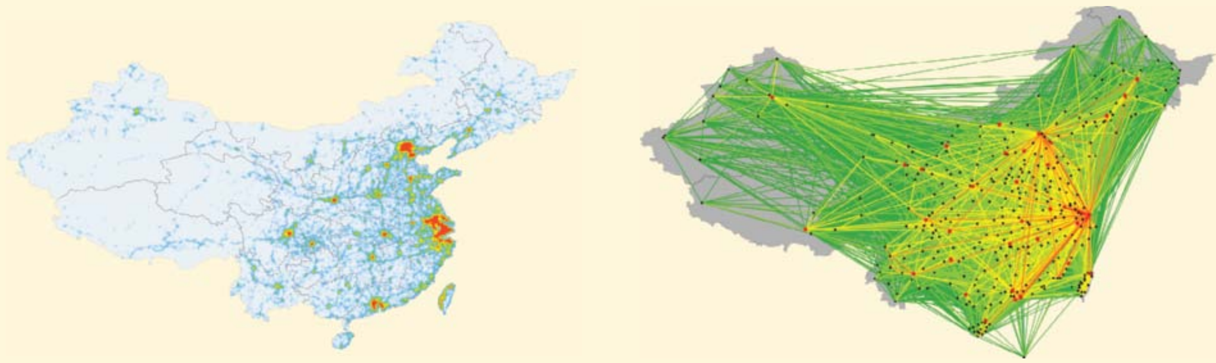

Courtesy: Yu Liu

(A)


(B)

**Figure 10.2** Relatedness of provinces of China, measured using toponym cooccurrences in Web documents. (A) Chord graph depicting the relatedness between provinces. For each city, only the five top relations are depicted. (B) Regionalization based on relatedness between provinces. Two-region clustering is delineated by heavy lines. Six-region clusters are represented by different colors, and the 14 regions are distinguished using different fill patterns. The clustering result identifies several well-known geographical zones, such as northeastern China (including Heilongjiang, Jilin, and Liaoning) and southwestern China (including Sichuan, Chongqing, Guizhou, Yunnan, and Guangxi). Note that Taiwan is blank due to data absence, and the South China Sea Islands are not shown for simplicity.

Courtesy: Yu Liu



(A) (B)

**Figure 10.3** Spatial distribution of check-in records collected from a social media Web site over the course of one year. The dataset covers 521,000 users and 370 cities in China. (A) Density map of all check-in points, clearly depicting the distributions of cities and transportation networks in China. (B) From the users' movements, we can obtain the interactions between the 370 cities. The red lines indicate stronger connections. Note that the South China Sea Islands are not shown for simplicity.

computers are now connected, has enabled a new vision of *online* GI systems, in which the component parts no longer need to be colocated. New technologies are making it possible for a GI project to be conducted not only on the desktop but also anywhere the user chooses to be, using data located anywhere on the network, and using services provided by remote sites on the network.

**Online GI systems allow the six component parts to be at different locations.**

Although some of this chapter's discussion will be concerned with online services that are distributed across several independently managed sites, with all that implies about the need for standards and interoperability, much of the discussion will be about a more integrated architecture known as the Cloud. Cloud computing, as offered by companies such as Amazon or Microsoft, provides a massive online computing resource that is effectively transparent to the user, meaning that very little effort is needed to migrate from the user's own desktop to the Cloud: Software developed for the desktop is expected to operate equally easily in the Cloud. The advantages are obvious: Many users, say within one organization, are able to access the application simultaneously; data and software updates no longer need to be distributed to individual computers, but can be hosted in the Cloud; workloads can be distributed across the Cloud's resources much more cost-effectively, lowering costs to all users; and the responsibility for maintaining the Cloud no longer falls to the user's organization. Note that although this chapter refers frequently to "the Cloud," there are in reality many Clouds, some owned by organizations and restricted to members, and others offered commercially for general purposes.

In Chapter 6 we discussed some aspects of this new concept of online GI systems and services. Many of those services are now available through smartphones, including wayfinding and mapping, and searches that are geographically targeted ("find the nearest coffee shop to my current location"). In Section 6.6.6 we termed these "handheld GI systems and services." Statistical agencies such as the U.S. Bureau of the Census or the U.S. Department of Agriculture are now using handheld devices widely for field data collection and routinely equip their field crews with such devices, allowing them to record data in the field and to upload the results wirelessly or back in the office; the field crews of utility companies use similar systems to record the locations and status of transformers, poles, and switches. Figures 10.4 and 10.5 show

**Figure 10.4** A soldier using a simple GIS in the field to collect data. The device on the pole is a GPS antenna, used to georeference data as they are collected.

© AFP/Getty Images, Inc.

© Sean Locke/iStockphoto

**Figure 10.5** Recording data in the field, an increasingly common task in many occupations. In this case the device is used to record the location and other details of a delivery.

examples of such activities. Vendors are also offering various forms of online GI systems (Sections 6.6.2 and 6.6.3) that begin to offer services comparable to the analytic power of the traditional desktop.

Certain concepts need to be clarified at the outset. First, there are four distinct locations of significance to distributed and online GI systems and services:

- The location of the user and of the interface through which the user submits requests and obtains information, denoted by *U* (e.g., the handheld device in Figure 10.4).
- The location of the data being accessed by the user, denoted by *D*. Traditionally, data had to be moved to the user's computer before being used, but today's technology allows data to be accessed directly from data warehouses and archives, or indirectly through portals.
- The location where the data are processed, denoted by *P*. In Section 1.5.2.3 we introduced the concept of a GI service, a processing capability accessed at a remote site rather than provided locally by the user's desktop GI system. A Cloud will include many servers, perhaps distributed over a wide area.
- The area that is the focus of the GI project, or the *subject* location, denoted by *S*. All GI projects necessarily study some area, obtain data as a representation of the area, and apply GI system processes to those data.

In traditional GI projects three of these locations—*U*, *D*, and *P*—are the same because both the data and processing occur at the user's desktop. The subject location could be anywhere in the world, depending on the project. But in distributed and online GI projects there is no longer any need for *D* and *P* to be the same as *U*. Moreover, it is possible for the user to be located in the subject area *S* and be able to see, touch, feel, and even smell it, rather than being in a distant office. The GI interface might be held in the user's hand, or stuffed in a backpack, or mounted in a vehicle.

**In distributed and online GI projects, the user location and the subject location can be the same.**

Critical to distributed and online GI projects are the standards and specifications that make it possible for devices, data, and processes to operate together, or *interoperate*. Some of these are universal, such as ASCII, the standard for the coding of characters (Box 3.1), and XML, the extensible markup language that is used by many services on the Web. Others are specific to GI, and many of these standards have been developed through the Open Geospatial Consortium (OGC, www.opengeospatial.org), an organization set up to promote openness and interoperability in the world of GI. Among the many successes of OGC over the past decade are the simple feature specification, which standardizes many of the terms associated with primitive elements of GIS databases (polygons, polylines, points, etc.; see also Chapter 7); Geography Markup Language (GML), a version of XML that handles geographic features and enables open-format communication of geographic data; and specifications for Web services (Web Map Service, WMS; Web Feature Service, WFS; and Web Coverage Service, WCS) that allow a user's software to request data automatically from remote servers.

Distributed and online GI systems and services reinforce the notion that today's computer is not simply the device on the desk, with its hard drive, processor, and peripherals, but something more extended. The slogan "The Network Is the Computer" provided an early vision for at least one company—Sun Microsystems (www.sun.com)—and propelled many major developments in computing. The term *cyberinfrastructure* describes a contemporary approach to the conduct of science, relying on high-speed networks, massive processors, and distributed networks of sensors and data archives (www.cise.nsf.gov/sci/reports/toc.cfm). By integrating the world's computers it is possible to provide the kinds of massive computing power that are needed by projects such as SETI (the Search for Extra-Terrestrial Intelligence, www.seti.org), which processes terabytes of data per day in the search for anomalies that might indicate life elsewhere in the universe, and makes use of computer power wherever it can find it on the Internet (Figure 10.6). *Grid* computing is a generic term for such a fully integrated worldwide network of computers and data, and as we have already seen, the Cloud represents the convergence of a number of separate trends into a new vision for online computing.

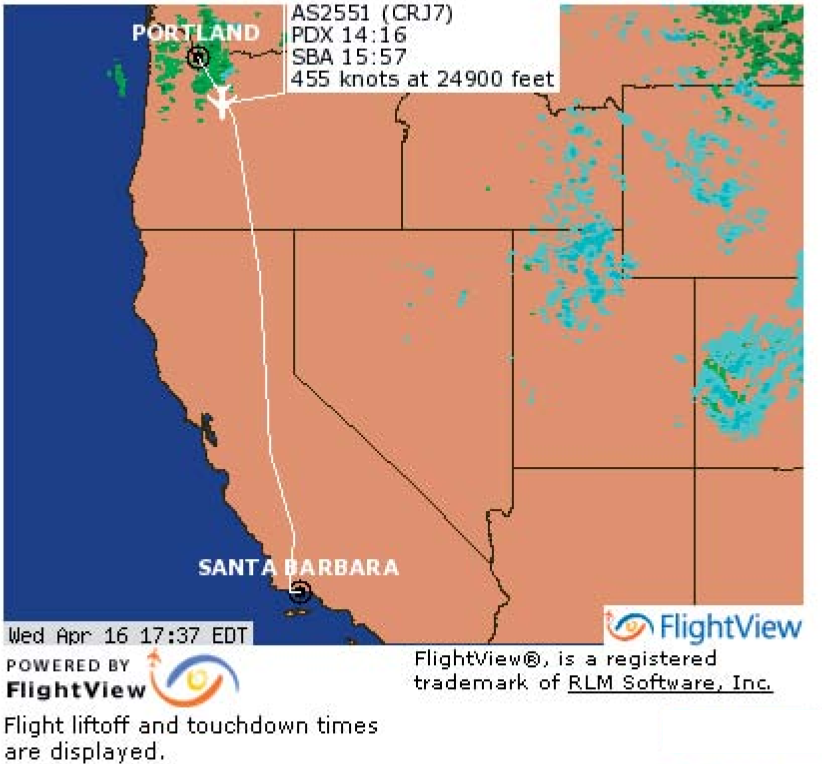


**Figure 10.6** The power of this home computer would be wasted at night when its owner is sleeping, so instead its power has been "harvested" by a remote server and used to process signals from radio telescopes as part of a search for extraterrestrial intelligence.

In the early days of GI systems, all the data were derived from paper maps and described features on the Earth's surface that were largely static and unchanging. But an increasing amount of real-time data is becoming available, thanks in part to GPS, and is being fed over the Internet to users. It is now routine, for example, to keep track of the progress of an arriving flight before going to the airport (Figure 10.7), or to make available real-time data on traffic congestion that are obtained by monitoring the speed of vehicles using GPS (see, for example, www.inrix.com). Real-time data are available from surveillance cameras, networks of environmental sensors, and an increasing range of small, cheap devices. RFID (radio frequency identification) allows the tracking of objects that have been implanted or tagged with small sensors and is now widely used in retailing, livestock management, and building construction. Many mobile phones carry RFID tags, as do new passports, items purchased from major retailers, vehicles using automatic highway toll gates, and even some oddly motivated individuals (Figure 10.8). A QR (quick response) code on an object (Figure 10.9) allows a sensor such as a mobile phone to immediately link to an Internet site where additional information about the object can be found. All of this raises the prospect that at some point in the future, it will be possible to know where *everything* is, or at least everything that is important in some specific area of human activity.

In Chapter 4 we introduced the concept of a *mashup*, which describes linking Web sites to create new services that none of the component sites can provide alone and has special significance when the linking is done through geographic location, when it is akin to overlay (see Section 13.2.4). Mashup, and the linking of services in general, is a key concept of the GeoWeb. Another is *service-oriented architecture* (SOA), the

**Figure 10.7** Tracking a flight: the real-time location of Alaska Airlines Flight 2551 from Portland to Santa Barbara, together with a base map and a depiction of current weather (light blue shows snow).




Steven Puetzer/Photolibrary Group Limited

**Figure 10.8** RFID (radio frequency identification) uses small tags to identify and track various kinds of objects.

notion that any complex computer application can be decomposed into component parts, and that each of these parts can be provided by services that are distributed over the Internet. For example, suppose one is interested in developing an application to display current news stories, using a map interface that allows the user to zoom to any part of the world and to read about current events there. The service might combine a real-time feed of news stories, including text and images, from Web site A; sending the feed to a second site, B, that specializes in finding place-names in text and converting them to latitude and longitude; sending the results to a third site, C, that specializes in generating cartographic displays; and finally connecting the user with this third site. The locations of each of the three sites could be anywhere in the world that is connected to the Internet. SOA requires standards to ensure that the various services are interoperable and mechanisms for searching out services that meet specific requirements. A more elaborate example of SOA is given in Section 10.4.1.

**Figure 10.9** A QR code.

NASA

## 10.2 Distributing the Data

Since its popularization in the early 1990s, the Internet has had a tremendous and far-reaching impact on the accessibility of GI and on the ability of GI users to share datasets. As we saw in Chapter 8, a large and increasing number of Web sites offer GI for free, for sale, or for temporary use and also provide services that allow users to search for datasets that satisfy certain requirements.

In the past few years many private citizens have become involved in the distributed creation and dissemination of geographic information, in a process known as *volunteered geographic information* (VGI; see also Section 1.5.6). The notion that the content of the Web is increasingly created by its users has been termed Web 2.0 (introduced in Section 1.5.5), and today the Web is littered with thousands of sites that support the creation of VGI. Section 4.2 cited one prominent example, Wikimapia, in effect a place-name index created entirely by volunteers and with rich descriptive information. Another example is Open Street Map (OSM), one of many efforts around the world to enlist volunteers in the creation of open, free digital maps that is especially important in areas where such maps are not freely available (Table 1.4). VGI is becoming an important source of geographic information, particularly information that is difficult to obtain from any other source. On the other hand, its contributors have no authority, VGI sites do not generally conduct formal quality testing, and there is no equivalent of the trust that people place in traditional mapping sources (Chapter 17).

In effect, in a period of little more than two decades we have gone from a situation in which geographic data were available only in the form of printed maps from map libraries and retailers, to one in which petabytes (Table 1.2) of information are available for download and use at electronic speed (about 1.5 million CDs would be required to store 1 petabyte). For example, the NASA-sponsored EOSDIS (Earth Observing System Data and Information System; earthdata.nasa.gov) archives and distributes the geographic data from the EOS series of satellites, acquiring new data at over a terabyte per day, with an accumulated total of more than 1 petabyte at this site alone.

**Some GIS archives contain petabytes of data.**

The vision of the GeoWeb goes well beyond the ability to access and retrieve remotely located data, however, because it includes the concepts of *search*, *discovery*, and *assessment*. In the world of distributed GI systems, how do users search for data, discover their existence at remote sites, and assess their fitness for use? Four concepts are important in this respect: *object-level metadata*, *geolibraries*, *geoportals*, and *collection-level metadata*.

## 10.2.1 Object-Level Metadata

Strictly defined, metadata are data about data, and *object-level* metadata (OLM) describe the contents of a single dataset. We need information about data for many purposes, and OLM try to satisfy them all. First, we need OLM to automate the process of search and discovery over archives. In that sense OLM are similar to a library's catalog, which organizes the library's contents by author, title, and subject, and makes it easy for a user to find a book. But OLM are potentially much more powerful because a computer is more versatile than the traditional catalog in its potential for re-sorting items by a large number of properties, going well beyond author, title, and subject, and including geographic location. Second, we need OLM to determine whether a dataset, once discovered, will satisfy the user's requirements—in other words, to assess the fitness of a dataset for a given use. Does it have sufficient spatial resolution and acceptable quality? Such metadata may include comments provided by others who tried to use the data, or contact information for such previous users (users often comment that the most useful item of metadata is the phone number of the person who last tried to use the data). Third, OLM must provide the information needed to handle the dataset effectively. This may include technical specifications of format, or the names of software packages that are compatible with the data, along with information about the dataset's location and its volume. Finally, OLM may provide useful information on the dataset's contents. In the case of remotely sensed images, this may include the percentage of cloud obscuring the scene, or whether the scene contains particularly useful instances of specific phenomena, such as hurricanes.

**Object-level metadata are formal descriptions of datasets that satisfy many different requirements.**

OLM generalize and abstract the contents of datasets, and we would therefore expect the datasets to be smaller in volume than the data they describe. In reality, however, it is easy for the complete description of a dataset to generate a greater volume of information than the actual contents. OLM are also expensive to generate because they represent a level of understanding of the data that is difficult to assemble and require a high level of professional expertise. Generation of OLM for a geographic dataset can easily take much longer than it takes to catalog a book, particularly if it is necessary to deal with technical issues such as the precise geographic coverage of the dataset, its projection and datum details (Chapter 4), and other properties that may not be easily accessible. Thus the cost of OLM generation, and the incentives that motivate people to provide OLM, are important issues.

For metadata to be useful, it is essential that they follow widely accepted standards. If two users are to be able to share a dataset, they must both understand the rules used to create its OLM, so that the custodian of the dataset can first create the description and so that the potential user can understand it.

The most widely used standard for OLM is the U.S. Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata, or CSDGM, first published in 1993 and now the basis for many other standards worldwide. Box 10.2 lists some of its major features. As a *content standard*, CSDGM describes the items that should be in an OLM archive, but does not prescribe exactly how they should be formatted or structured. This allows developers to implement the standard in ways that suit their own software environments, but guarantees that one implementation will be understandable to another—in other words, that the implementations will be *interoperable*. For example, Esri's ArcGIS provides several formats for OLM, including one using the widely recognized XML standard and another using Esri's own format.

CSGDM was devised as a system for describing geographic datasets, and most of its elements make sense only for data that are accurately georeferenced and represent the spatial variation of phenomena over the Earth's surface. As such, its designers did not attempt to place CSGDM within any wider framework. But in the past decade a number of more broadly based efforts have also been directed at the metadata problem and at the extension of traditional library cataloging in ways that make sense in the evolving world of digital technology.

One of the best known of these is the Dublin Core (see Box 10.3), which is the outcome of an effort to find the minimum set of properties needed to support search and discovery for datasets in general, not only geographic datasets. Dublin Core treats both space and time as instances of a single property, "coverage," and unlike CSGDM does not lay down how such specific properties as spatial resolution, accuracy, projection, or datum should be described.

The principle of establishing a minimum set of properties is sharply distinct from the design of CSGDM, which was oriented more toward the capture of all knowable and potentially important properties of geographic datasets. Of direct relevance here is the problem of cost, and specifically the cost of capturing a full CSGDM metadata record. Although many organizations have wanted to make their data more widely available and have been driven to create OLM for their datasets, the cost of determining the full set of CSDGM elements is often highly discouraging. There is interest therefore in a concept of *light metadata*, a limited set of properties that is both comparatively cheap to capture and still useful to support search and discovery. Dublin Core represents this approach and thus sits at the opposite end of a spectrum from CSGDM. Every organization must somehow determine where its needs lie on this spectrum, which ranges from light and cheap to heavy and expensive.

**Light, or stripped-down, OLM provide a short but useful description of a dataset that is cheaper to create.**

## Technical Box 10.2

### Major Features of the U.S. Federal Geographic Data Committee's Content Standards for Digital Geospatial Metadata

1. Identification Information—basic information about the dataset.
2. Data Quality Information—a general assessment of the quality of the dataset.
3. Spatial Data Organization Information—the mechanism used to represent spatial information in the dataset.
4. Spatial Reference Information—the description of the reference frame for, and the means to encode, coordinates in the dataset.
5. Entity and Attribute Information—details about the information content of the dataset, including the entity types, their attributes, and the domains from which attribute values may be assigned.
6. Distribution Information—information about the distributor of and options for obtaining the dataset.
7. Metadata Reference Information—information on the currentness of the metadata information and the responsible party.
8. Citation Information—the recommended reference to be used for the dataset.
9. Time Period Information—information about the date and time of an event.
10. Contact Information—identity of, and means to communicate with, person(s) and organization(s) associated with the dataset.

**Technical Box 10.3**

### The 15 Basic Elements of the Dublin Core Metadata Standard

1. TITLE. The name given to the resource by the CREATOR or PUBLISHER.
2. AUTHOR or CREATOR. The person(s) or organization(s) primarily responsible for the intellectual content of the resource.
3. SUBJECT or KEYWORDS. The topic of the resource, or keywords, phrases, or classification descriptors that describe the subject or content of the resource.
4. DESCRIPTION. A textual description of the content of the resource, including abstracts in the case of document-like objects or content description in the case of visual resources.
5. PUBLISHER. The entity responsible for making the resource available in its present form, such as a publisher, a university department, or a corporate entity.
6. OTHER CONTRIBUTORS. Person(s) or organization(s) in addition to those specified in the CREATOR element who have made significant intellectual contributions to the resource, but whose contribution is secondary to the individuals or entities specified in the CREATOR element.
7. DATE. The date the resource was made available in its present form.
8. RESOURCE TYPE. The category of the resource, such as home page, novel, poem, working paper, technical report, essay, or dictionary.
9. FORMAT. The data representation of the resource, such as text/html, ASCII, Postscript file, executable application, or JPEG image.
10. RESOURCE IDENTIFIER. String or number used to uniquely identify the resource.
11. SOURCE. The work, either print or electronic, from which this resource is delivered, if applicable.
12. LANGUAGE. Language(s) of the intellectual content of the resource.
13. RELATION. Relationship to other resources.
14. COVERAGE. The spatial locations and temporal durations characteristics of the resource.
15. RIGHTS MANAGEMENT. The content of this element is intended to be a link (a URL or other suitable URI as appropriate) to a copyright notice, a rights-management statement, or perhaps a server that would provide such information in a dynamic way.

## 10.2.2 Geolibraries and Geoportals

The use of digital technology to support search and discovery opens up many options that were not available in the earlier world of library catalogs and bookshelves. Books must be placed in a library on a permanent basis, and there is no possibility of reordering their sequence—but in a digital catalog it is possible to reorder the sequence of holdings in a collection almost instantaneously. So although a library's shelves are traditionally sorted by subject, it would be possible to re-sort them digitally by author name or title, or by any property in the OLM catalog. Similarly, the traditional card catalog allowed only three properties to be sorted—author, title, and subject—and discouraged sorting by multiple subjects. But the digital catalog can support any number of subjects.

Of particular relevance to GI users is the possibility of sorting a collection by the coverage properties: location and time. Both the spatial and temporal dimensions are continuous, so it is impossible to capture them in a single property analogous to author that can then be sorted numerically or alphabetically. In a digital system this is not a serious problem, and it is straightforward to capture the coverage of a dataset and to allow the user to search for datasets that cover an area or time of interest defined by the user. Moreover, the properties of location and time are not limited to geographic datasets because many types of information are associated with specific areas on the Earth's surface or with specific time periods. Searching based on location or time would enable users to find information about any place on the Earth's surface, or any time period—and to find reports, photographs, or even pieces of music, as long as they possessed geographical and temporal *footprints*.

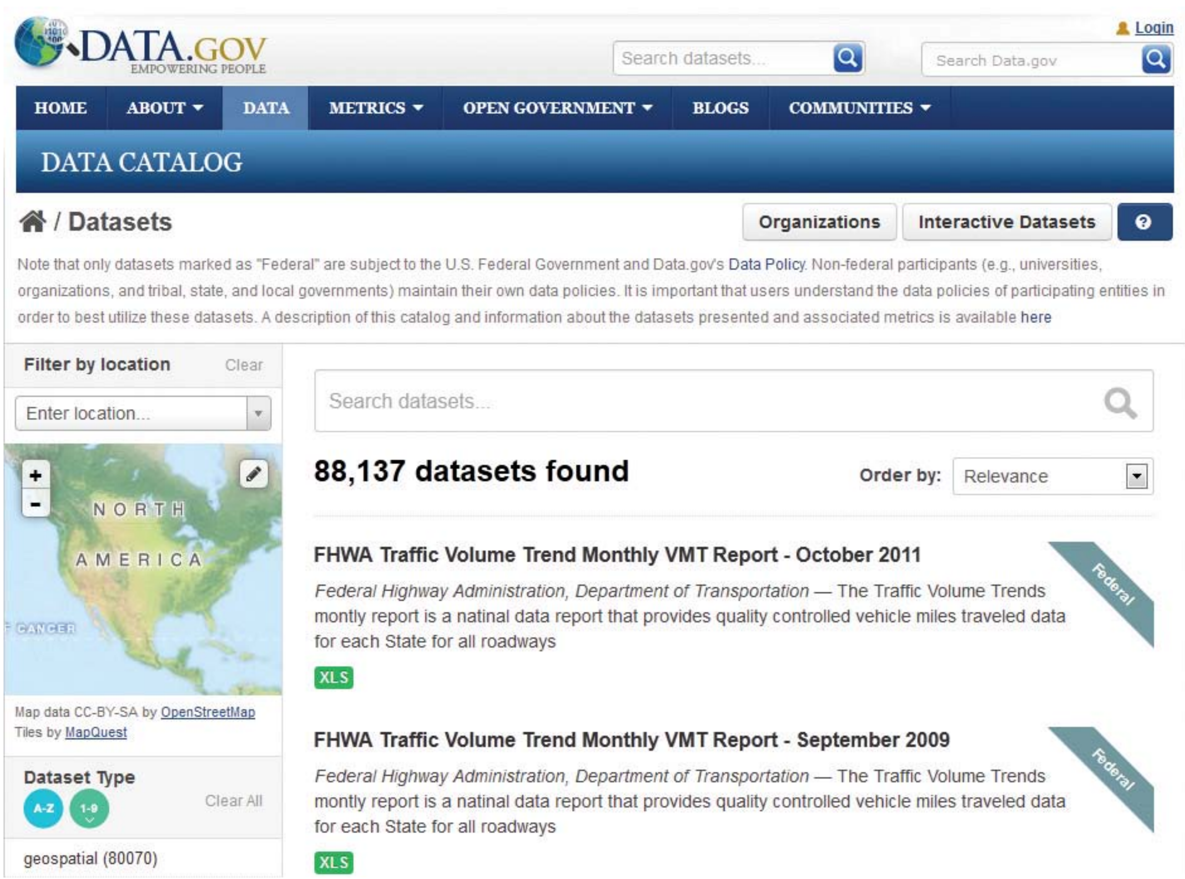
The term *geolibrary* has been coined to describe digital libraries that can be searched for information about any user-defined geographic location. A U.S. National Research Council report (see Further Reading) describes the concept and its implementation, and many instances of geolibraries can be found on the Web.

**A geolibrary can be searched for information about a specific geographic location.**

Although geolibraries are very useful sources of data, it is often difficult for the user to predict *which* geolibrary is most likely to contain a given dataset, or the closest approximation to it. Rather than have to resort to trial and error, it would be much better if it were possible to search a single site for *all* datasets. This is the goal of a *geoportal*, which can be defined as a single point of entry to a distributed collection of geolibraries, with a single catalog, just as many networks of libraries provide a single *union catalog* to all their holdings. There is a similarity here to the Web search engines such as Google, which provide a single point of entry for search over a large proportion of the entire Web. However, conventional search engines are not well designed for finding geographic datasets.

The Geospatial One-Stop, now incorporated into data.gov and accessible at geo.data.gov, is a good example of a geoportal, and in many ways it represents the state of the art in searching for geographic data. It was initiated by the U.S. government as a single point of entry to the holdings of government agencies, but it also catalogs many other geolibraries (though its coverage is essentially limited to the United States). A user visiting geo.data.gov is presented with several ways of specifying requirements (Figure 10.10) and is provided with a list of datasets that appear to satisfy them, together with links to the geolibraries that contain the data. Mechanisms are provided for handling datasets that have usage restrictions and may require licensing or the payment of fees. The site receives tens of thousands of visitors per day and

**Figure 10.10** The U.S. government's geoportal geo.data.gov. The user is able to search through all datasets registered with the portal, selecting by geographic location and many other categories.



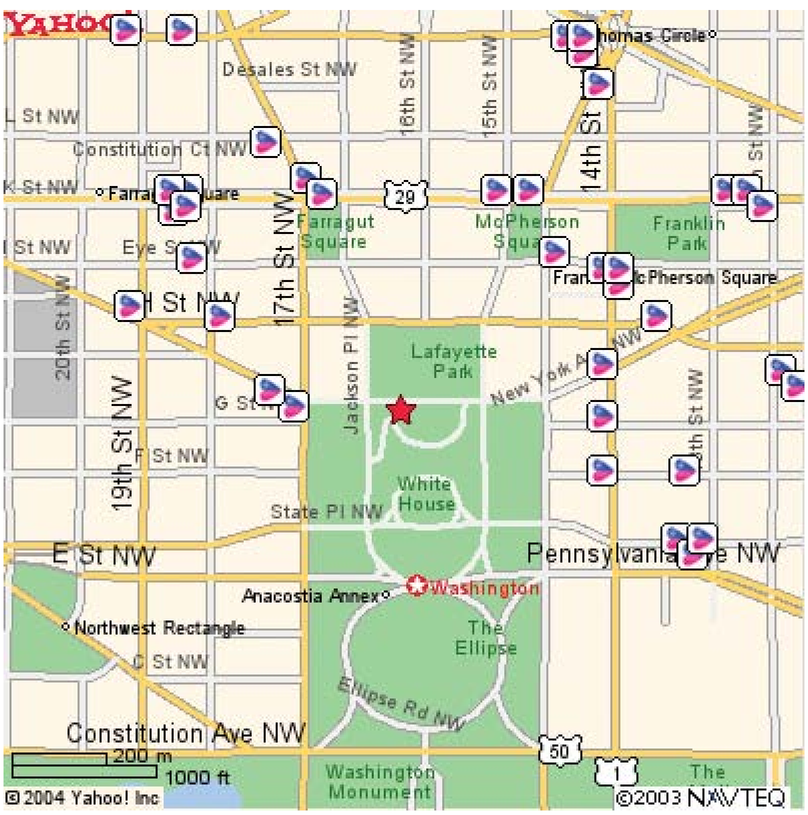
catalogs the contents of over a thousand geolibraries. It also supports automatic harvesting of catalog contents from geolibraries that are willing to contribute in this way.

Contemporary geoportals support several forms of use. The most traditional is the download, when an entire dataset is simply transferred to the user's hard drive and then incorporated in some form of analysis. But it is also possible to rely on many geoportals for simple online functions such as display, in which case the user needs no more than a Web browser. Finally, it may be possible to use a dataset "live" if the geoportal supports the appropriate standards. In this case the user's GI system behaves as if the geoportal's data were local, but sends requests over the Internet rather than to the local hard drive when specific data are needed in response to an operation. In this mode there is no need to download an entire dataset, and standards take care of such potential interoperability problems as differences of coordinate system or geographic coverage. Live data standards include the Open Geospatial Consortium's WMS, WFS, and WCS.

## 10.3 The Mobile User

The computer has become so much a part of our lives that for many people it is difficult to imagine life without it. We increasingly need computers to shop, to communicate with friends, to obtain the latest news, and to entertain ourselves. In the early days, the only place one could use computers was in a computing center, within a few meters of the central processor. Computing had extended to the office by the 1970s and to the home by the 1990s. The portable computers of the 1980s opened the possibility of computing in the garden, at the beach, or "on the road" in airports and airplanes. Wireless communication services such as WiFi (the wireless access technology based on the 802.11 family of standards) now allow broadband communication to the Internet from "hotspots" in many hotels, restaurants, airports, office buildings, and private homes (Figure 10.11). The range of mobile computing devices is also multiplying rapidly, from the relatively cumbersome but powerful laptop weighing several kilograms to the tablet, and the mobile

**Figure 10.11** Map of WiFi (802.11) wireless broadband "hotspots" within 1 mile of the White House (1600 Pennsylvania Avenue, Washington, DC) and using the T-Mobile Internet provider.




Courtesy: Krzysztof Janowicz

**Figure 10.12** Google's Glass, an example of a wearable, hands-free device, modeled here by Krzysztof Janowicz, University of California, Santa Barbara.

phone weighing a hundred or so grams. In many ways the ultimate endpoint of this progression is the *wearable* computer, a device that is worn on the body or fully embedded in the user's clothing, goes everywhere, and provides ubiquitous computing service. Such devices are already obtainable, including Google's Glass (Figure 10.12), which is being promoted as a hands-free phone, a camera, and a simple computing device.

**Computing has moved from the computer center and is now close to, even inside, the human body. The ultimate form of mobile computing is the wearable computer.**

What possibilities do such systems create for GI? Of greatest interest is the case in which the user is situated in the subject location; that is, *U* is contained within *S*, and the system is used to analyze immediate surroundings. A convenient way to think about the possibilities is by comparing *virtual* reality with *augmented* reality.

## 10.3.1 Virtual Reality and Augmented Reality

One of the great strengths of GI systems is the window they provide on the world. A researcher in an office in Nairobi, Kenya, might use a GI system to obtain and analyze data on the effects of salinization in the Murray Basin of Australia, combining images from satellites with base topographic data, data on roads, soils, and population distributions, all obtained from different sources via the Internet. In doing so, the researcher would build a comprehensive picture of part of the world he or she might never have visited, all through the medium of digital geographic data and the GI system, and might learn almost as much through this window as from actually being there. The expense and time of traveling to Australia would be avoided, and the analysis could proceed almost instantaneously. Some aspects of the study area would be missing, of course—aspects of culture, for example, that can best be experienced by meeting with the local people (see the discussion of representation in Chapter 3).

Research environments such as this are termed *virtual realities* because they replace what humans normally gather through their senses—sight, sound, touch, smell, and taste—by presenting information from a database. In most GI system applications, such as those discussed in Section 12.4.3, only one of the senses, sight, is used to create this virtual reality, or VR. In principle it is possible to record sounds and store them in a GI database as attributes of features, but in practice very little use is made of any sensory channel other than vision. Moreover, in most GI system applications the view presented to the user is the view from *above*, even though our experience with looking at the world from this perspective is limited (for most of us, it is restricted to times when we requested a window seat in an airplane). GI systems have been criticized for what has been termed their *God's-eye*, or a *privileged* view by some writers, on the basis that it distances the researcher from the real conditions experienced by people on the ground.

**Virtual environments can place the user in distant locations.**

More elaborate VR systems are capable of *immersing* the user by presenting the contents of a database in a three-dimensional environment using special eyeglasses or by projecting information onto walls surrounding the user and effectively *transporting* the user into the environment represented in the database. But even the standard personal computer is capable of creating remarkably close approximations to the physical appearance of the geographic landscape, and services such as Google Earth and Microsoft's Bing Maps have pushed the limits dramatically in recent years. Such *geobrowsers* or *virtual globes* create three-dimensional renderings of the Earth and rely on the powerful graphics capabilities now available in the standard office or home computer to support real-time movement—a "magic carpet ride" that even a child of 10 can learn to experience. Recent versions allow a smooth transition from the "God's-eye view" from above to street views assembled from hundreds of millions of

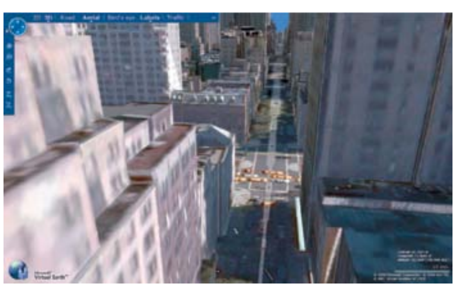



**Figure 10.13** View along East 64th Street in Manhattan generated in Google Maps using street-level photographs.

photographs (Figure 10.13), and the 3-D extension of Microsoft's Bing Maps presents photorealistic exteriors of buildings (Figure 10.14).

Roger Downs, professor of geography at Pennsylvania State University, uses Johannes Vermeer's famous painting popularly known as *The Geographer* (Figure 10.15) to make an important point about virtual realities. On the table in front of the figure is a map, taken to represent the geographer's window on a part of the world that happens to be of interest. But the subject figure is shown looking out the window, at the real world, perhaps because he needs the information he derives from his senses to understand the world as shown on the map. The idea of combining information from a database with information derived directly through the senses is termed *augmented reality*, or AR. In terms of the locations of computing discussed earlier, AR is clearly of most value when the location of the user *U* is contained within the subject area *S*, allowing the user to augment what can be seen directly with information retrieved about the same area from a database. This might include historic information, or predictions about the future, or information that is for some other reason invisible to the user.

**Figure 10.14** Simulated view along East 64th Street in Manhattan generated in Microsoft's Bing Maps 3D, with three-dimensional buildings and exterior textures.



AR can be used to augment or replace the work of senses that are impaired or absent. A team led by the late Reginald Golledge, professor of geography at the University of California, Santa Barbara, experimented with AR as a means of helping visually impaired people to perform the simple task of navigation. The system, which was worn by the user (Figure 10.16), included a differential GPS for accurate positioning, a GI database that included very detailed information on the immediate environment, a compass to determine the position of the user's head, and a pair of earphones. The user gave the system information to identify the desired destination, and the system then generated sufficient information to replace the

**Figure 10.15** Johannes Vermeer's painting of 1669.

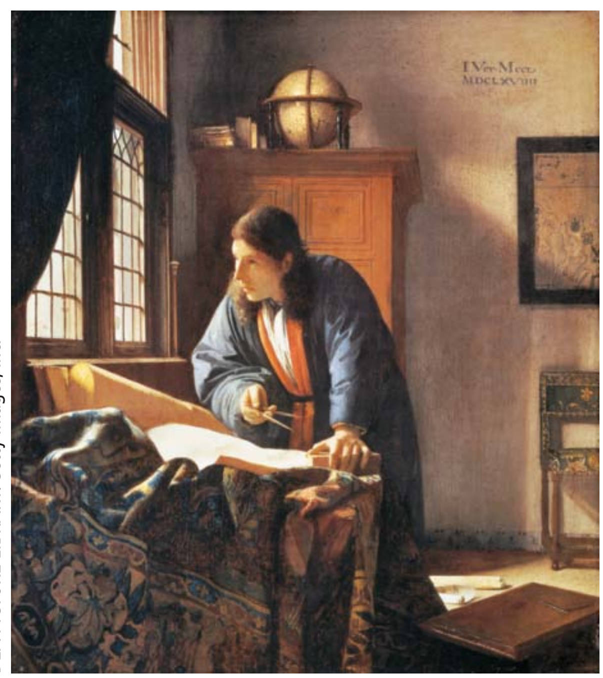




**Figure 10.16** The system worn here by the late Reg Golledge (a leader of the development team) uses a GI system and GPS to augment the senses of a visually impaired person navigating through a complex space such as a university campus.

normal role of sight, either through verbal instructions, or through the generation of stereo sounds that appeared to come from the appropriate direction.

**Augmented reality combines information from the database with information from the senses.**

Steven Feiner, professor of computer science at Columbia University, has demonstrated another form of AR that superimposes historic images and other information directly on the user's field of view. For example, in Figures 10.17 and 10.18 a user wearing a head-mounted device coupled to a wearable computer is seeing both the Columbia University main library building and an image generated from a database showing the building that occupied the library's position prior to the university's move to this site in 1896: the Bloomingdale Insane Asylum.

## 10.3.2 Location-Based Services

A location-based service (LBS) is an information service provided by a device that knows where it is and is capable of modifying the information it provides based on that knowledge. Traditional computing devices, such as desktops or laptops, have no way of

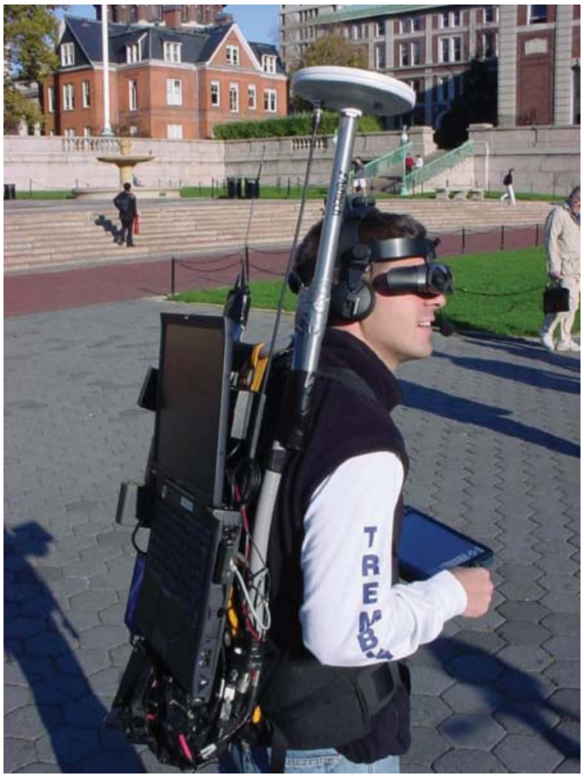




**Figure 10.17** An augmented reality system developed at Columbia University by Professor Steve Feiner and his group.

knowing where they are, and their functions are in no way changed when they are moved. But increasingly the essential information about a device's location is available, from a GPS or RFID tag incorporated in the device or from the device's IP address (Section 4.4), and is being used for a wide range of purposes.

**A location-based service is provided by a computing device that knows where it is.**

The simplest and most obvious form of locationally enabled device is the GPS receiver, and any device that includes a GPS capability is capable in principle of providing LBS. The most ubiquitous LBS-capable devices are the modern mobile phone and tablet. A variety of methods exist for determining device locations, including GPS embedded in the device itself, measurements made by the mobile phone of signals received from towers, and measurements made by the towers of signals received from the device. As long as the device is on, the operator of the mobile network is able to pinpoint its location at least to the accuracy represented by the size of the cell, on the order of 10 km, and frequently much more accurately, even to 10 m.

One of the strongest motives driving this process is emergency response. A large proportion of emergency calls come from mobile phones, and although the location of each landline phone is likely to be recorded in a database available to the emergency

**Figure 10.18** The user's field of view of Feiner's AR system, showing the Columbia University main library and the insane asylum that occupied the site in the 1800s.

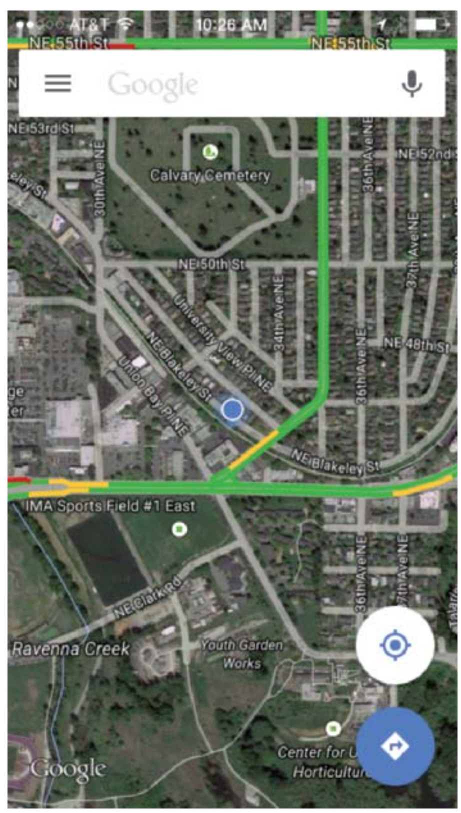


responder, in a significant proportion of cases the user of a mobile phone is unable to report his or her current location to sufficient accuracy to enable effective response. Several well-publicized cases have drawn attention to the problem. The magazine *Popular Science*, for example, reported a case of a woman who lost control of her car in Florida in February 2001, skidding into a canal. Although she called 911 (the emergency number standard in the United States and Canada), she was unable to report her location accurately and died before the car was found.

**Emergency services provide one of the strongest motivations for LBS.**

There are many other examples of LBS that take advantage of locationally enabled mobile phones. A Yellow Pages service responds to a user who requests information on businesses that are close to his or her current location (Where is the nearest pizza restaurant? Where is the nearest hospital? Where is the nearest WiFi hotspot?) by sending a request that includes location to a suitable Web server. The response might consist of an ordered list presented on the mobile phone screen, or a simple map centered on the user's current location (Figure 10.19). A trip planner gives the user the ability to find an optimum driving route from the current location to some defined destination. Similar services are now being provided by public transport operators, and in some cases these services make use of GPS transponders on buses and trains to provide information on actual, as distinct from scheduled, arrival and departure times. Some social networking sites allow a mobile phone user to display a map showing the current locations of nearby friends (provided their mobile phones are active and they are also registered for this service). Geocaching (see Figure 12.6) is a type of orienteering in which contestants navigate their way to sites using GPS and conventional navigational means.

**Figure 10.19** A map displayed on a mobile phone and centered on the user's current location.



© Cihan Taskin/iStockphoto

**Figure 10.20** Credit card in use.

Direct determination of location, using GPS or measurement to or from towers, is only one basis on which a computing device might know its location, however. Other forms of LBS are provided by fixed devices and rely on the determination of location when the device was installed. For example, many point-of-sale systems that are used by retailers record the location of the sale, combining it with other information about the buyer obtained by accessing credit-card or store-affinity-card records (Figure 10.20). In exchange for the convenience or financial inducement of a card, the user effectively surrenders some degree of location privacy to the company whenever the card is used. One benefit is that it is possible for the company to analyze transactions, looking for patterns of purchase that are outside the card user's normal buying habits, perhaps because they occur in locations that the user does not normally frequent, or at anomalous times. Many of us will have experienced the embarrassment of having a credit card transaction refused in an unfamiliar city because the techniques used by the company have flagged the transaction as an indicator that the card might have been stolen. In principle, a store-affinity card gives the company access to information about buying habits and locations, in return for a modest discount.

**Location is revealed every time a credit, debit, or store-affinity card is used.**

## 10.3.3 Issues in Mobile GIS

Using GI systems in the field or "on the road" is very different from using them in the office. First, the location of the user is important and is directly relevant to the application. It makes good sense to center maps on the user's location, to provide the capability for maps that show the view from the user's location rather than from above, and to offer maps that are oriented to the user's direction of travel, rather than north. Second, the field environment may make certain kinds of interaction impractical or less desirable. In a moving vehicle, for example, it would be dangerous to present the driver with visual displays, unless perhaps these are directly superimposed on the field of view (on the windshield). Instead, such systems often provide instructions through computer-generated speech and use speech recognition to receive instructions. With wearable devices that provide output on minute screens built into the user's eyeglasses, there is no prospect of conventional point-and-click with a mouse, so again voice communication may be more appropriate. On the other hand, many environments are noisy, creating problems for voice recognition.

One of the most important limitations to mobility remains the battery, and although great strides have been made in recent years, battery (or other wireless energy) technology has not advanced as rapidly as other components of mobile systems, such as processors and storage devices. Batteries typically account for the majority of a mobile system's weight, and they limit operating time severely.

**The battery remains the major limitation to LBS and to mobile computing in general.**

Although broadband wireless communication is possible using WiFi, connectivity remains a major issue. Wireless communication techniques tend to be

- Limited in spatial coverage. WiFi hotspots are limited to a single building or perhaps a hundred meters from the router; mobile phone–based techniques have wider coverage but lower bandwidth (communication speeds), and only the comparatively slow satellite-based systems approach global coverage.
- Noisy. Mobile phone and WiFi communications tend to "break up" at the edges of coverage areas, or as devices move between cells, leading to errors in communication.
- Limited in temporal coverage. A moving device is likely to lose signal from time to time, whereas some devices, particularly recorders installed in commercial vehicles and those used for surveys, are unable to upload data until within range of a home system or fixed beacon.
- Insecure. Wireless communications often lack adequate security to prevent unwanted intrusion. It is comparatively easy, for example, for someone to tap into a wireless session and obtain sensitive information.
- Progress is being made on all these fronts, but it will be some years before it becomes possible to use GI systems and services as efficiently, effectively, and safely anywhere in the field as it currently is in the office.

## 10.4 Distributing the Software: GI Services

This final section addresses distributed processing, the notion that the actual operations of a GI system might be provided from remote sites, rather than by the user's own computer. Despite the move to component-based software (Section 6.3.3), it is still true that almost all the operations performed on a user's data are executed in the same computer, the one on the user's desk. Each copy of a popular GI system includes the same functions, which are replicated in every installation of that system around the world. When new versions are released, they must be copied and installed at each site, and because not all copies are replaced, there is no guarantee that the system used by one user is identical to the system used by another, even though the vendor and product name are the same.

A GI service is defined as a program executed at a remote site that performs some specific task. The execution of the program is initiated remotely by the user, who may have supplied data, or may rely on data provided by the service, or both. A simple example of a GI service is that provided by wayfinding sites such as MapQuest (www.mapquest.com) or Yell.com. The user's current location and desired destination are provided to the service (the current location may be entered by the user or provided automatically from GPS), but the data representing the travel network are provided by the GI service. The results are obtained by solving for the shortest path between the origin and the destination (often a compromise between minimizing distance and minimizing expected travel time), a function that exists in many GI systems but in this case is provided remotely by the GI service. Finally, the results are returned to the user in the form of driving instructions, a map, or both.

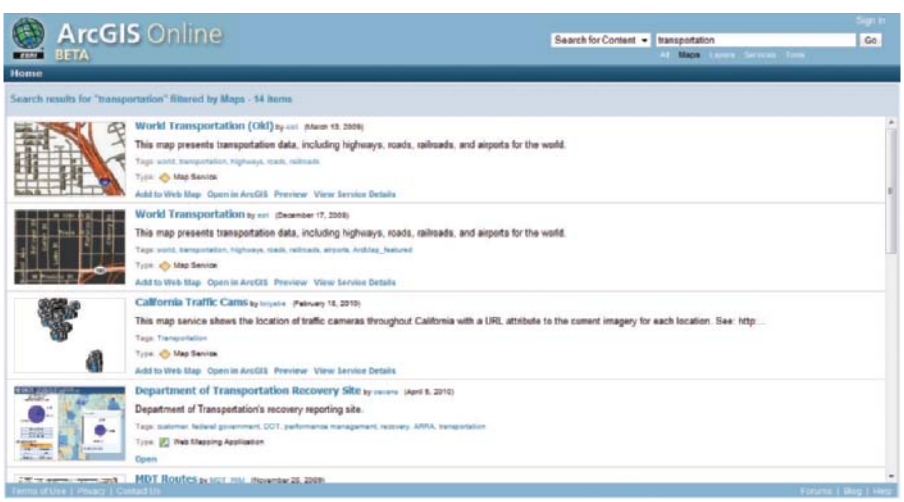
**A GI service replaces a local function with one provided remotely by a server.**

In principle, any GI system function could be provided in this way, based on server software (Section 6.6.3). In practice, however, certain functions tend to have attracted more attention than others. One obvious problem is commercial: How would a GI service pay for itself, would it charge for each transaction, and how would this compare to the normal sources of income for GI system vendors based on software sales? Some services are offered free and generate their revenue by sales of advertising space or by offering the service as an add-on to some other service. MapQuest and Yell are good examples, generating much of their revenue through direct advertising and through embedding their service in other Web sites, such as travel services. Other services, such as Esri's ArcGIS Online, are offered on a subscription basis where the user is charged based on actual use.

In general, the characteristics that make a function suitable for offering as a GI service appear to be

- Reliance on a database that must be updated frequently and is too expensive for the average user to acquire. Both geocoding (Box 4.3) and wayfinding services fall into this category, as do gazetteer services (Section 4.10).
- Reliance on GI system operations that are complex and can be performed better by a specialized service than by a generic system.

The number of available GI services and GeoWeb services is growing steadily, creating a need for directories and other mechanisms to help users find them and standards and protocols for interacting with them. Esri's ArcGIS Online (arcgis.com; Figure 10.21) provides such a directory, in addition to its role as a geoportal to distributed data. The next section describes the state of the art in exploiting distributed GI services.




Courtesy: Esri

**Figure 10.21** ArcGISOnline provides a directory of remote GIServices. In this case a search for services related to transportation networks has identified fourteen maps and services.

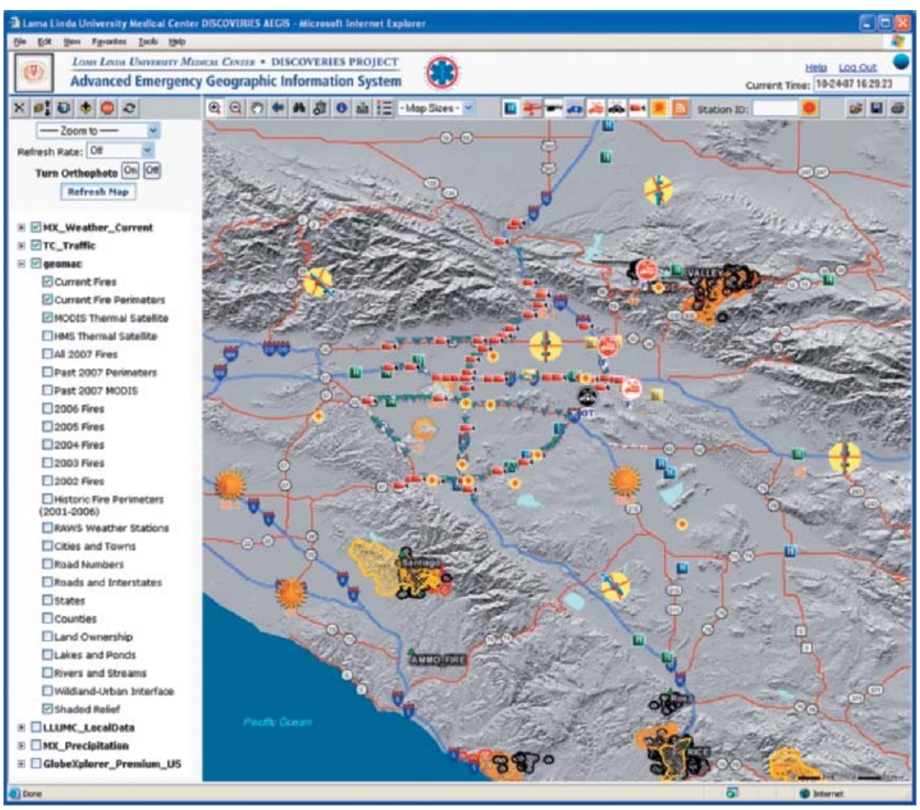
## 10.4.1 Service-Oriented Architecture

When managing recovery in the aftermath of disasters, it is essential that managers have a complete picture of the situation: a map showing the locations and status of incidents, rescue vehicles and hospitals, traffic conditions, weather, and many other relevant assets and factors. A report of the U.S. National Research Council discusses the use of GI technology in all aspects of emergency management (see Further Reading).

To illustrate the capabilities of current technology, particularly distributed GI services and the GeoWeb, the Loma Linda University Medical Center and Esri teamed up to develop the Advanced Emergency GIS (AEGIS). It employs mashups and a service-oriented architecture to gather a wide variety of relevant data and to present it in readily understood form to emergency managers and other users. Figure 10.22 shows a typical synoptic visualization during the series of wildfires that hit Southern California in late 2007. The various icons signal the availability of information on traffic conditions (real-time video feeds from cameras on the freeway network), traffic incidents, hospitals (e.g., the number of available beds in each hospital's emergency room), helicopters, the current footprint of each fire, and much other useful information, any of which can be displayed by clicking the icon.

Figure 10.23 shows the architecture of the system. Roughly ten servers maintained by the relevant agencies, such as the California Department of Transportation and the U.S. Geological Survey, provide real-time feeds of data. The server in the center polls each server at regular intervals, with a frequency depending on the rate of change of the respective server's data, which may range from seconds in the case of emergency vehicle locations to months in the case of base mapping. The format of data contributed by each server varies and in some cases requires the use of an additional service, such as geocoding, or special software to extract relevant information. The central server integrates all the feeds into a composite mashup and distributes the result to users such as the one symbolized on the right. Several standards are involved, including SOAP (Simple Object Access Protocol), RSS (originally Rich Site Summary), and HTTP (Hypertext Transfer Protocol).

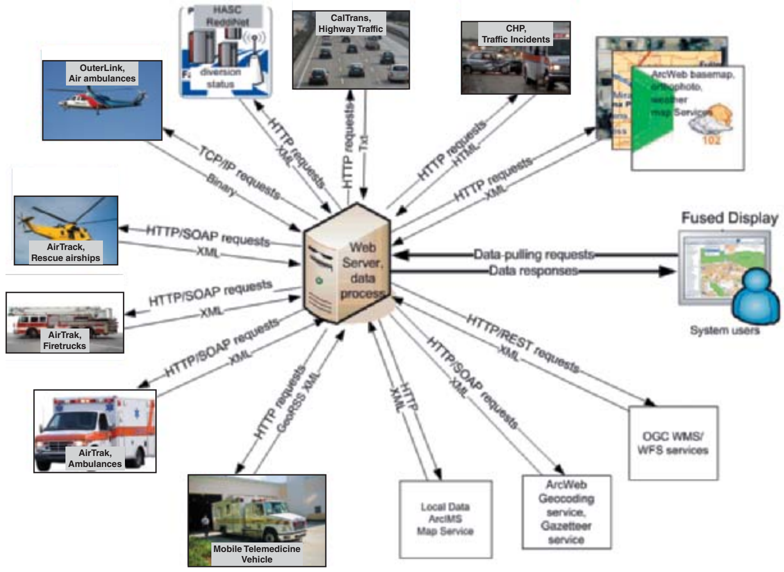
It would be dishonest to suggest that this kind of application is easy to develop. Although standards exist, there are a large number of them, and each requires its own specialized approach. In this example the most problematic case is the incident report feed from the California Department of Transportation, which requires analyzing the text of Web pages posted by the agency using a specially developed program. Moreover, AEGIS relies on the correct operation of many servers and feeds and is clearly vulnerable to power and network outages. However, there is no doubt that in future it will be easier to develop complex SOA-based systems, given the obvious power and immediacy that is possible.




Courtesy: Esri

**Figure 10.22** The Advanced Emergency GIS (AEGIS) in operation during the wildfire outbreak in Southern California, late 2007.

Figure 10.23 The service-oriented architecture of AEGIS.




Courtesy: Esri

## 10.5 Prospects

Distributed, online GI services offer enormous benefits: They reduce duplication of effort, allow users to take advantage of remotely located data and services through simple devices, and provide ways of combining information gathered through the senses with information obtained from digital sources. Many issues continue to impede progress, however: complications resulting from the difficulties of interacting with devices in field settings; limitations placed on communication bandwidth and reliability; and limitations inherent in battery technology. Some older issues have largely been overcome, including problems of incompatible georeferencing systems (Chapter 4), but other incompatibilities continue to exist, including differences in the meaning of terms used to describe geographic features. Perhaps more problematic than any of these at this time is the difficulty of imagining the full potential of distributed GI services. We are used to GI technology on the desktop and conscious that we have not yet fully exploited its potential. So it is hard to imagine what might be possible when the technology can be carried anywhere and its information combined with the window on the world provided by our senses.

More broadly, the GeoWeb offers a vision of a future in which geographic location is central to a wide range of applications of information technology, and one in which individuals are able to access detailed information about the dynamic state of transportation, weather, and a host of other socially important domains. We are just beginning to see the potential of this vision and to recognize that although its benefits are often compelling, it raises invasion of privacy and surveillance issues that society as a whole will have to address (Box 10.4).

**Biographical Box 10.4**

### Sarah Elwood and the social implications of the GeoWeb

Sarah Elwood (Figure 10.24) is a professor of geography at the University of Washington whose research bridges GI science, urban geography, and critical poverty studies. In the 1990s and 2000s, she and other critical scholars showed that GI systems and services are more than digital technologies for storing and representing spatial data. Rather, they are also an array of social and political practices for the production, regulation, and use of geographic information, and a way of knowing and making knowledge. This tradition of social critique of GI technologies emphasizes that geospatial data, geographic technologies, and geovisual representations are nonneutral. They tend to advance the interests of society's most powerful people and places, and as such, geographers and others have an important role to play in developing alternative applications of GI technologies that combat inequality.

Sarah argues in her work that these core critiques can help us better understand the societal implications of the GeoWeb. For instance, looking at the social and political relationships of data creation we see a new world of bottom-up, user-generated and user-modified data, produced and circulated online, transforming the nature of privacy. Personal information from the GeoWeb can be assembled into detailed time–space profiles of individuals without their knowledge or consent. The role of global corporations as GI actors makes it difficult for governments to regulate or even know about these problems. The GeoWeb's new forms of immediate, individualized, and photorealistic representation (e.g., Google Street View) shape what we know or assume about other people and places in ways that are different from the more abstracted and generalized visual language of conventional cartography. Her research shows that activists, disadvantaged social groups, even schoolchildren can use the GeoWeb to create new kinds of geovisual politics, creating multimedia visual and locative representations that rely on humor, parody, witness, and storytelling.

Courtesy: Sarah Elwood

**Figure 10.24** Sarah Elwood, University of Washington.

## Questions for Further Study

1. Design a location-based game based on GPS-enabled mobile phones.
2. Find a selection of geolibraries on the Web and identify their common characteristics. How do each of them allow the user to (a) specify locations of interest, (b) browse through datasets that have similar characteristics, and (c) examine the contents of data sets before acquiring them?
3. To what extent do citizens have a right to locational privacy? What laws and regulations control the use of locational data on individuals?
4. How many computers are there currently in your home, and how many items in your home have RFID tags? If you have a mobile phone, is it GPS-enabled? Do you have other GPS devices in the home?

## Further Reading


Fu, P. and Sun, J. 2011. *Web GIS: Principles and Applications*. Redlands, CA: Esri Press.

National Research Council. 1999. *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press.

National Research Council. 2007. *Successful Response Starts with a Map: Improving Geospatial Support for Disaster Management*. Washington, DC: National Academy Press.

Scharl, A. and Tochtermann, K. (eds.). 2007. *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 Are Shaping the Network Society*. Berlin: Springer.

Sui, D. Z., Elwood, S., and Goodchild, M. F. (eds.). 2013. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice.* New York: Springer.