International Journal of Spatial Data Infrastructures Research, 2007, Vol. 2, 24-32.

Editorial:

[1]Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0

Michael F. Goodchild
National Center for Geographic Information and Analysis, and Department of Geography, University of California, Santa Barbara, CA 93106-4060, USA.
good@geog.ucsb.edu

1. INTRODUCTION

Much progress has been made in the past two decades, and increasingly since the popularizing of the Internet and the advent of the Web, in exploiting new technologies in support of the dissemination of geographic information. Data warehouses, spatial data libraries, and geoportals have proliferated, and today's users of geographic information have a wealth of potential sources that can be searched for suitable data sets. Standards have been established, issues of syntactic interoperability have been largely addressed, and rich descriptions are available in metadata to allow the suitability of a given data set to be assessed. Table digitizers used to be an essential asset for any spatial data center in the days when most sources of geographic information were in the form of paper maps, and skill in digitizing was a major part of any introduction to geographic information systems (GIS). Today, however, users rely heavily on digital sources, and virtually all digitizing is heads-up on-screen.

Despite this picture of progress, however, there remain many issues in the development of spatial data infrastructures, and these of course form the primary rationale for this journal and for projects such as INSPIRE (http://inspire.jrc.it). Surveys by Masser (1998, 2007), Burrough and Masser (1998), Onsrud (2007), and Goodchild, Fu, and Rich (2007) have addressed the thorny problems of semantic interoperability, the economics of spatial data sharing, and legal issues, among others. What is largely missing from these discussions, however, is a concern for the basic supply of geographic information, and trends affecting the processes by which it is acquired and compiled. The Global Positioning System (GPS) has revolutionized the processes of surveying, allowing the rapid and accurate determination of absolute position on the Earth's surface, and remote

sensing provides a massive and constant flow of Earth imagery. But many other types of geographic information are not visible from above, or cannot be extracted from imagery by any automated process. They include the names humans attach to features, otherwise known as geonames or gazetteer entries; environmental information, including measures of air quality; cultural information, including information on the use of land and buildings; and population information, including population density and socioeconomic measures.

Some years ago Estes and Mooneyhan (1994) called attention to what they termed the *mapping myth* – the mistaken belief that the world is well mapped, and that maps are constantly being updated and becoming more accurate. In reality, they argue, topographic mapping as a governmentally sponsored activity peaked in most countries in the middle of the past century, and has been declining ever since. Programs of map updating are seriously lagging in many countries, including the U.S., and few if any efforts exist to improve the levels of detail of existing maps. Mapping is costly and labor-intensive, with the exception of those themes that can be readily gathered through automated means, and governments are less and less willing to absorb the costs. Exceptions exist, of course, especially in the U.K. where the actions of central government appear to have placed the Ordnance Survey on a firm financial basis, at least for now.

Recent commentaries, including several in the popular press (see, for example, Ratcliff, 2007; Helft, 2007), have suggested that a new, third option might effectively supplement the traditional efforts of mapping agencies and the power of remote sensing, filling an obvious gap in the current processes of acquiring geographic information. No consensus yet exists on what to call this third option, which combines elements of Web 2.0 (Scharl and Tochtermann, 2007), collective intelligence (Smith, 1994), and neogeography (Turner, 2006). Calling it volunteered geographic information (VGI) captures what is perhaps its most important aspect, and will be used in the following sections, which address some of the social and research issues it raises.

2. SIX BILLION SENSORS

The Earth's surface is currently occupied by more than six billion humans. Each human being begins acquiring geographic knowledge at an early age, and by adulthood has constructed elaborate mental understanding of the areas where he or she lives and works, as well as of areas that may have been visited or learned about. Such knowledge includes placenames, topographic features, and transport networks – indeed many of the themes that are so difficult to acquire by automated means. The knowledge will have been acquired through up to five functioning senses, augmented by books, magazines, television, and the Internet. Indeed, one might think of humanity as a large collection of intelligent,

mobile sensors, equipped with abilities to interpret and integrate that range from the rudimentary in the case of young children to the highly developed skills of field scientists. These abilities can be augmented with devices that collect other geographic information, from cellphones enabled with GPS, vehicles that track position, digital cameras, or sensors that monitor atmospheric pollution and are carried on the body. Specialists may be trained to observe particular types of geographic information, as for example when surveyors collect information on position, maintenance workers for a utility company collect information on the condition of distributed assets, or soldiers in the field collect information on artillery damage or the enemy's current positions. In summary, the six billion humans constantly moving about the planet collectively possess an incredibly rich store of knowledge about the surface of the Earth and its properties.

Historically, only a very small proportion of this information has been tapped for map-making. Professionals working for mapping agencies rely on interviews with local residents for verification, particularly of placenames, and statistical agencies often interview residents about socioeconomic variables. But with some exceptions, only a very small fraction of human knowledge of the planet makes its way through the various processes used to acquire, assemble, and disseminate geographic information.

Several problems have blocked the effective use of such sources in the past. First, acquisition of many types of geographic information is believed to require training, and thus to be beyond the abilities of amateurs. This is a comparatively recent development in the history of science, since few of the great scientists of the past – Darwin, Kelvin, von Humboldt, for example – had advanced degrees or other credentials that would have given their observations authority (and see Waller, 2002 for a commentary on the reliability of many such observations). Networks of amateur observers play important roles in the collection of weather data (see, for example, the GLOBE program, http://www.globe.gov) and in the Christmas Bird Count (http://www.audubon.org/bird/cbc/), but in both cases it is membership in or sponsorship by organizations that succeeds in conveying some degree of trust. The term *citizen science* is often used in such contexts. Nevertheless, many of the basic observations of topographic mapping are scarcely sophisticated, involving little more than the ability to identify placenames or street addresses, and to classify geographic features by type. Moreover the geographic expertise of a local resident seems inherently different from amateur expertise in other domains of knowledge. Drivers routinely trust driving directions given by local residents, for example, treating them in effect as professionals rather than amateurs.

Second, in the past there has been a general lack of the mechanisms by which such information might be communicated, assembled, integrated, and interpreted. It was possible for the 19th Century compilers of the Oxford English

Dictionary to rely on enormous amounts of written communication by mail with individual contributors (Winchester, 1998). But a similar process would clearly be impossible with respect to the vast numbers of local observers who would be needed to collect detailed geographic information, and not surprisingly there is little evidence of efforts to make such a process work in the past. Instead communication has mostly been with local governments and agencies, who are able to provide detailed geographic information that can be verified and integrated at the national level. Even so, the number of links needed with local agencies can be daunting, since there are on the order of 10^5 local government agencies in the U.S. alone.

Trust is clearly an important issue in any large-scale program for information acquisition. Companies selling digital street maps often rely on local observers, some of them equipped with GPS tracking, to acquire information on recent developments, but here it is the contractual relationship implied by compensated employment that provides the quality assurance, as it would if street-map data were acquired by tracking the vehicles of a package delivery service.

3. WEB 2.0

At the core of this third option is the collection of services known somewhat vaguely as Web 2.0. Whereas the early Web was primarily one-directional, allowing a large number of users to view the contents of a comparatively small number of sites, the new Web 2.0 is a bi-directional collaboration in which users are able to interact with and provide information to central sites, and to see that information collated and made available to others. Wikipedia (http://www.wikipedia.org; Dee, 2007) provides a compelling and well-known example, in which individuals are able to provide the contents of a vast encyclopedia that is managed by a comparatively small group of reviewers and administrators. Such services typically provide for extensive loosely structured metadata; in the case of Wikipedia, for example, users are able to access the complete history of any entry, including all previous versions and edits.

The issues provoked by Web 2.0 services are immediately evident when one compares Wikipedia with traditional mechanisms for compiling encyclopedias. Instead of an elaborate administrative structure that recruits a number of contributors, waits for their inputs, compensates them, and edits, compiles, and prints the results, a process that often can last for years, Wikipedia is assembled continuously, contributions appearing instantaneously. Contributors are entirely volunteers, and in many cases without any professional qualifications. Errors are often caught by users, or by reviewers, and result in edits. But Wikipedia lacks the authority conveyed by a recognized publisher, by an extensive process of review and edit by experts, and by the qualifications of its contributors. Moreover

the Internet and Web have a long history of subversive behavior, which has undoubtedly led to deliberate misinformation making its way into Wikipedia on occasion.

Many examples already exist of Web 2.0 services designed to acquire, assemble, and publish geographic information. Wikimapia (http://www.wikimapia.org) is a service operating on similar lines to Wikipedia, allowing citizens to provide descriptions of places of interest to them, along with geographic coordinates. Each entry is comprised of a rectangle aligned with latitude and longitude, together with a text description. At time of writing there were 4.2 million entries, including descriptions of most of the buildings on the campus of the University of California, Santa Barbara, along with numerous places of interest in the surrounding area. Entries are vetted, again by a group of volunteers, and must meet a number of criteria. 4.2 million is an interesting number in this context, because it is roughly the size of the world's largest gazetteers, which are lists of recognized placenames with geographic locations. For example, the Alexandria Digital Library gazetteer (middleware.alexandria.ucsb.edu/client/gaz/adl/index.jsp) is approximately of this size, having been compiled from various official US Government sources. Traditional gazetteer entries are highly structured, consisting of triples of the form <name, location, type> (Hill, 2006), and using a controlled vocabulary to define types. By contrast, Wikimapia is a volunteered gazetteer, produced entirely by individual citizens, and potentially providing much richer descriptions of places that may include hyperlinks. Other sites in this genre include Flickr (www.flickr.com), with its collection of over 21 million (at time of writing) geo-referenced photographs; and the increasing proportion of entries in Wikipedia that have been geo-referenced.

At a higher level of sophistication are projects in which volunteers contribute substantial technical content. For example, OpenStreetMap (www.openstreetmap.org) is building a public-domain street map of the entire world through volunteer effort. Each contributor develops a map of his or her local streets using GPS tracking; and individual contributions are assembled and reconciled into a single patchwork. Extensive metadata is incorporated, since each piece of the patchwork may have different levels of accuracy and may have been acquired at different dates. Some level of expertise is required in the use of GIS and the project's software, in the basic principles of geographic measurement, and in the project's system for classifying streets. In a similar vein Inrix's Dust Network is tracking some 500,000 vehicles on U.S. highways to provide real-time data on congestion (http://www.inrix.com).

At a third level of sophistication are those services that allow contributors to make their own comparatively complex information available to others within easy-to-use Web 2.0 environments. Google Earth is perhaps the best-known of

these services, its client software having been downloaded more than 100 million times since its first release in 2005. Google Earth's Application Program Interface (API) allows any user to create and publish new content, in the form of layers that can be viewed over the Google Earth imagery base, or *mashed* with it. Tens of thousands of sources, many of them developed by citizens with no prior experience in geographic information technologies, have taken advantage of this mechanism in recent months, so that today it is possible to find on the Web overlays depicting all of the places found in the life and novels of Jane Austen (bbs.keyhole.com/ubb/showflat.php/Cat/0/Number/411188/an/0/page/0), historic maps of many areas of the world (many maps from the David Rumsey collection, www.davidrumsey.com, are available in Google Earth's Featured Content), the campaigns of Alexander the Great (bbs.keyhole.com/ubb/download.php?Number=126402), three-dimensional representations of the buildings of central London (bbs.keyhole.com/ubb/download.php?Number=420893), and the subway system of many cities (bbs.keyhole.com/ubb/showthreaded.php/Cat/0/Number/579229/page/vc/vc/1). All of these are viewable using the intuitive interface of Google Earth. All of them have been volunteered, in many cases by citizens with purely altruistic motives.

Surveys and summaries of the field of VGI are still few and far between. The Where 2.0 conference series (conferences.oreillynet.com/where) has become an excellent forum for what is increasingly termed neogeography. A book by Turner (2006) provides an overview, and more specialized books can be found, for example on using such services to produce novel maps (Erle, Gibson, and Walsh, 2005). Numerous sites such as Google Earth Hacks (www.gearthhacks.com) provide useful information on novel applications. A more academic review has recently been edited by Scharl and Tochtermann (2007), and several others are in preparation.

4. ASSESSMENT

The worlds of VGI and the traditional mapping agencies could not be more different. The latter represent the top-down, authoritarian, centrist paradigm that has existed for centuries, in which professional experts produce, dissemination is radial, and amateurs consume. Expertise in this world is measured with objective indicators such as advanced degrees; progress requires consensus and is therefore slow and deliberate; and costs rise steadily. The world of VGI is chaotic, with little in the way of formal structures. Information is constantly being created and cross-referenced, and flows in all directions, since producers and consumers are no longer distinguishable. Timescales are enormously compressed, and a site such as Wikimapia can go from zero to millions of entries in a matter of months. What is perhaps most surprising about the world of VGI is

the fact that tens of thousands of citizens are willing to spend large amounts of time contributing, without any hope of financial reward, and often without any assurance that anyone will ever make use of their contributions. The same kind of motivation drives the world of blogs, and is one of the most interesting of the many new kinds of social behavior that have emerged with the rise of the Internet.

Like any large-scale activity, VGI is having its own effects on geospatial standards. KML, the language of the Google Earth API, is now the subject of an agreement with the Open Geospatial Consortium, which hopes to adopt it as a standard for communication with virtual globes (http://www.opengeospatial.org/pressroom/newsletters/200701/). Google Earth's imagery is increasingly used as a framework layer, since it makes it so easy to determine latitude and longitude of any recognizable feature. But in areas of Santa Barbara at time of writing the Google Earth imagery was variously misregistered by 20m to the east and 40m to the west, and an entire strip 60m wide was missing around longitude 119 degrees 45 minutes 17 seconds West (e.g., see the discontinuity in U.S. Highway 101 around latitude 34 degrees 26 minutes 21 seconds North). Should Google decide to improve the registration of its imagery the effect would be comparable to that of shifting the North American Datum from NAD 27 to NAD 83 – all features that had been geo-referenced using its imagery would suddenly be in the wrong place, in the sense of being no longer consistent with the imagery.

Every human is able to act as an intelligent sensor, perhaps equipped with such simple aids as GPS or even the means of taking measurements of environmental variables. The notion that citizens might be useful and effective sources of scientifically rigorous observations has a long history, and it is only recently that the scientific community has come to dismiss amateur observation as a legitimate source. The scientific observers of earlier centuries were certainly amateurs by today's standards, with little formal training in measurement technique, little in the way of theory to frame their observations, and few advanced degrees. Today, practices that are often termed citizen science are widely recognized and respected in some areas. But in general it is only through formal institutional frameworks that the contemporary scientific community is willing to accept volunteered information as reliable.

Nevertheless, it is clear from even the most cursory examination of some of these sources of VGI that most contributors are well-meaning, and that the vast majority of the information they provide is of useful quality. Moreover, such volunteering appears to provide the only feasible solution to what is in reality a dramatic decline in the supply of geographic information worldwide. Despite massive investment, remote sensing provides only a partial solution to this problem, since many attributes, including placenames, cannot be seen from

above. But 6 billion citizen observers, equipped with the means to upload their observations, could provide a very effective replacement. The willingness to do so is clearly there, as is the technology to integrate their inputs. But largely missing at this point are the mechanisms needed to ensure quality, to detect and remove errors, and to build the same level of trust and assurance that national mapping agencies have traditionally enjoyed.

REFERENCES

Burrough, P.A. and I. Masser, editors, 1998. *European Geographic Information Infrastructures: Opportunities and Pitfalls.* London: Taylor and Francis.

Dee, J., 2007. All the news that's fit to print out. *New York Times Magazine*, July 1.

Erle, S., R. Gibson, and J. Walsh, 2005. *Mapping Hacks: Tips and Tools for Electronic Cartography*. Sebastopol, CA: O'Reilly Media.

Estes, J.E. and W. Mooneyhan, 1994. Of maps and myths. *Photogrammetric Engineering and Remote Sensing* **60**(5): 517-524.

Goodchild, M.F., P. Fu, and P. Rich, 2007. Sharing geographic information: an assessment of the Geospatial One-Stop. *Annals of the Association of American Geographers* **97**(2): 249-265.

Helft, M., 2007. Mapmaking for the masses, online. *International Herald Tribune*, July 26.

Hill, L.L., 2006. *Georeferencing: The Geographical Associations of Information*. Cambridge, MA: MIT Press.

Masser, I., 1998. *Governments and Geographic Information*. London: Taylor and Francis.

Masser, I., 2007. *Building European Spatial Data Infrastructures*. Redlands, CA: ESRI Press.

Onsrud, H.J., 2007. *Research and Theory in Advanced Spatial Data Infrastructure Concepts.* Redlands, CA: ESRI Press.

Ratcliff, E., 2007. Google Maps is changing the way we see the world. *Wired* 15.06, http://www.wired.com/techbiz/it/magazine/15-07/ff_maps

Scharl, A. and K. Tochtermann, editors, 2007. *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society*. London: Springer.

Smith, J.B., 1994. *Collective Intelligence in Computer-Based Collaboration*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Turner, A., 2006. *Introduction to Neogeography.* Sebastopol, CA: O'Reilly Media.

Waller, J., 2002. *Fabulous Science: Fact and Fiction in the History of Scientific Discovery*. New York: Oxford University Press.

Winchester, S., 1998. *The Professor and the Madman: A Tale of Murder, Insanity, and the Making of the Oxford English Dictionary*. New York: HarperCollins.