

Citizens as sensors: the world of volunteered geography

Michael F. Goodchild

Published online: 20 November 2007
© Springer Science+Business Media B.V. 2007

Abstract In recent months there has been an explosion of interest in using the Web to create, assemble, and disseminate geographic information provided voluntarily by individuals. Sites such as Wikimapia and OpenStreetMap are empowering citizens to create a global patchwork of geographic information, while Google Earth and other virtual globes are encouraging volunteers to develop interesting applications using their own data. I review this phenomenon, and **examine associated issues: what drives people to do this, how accurate are the results, will they threaten individual privacy, and how can they augment more conventional sources?** I compare this new phenomenon to more traditional citizen science and the role of the amateur in geographic observation.

Keywords Geographic information · Web 2.0 · Virtual globe · Privacy · Citizen science

Introduction

In 1507 in St-Dié-des-Vosges, **Martin Waldseemüller drew an outline of a new continent and labeled it America** (Fig. 1). It appears that he was influenced by new books being circulated in Europe at the time, and particularly by the Soderini Letter and its purported author Amerigo Vespucci, and the latter's claims to the continent's discovery. Although Waldseemüller withdrew the name on a later map, and although many scholars and a new biography by Felipe Fernández-Armesto (2006) cast doubt on the authenticity of the Letter, the feminine form of **Vespucci's first name stuck, and was eventually adopted as the authoritative name of not one but two continents.**

By today's standards this act of naming by an obscure cartographer would attract little or no attention. **Modern naming in developed countries is closely regulated by a hierarchy of committees** that in the U.S. extend from the local to the national level (Monmonier 2006). The Board on Geographic Names was established in 1890 for the purpose of standardizing the use of names within the federal government, and thus within the national mapping agencies. In English the term *gazetteer*, the common term for a placename index, is itself rooted in official authority, and **renaming of geographic features requires a lengthy process of review, and is virtually impossible** except in special circumstances. Geographic naming has been centralized and standardized, and assigns no role to obscure individuals like Waldseemüller, who

M. F. Goodchild
National Center for Geographic Information and Analysis,
University of California, Santa Barbara, Santa Barbara,
CA 93106-4060, USA

M. F. Goodchild (✉)
Department of Geography, University of California, Santa
Barbara, Santa Barbara, CA 93106-4060, USA
e-mail: good@geog.ucsb.edu

Fig. 1 An extract from the Waldseemüller map of 1507



would certainly be amazed to learn that his map was recently acquired by the U.S. Library of Congress for \$10 million.

Nevertheless, the events of 1507 provide an early echo of a remarkable phenomenon that has become evident in recent months: the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information, a function that for centuries has been reserved to official agencies. They are largely untrained and their actions are almost always voluntary, and the results may or may not be accurate. But collectively, they represent a dramatic innovation that will certainly have profound impacts on geographic information systems (GIS) and more generally on the discipline of geography and its relationship to the general public. I term this *volunteered geographic information (VGI)*, a special case of the more general Web phenomenon of *user-generated content*, and it is the subject of this paper.

The evolving world of VGI

One of the more compelling examples of VGI is *Wikimapia*, which adapts some of the procedures that have been so successful in the creation of the *Wikipedia* encyclopedia and applies them to the creation of a gazetteer. Anyone with an Internet connection can select an area on the Earth's surface and provide it with a description, including links to other sources. Anyone can edit entries, and volunteer

reviewers monitor the results, checking for accuracy and significance. At time of writing *Wikimapia* had 4.8 million entries compared to *Wikipedia*'s 7 million, describing features ranging in size from entire cities to individual buildings (each entry's geographic extent is defined by ranges of latitude and longitude). Some descriptions are extensive and include hyperlinks; for example, the entry for *Madinah* (Saudi Arabia) includes a picture of the *Masjid-e-Nabawi* and a link to the city's *Wikipedia* entry. Other entries describe features within the city (Fig. 2) or in the surrounding area.

Similar in some respects is the *Flickr* site, which allows users to upload and locate photographs on the Earth's surface by latitude and longitude. At time of writing roughly 2.8 million photographs were being contributed each month to the site. Figure 3 shows one of the more than 2,500 volunteered photographs of *Uluru* (Ayer's Rock) in central Australia.

At a rather different level of sophistication is *MissPronouncer*, a site created by Jackie Johnson to help people pronounce some of the more distinctive Wisconsin placenames. A full-time radio broadcaster, Ms Johnson developed the site in her spare time, and offers audio recordings of the correct pronunciation of almost 2,000 places in the state. Phonic representations of placenames have the advantage that they are not subject to problems over differences of alphabet (Beijing versus 北京, Baghdad versus دادغوب), though the phonic rendering of common placenames may vary from one language to another (e.g., Paris, Moscow).

Fig. 2 The Wikimapia coverage of Madinah, Saudi Arabia. Each box denotes the availability of information describing the feature outlined by the box

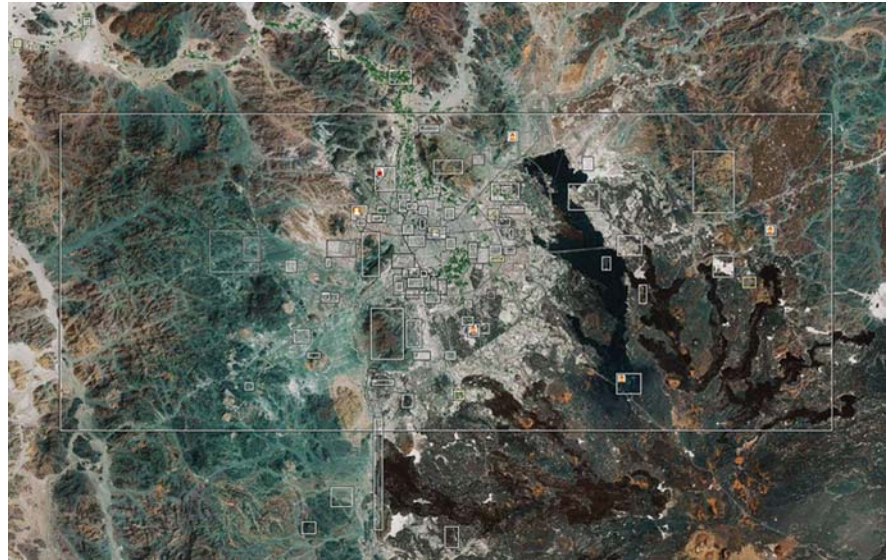
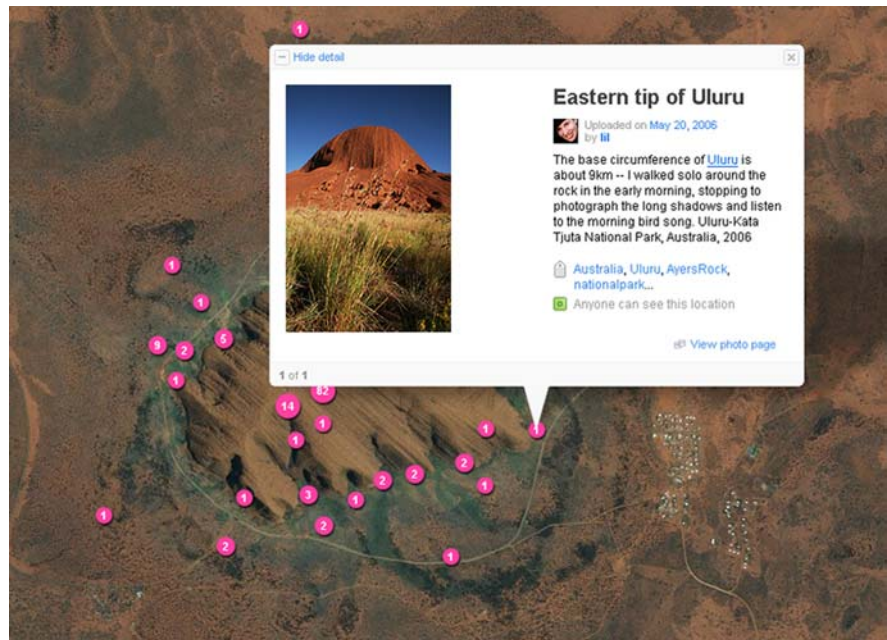


Fig. 3 Information from the Flickr site for the area of Uluru (Ayer's Rock) in central Australia. Each symbol denotes the availability of a photograph; at time of writing there were more than 2,500 available for the area shown. Descriptive information is displayed for one such entry

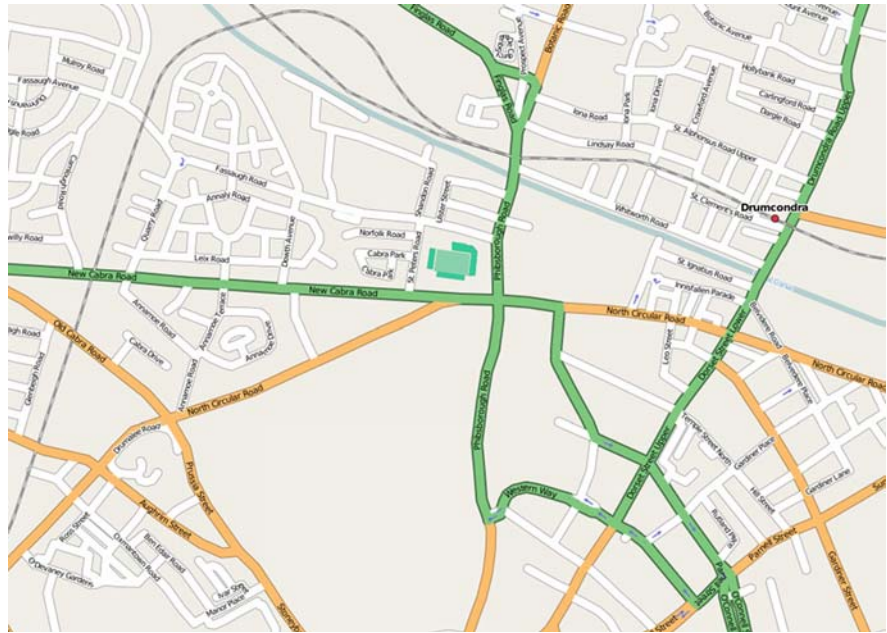


Other VGI activities focus on the creation of more elaborate representations of the Earth's surface. **OpenStreetMap** is an international effort to create a free source of map data through volunteer effort. Figure 4 shows the map for part of Dublin at time of writing. Note the incomplete nature of the map, with major streets, railways, and parks shown but with minor street detail in some areas but not others, and some streets named but not others. Dublin famously lacks a cheap, readily available digital street map, as

do many other cities around the world, so this volunteer effort can potentially fill a yawning gap in the availability of digital geographic information.

When Google acquired the software previously known as Earthviewer, rebranded it, redesigned the user interface, and published an application program interface, it created a service that had immediate appeal to millions. I have described the **Google Earth** phenomenon as the "democratization of GIS" (Butler 2006), because it has opened some of the more

Fig. 4 Part of the OpenStreetMap coverage of Dublin, Ireland. Note the missing street names and areas where no coverage is available



straightforward capabilities of GIS to the general public. Whereas the creation of a “fly-by” was previously one of the more sophisticated GIS tasks, it is now possible for a child of ten to create one in ten minutes. Google Earth and Google Maps popularized the term *mash-up*, the ability to superimpose geographic information from sources distributed over the Web, many of them created by amateurs. For example, Fig. 5 shows a Google Earth mash-up of the Soho area of London during the 1854 cholera outbreak made famous by Dr. John Snow (Johnson 2006). It combines a street map of London from 1843 (from the on-line private collection of David Rumsey, a San Francisco map collector) with on-line data on the water sources and cholera deaths from my own Web site. Readily available software makes this kind of mash-up remarkably easy (see, for example, Brown 2006) and well within the capabilities of the general public. As a result, the number of available mash-ups has reportedly reached the hundreds of thousands, and the number of downloads of the Google Earth software exceeds a 100 million.

These are just a few examples of a phenomenon that has taken the world of geographic information by storm and has the potential to redefine the traditional roles of mapping agencies and companies. In the next section I examine some of the technologies that have combined to make this possible. This is then followed

by a discussion of relevant concepts and issues, and then by an analysis of the usefulness of VGI.

Enabling technologies

Web 2.0

To understand VGI, we must first ask about the technologies that make it possible. Early concepts of the Web stressed the ability of users to access remote sites through simple interfaces known as *browsers* (Mosaic, launched in 1992, was the first widely available browser). One could *surf* the Web by following *hyperlinks*, typically highlighted words that when *clicked* would initiate a download from another page or site. Web pages consisted primarily of text, but graphic images could also be included, taking advantage of the recently expanded graphics capabilities of personal computers. In all of this, however, the relationship between user (*client*) and Web page (located on a *server*) was essentially one-way; the user's only role was to initiate the downloading of content.

In time it became possible for the user's role to extend somewhat. Protocols were developed that allowed users to access information stored in a server's databases, and even to add records to such

Fig. 5 A Google Earth mash-up of the area of Soho, London. The contemporary imagery base has been obscured by an 1843 map from the David Rumsey collection. Superimposed on this are the **deaths (green)** from cholera in the outbreak of 1854, and the **water sources (red)**. Dr John Snow showed that the central source was responsible for the outbreak



databases by completing forms. Airline reservation sites (e.g., Expedia), eBay, and Craig's List all exploit this capability. By the early 2000s this ability of users to supply content to Web sites had grown in sophistication to the point where it became possible to construct sites that were almost entirely populated by user-generated content, with very little moderation or control by the site's owners and very little restriction on the nature of content. In some cases users could even edit the content created by others. Blogs and Wikis fall into this category, as do the sites reviewed in the previous section. Collectively, they have been termed Web 2.0. First and foremost, then, VGI is a result of the growing range of interactions enabled by the evolving Web.

Georeferencing

GIS relies on the ability to specify location on the Earth's surface using a small number of well-defined

and interoperable systems, of which latitude and longitude is by far the most universal. Most countries have some form of national grid that provides an alternative local coordinate system, and the Universal Transverse Mercator (UTM) system has been adopted for the geographic coordinates needed by many military agencies. All of these are specialized, however, and in normal human discourse it is place-names that provide the basis of geographic referencing. Very few people know the latitude and longitude of their home, let alone its UTM coordinates. To enable the creation of geographic data by the general public, therefore, it is necessary to have a range of readily available tools for identifying the coordinates of locations on the Earth's surface.

Several tools now supply this need, and collectively enable VGI. The Global Positioning System (GPS) can be accessed by a wide range of consumer products, allowing location to be measured in many standard coordinate systems. Cameras can be enabled with GPS, so that digital photographs can be

automatically tagged with coordinates. Some GPS receivers store entire tracks that can later be uploaded in digital form, and similar capabilities can be built into mobile phones. Coordinates can also be obtained through a process known as *geocoding*. Any recognized street address can be matched to a digital street file in a service available in most GIS software as well as on the Web.

A technically simpler option is to use the imagery available through Google Earth, Google Maps or similar services to select a location visually, and to record its coordinates by clicking. Several services allow this approach to be used to create digital records of entire streets and other features by following (*digitizing*) the features on the screen; the results are then uploaded and compiled into composite digital maps. OpenStreetMap has already been cited as an example of this approach.

Geotags

A *geotag* is a standardized code that can be inserted into information in order to note its appropriate geographic location. Geotags have been inserted into many Wikipedia entries, when the contents relate to a specific location on the Earth's surface, and several sites allow such entries to be accessed from maps. For example, Fig. 6 shows the result of searching the Geonames site for Wikipedia entries in French in the region of Alsace-Lorraine; clicking on the symbol beside St-Dié-des-Vosges brings up the town's Wikipedia description. At time of writing there were

over 60,000 geotagged entries in the Wikipedia French-language resource alone.

GPS

The Global Positioning System is arguably the first system in human history to allow direct measurement of position on the Earth's surface. GPS receivers are easy to use, and provide virtually instantaneous estimates of location, often to better than 10 m accuracy. Incorporated in in-car navigation systems, GPS allows the current location of the vehicle to be compared to the contents of a digital street map. As a stand-alone device, a receiver is the basis of the popular sport of *geocaching*, which engages participants in finding hidden destinations based only on their coordinates. GPS has sparked a number of interesting VGI activities, such as the creation of maps by walking, cycling, or driving. Figure 7 shows the interesting map created by my colleague Val Noronha, who has installed a GPS in his car to keep track of his daily travels around his neighborhood in Goleta, California. The colors denote his average speed.

Graphics

It is easy to forget that high-quality graphics are a comparatively recent innovation in the history of computing. Dynamic visualization of three-dimensional objects, such as occurs with Google Earth,

Fig. 6 The Geonames site shows the geographic location geotagged in Wikipedia entries, allowing the encyclopedia to be accessed via maps



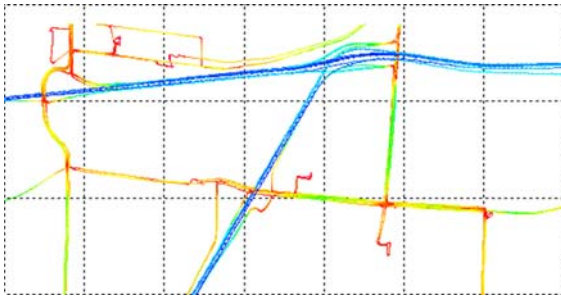


Fig. 7 Average driving speed logged by one car over an extended period around an area of Goleta, California

required a highly sophisticated and expensive computer as recently as 1995, and when *Earthviewer* appeared in 2000 only a few personal computers had the powerful graphics hardware needed to run it. Today, of course, lowly household computers have sufficient power, though devices built for video games, such as Wii, often have even greater power.

Broadband communication

Finally, VGI would be impossible without widespread access to the Internet, preferably via a high-capacity connection. Many households in developed countries now have such broadband connections, using a range of satellite, cable, and phone-line technologies.

Concepts

Spatial data infrastructure patchworks

It is easy to believe that the world is well mapped. Most countries have national mapping agencies that produce and update cartographic representations of their surfaces, and remote-sensing satellites provide regularly updated images. But in reality world mapping has been in decline for several decades (Estes and Mooneyhan 1994). The U.S. Geological Survey no longer attempts to update its maps on a regular basis, and many developing countries no longer sustain national mapping enterprises.

The decline of mapping has many causes (Goodchild et al. 2007). Governments are no longer willing to pay the increasing costs of mapping, and often look

to map users as sources of income. Remote sensing has replaced mapping for many purposes, but satellites are unable to sense many of the phenomena traditionally represented on maps, including the names of places. In the early 1990s the Mapping Science Committee of the U.S. National Research Council issued a report describing the concept of spatial data infrastructure (NRC 1993), which it defined as the aggregate of agencies, technologies, people, and data that together constituted a nation's mapping enterprise.

Among the many concepts introduced in the report was that of *patchwork*, the notion that national mapping agencies should no longer attempt to provide uniform coverage of the entire extent of the country, but instead should provide the standards and protocols under which numerous groups and individuals might create a composite coverage that would vary in scale and currency depending on need. The creation of the National Spatial Data Infrastructure (NSDI) was authorized by President Clinton under Executive Order 12906 in 1994, and has provided the policy umbrella for geographic information in the U.S. for the past 13 years.

VGI clearly fits the model of NSDI. A collection of individuals acting independently, and responding to the needs of local communities, can together create a patchwork coverage. Given a server with appropriate tools, the various pieces of the patchwork can be fitted together, removing any obvious inconsistencies, and distributed over the Web. The accuracy of each piece of the patchwork, and the frequency with which it is updated, can be determined by local need.

Humans as sensors

Recently a great deal of attention has been devoted to the concept of sensor networks. The observational objectives of Earth science, as well as the objectives of security and surveillance, can be addressed at least in part by the installation of networks of sensors across the geographic landscape. Commonly cited examples include the network of video monitors in many major cities, proposals to instrument the ocean and seabed with sensors in the interests of science and early warning of tsunamis, and networks of traffic sensors that can provide useful information to planners, as well as real-time pictures of congestion.

It is useful to distinguish three types of sensor networks. Most examples fit the first, a network of static, inert sensors designed to capture specific measurements of their local environments. Less commonly cited are sensors carried by humans, vehicles, or animals. For example, much useful research is emerging from projects that have equipped children with sensors of air pollution, in an effort to understand the factors affecting asthma. A third type of sensor network, and in many ways the most interesting, consists of humans themselves, each equipped with some working subset of the five senses and with the intelligence to compile and interpret what they sense, and each free to rove the surface of the planet.

This network of human sensors has over 6 billion components, each an intelligent synthesizer and interpreter of local information. One can see VGI as an effective use of this network, enabled by Web 2.0 and the technology of broadband communication.

Citizen science

The term *citizen science* is often used to describe communities or networks of citizens who act as observers in some domain of science. A perfect U.S. example is the Christmas Bird Count, an effort to enlist amateur ornithologists in conducting a mid-winter census of bird populations. Participants require a fairly high level of skill, and over the years a number of protocols have been established to ensure that the resulting data have high quality. An international example is Project GLOBE, an effort to enlist school-children and their teachers in providing a world-wide source of high-quality atmospheric observations. As with the Christmas Bird Count, a number of protocols and training programs have been established to ensure quality, and to collect, synthesize, and re-distribute the results.

Both of these projects require a fair degree of training and expertise. This need for expertise would be a limiting factor in any effort to extend VGI to such comparatively sophisticated mapping themes as land use, land cover, or soil class. Other forms of VGI are much less demanding, however, particularly those associated with place-names, streets, and other well-defined geographic features.

Participant populations

Sites such as Wikimapia are open to all, as are many other VGI efforts. The Christmas Bird Count and Project GLOBE, on the other hand, place restrictions on participation in order to ensure adequate expertise. The question of *who* may volunteer has much to do with the quality of the resulting information, and a range of possibilities exist. For many years companies producing digital street maps have relied on networks of local observers to provide rapid notice of new streets, changes of street names, etc., paying them as part-time workers. Inrix is collecting tracks from hundreds of thousands of trucks and other fleets, processing and compiling the results as a source of real-time information on the state of congestion and other short-term factors affecting travel on road networks. Military personnel are important potential sources of geographic information about local battlefield conditions that can be used to augment what is available from central mapping and imagery sources. Many farmers now have elaborate systems for mapping and monitoring their fields and crops (*precision agriculture*), and constitute a potential source of data that is in many cases much more detailed and current than that available from central agricultural agencies. In essence, such developments contribute to a growing reversal of the traditional *top-down* approach to the creation and dissemination of geographic information.

Early warning

Recent events such as the Indian Ocean tsunami or Hurricane Katrina have drawn attention to the importance of geographic information in all aspects of emergency management, and to the problems that arise in the immediate aftermath of the event before adequate overhead imagery becomes available for damage assessment and response planning (NRC 2007). Earth-observing satellites may not pass over the affected area for several days. Images from satellites and aircraft may be obscured by clouds and smoke. Conditions on the ground may prevent the rapid downloading of digital imagery because of a lack of power, Internet connections, or computer hardware and software.

On the other hand the human population in the affected area is intelligent, familiar with the area, and

increasingly able to report conditions through mobile phones, using voice, text, or pictures. To date there has been very little use of VGI in these situations, in part because of an almost complete lack of the tools needed to collect, synthesize, verify, and redistribute the information. However the potential to obtain almost immediate reports from geographically distributed observers on the ground will surely drive increased efforts to overcome these problems in the next few years.

Issues

Why do people do this?

In the mid 1990s the U.S. Federal Geographic Data Committee published its Content Standards for Digital Geospatial Metadata, a format for the description of geographic data sets. The project was very timely, given the rapid increase in the availability of geographic information via the Internet that occurred at that time. Metadata were seen as the key to effective processes of search, evaluation, and use of geographic information. Nevertheless, and despite numerous efforts and inducements, it remains very difficult to persuade those responsible for creating geographic data sets to provide adequate documentation. Even such a popular service as Google Earth has no way of informing its users of the quality of its various data layers, and it is virtually impossible to determine the date when any part of its image base was obtained. A recent news report concerned the apparent replacement of its coverage of New Orleans with pre-Katrina imagery, though its coverage of the Darfur region is updated almost daily.

Given this evident reluctance to provide documentation, it is perhaps surprising that the opportunity to create and publish VGI has engaged the interests of so many individuals. Why is it that citizens who have no obvious incentive are nevertheless willing to spend large amounts of time creating the content of VGI sites? What kinds of people are more likely to participate, and what drives them to be accurate (or inaccurate)?

Self-promotion is clearly an important motivator of Internet activity, and in its extreme form can lead to the exhibitionism of personal web-cams. Despite the vast resources of the Web, it is still possible to believe

that *someone* will be interested in one's personal site. The popularity of some blogs can be misread as suggesting that an audience exists for *any* blog.

At a different level many users volunteer information to Web 2.0 sites as a convenient way of making it available to friends and relations, irrespective of the fact that it becomes available to all. This may underlie the popularity of sites such as Picasa, which allow contributors of personal photographs to point others to them, but it scarcely explains the popularity of Flickr or Wikimapia, where content is comparatively anonymous. Contributors to OpenStreetMap may derive a certain personal satisfaction from seeing their own contributions appear in the patchwork, and from watching the patchwork grow in coverage and detail, but there can be no question of self-promotion in this essentially anonymous project.

Authority and assertion

The traditional mapping agencies have elaborate standards and specifications to govern the production of geographic information, and employ cartographers with documented qualifications. Over the years their products have acquired an authority that derives from each agency's reputation for quality. Google, on the other hand, has no such reputation in the geographic domain. Nevertheless users appear willing to ascribe authority to its products, perhaps because computerization carries authority per se, and perhaps because of the company's success in other areas, particularly its search engine.

At time of writing Google Earth's imagery over the campus of the University of California, Santa Barbara was mis-registered by approximately 20 m east–west. Further to the east in the City of Santa Barbara the mis-registration was approximately 40 m east–west in the opposite direction, and a swath approximately 60 m wide running north–south was missing from the coverage (Fig. 8). Any locations georeferenced from this imagery and incorporated into VGI will inherit these positional errors, and if Google re-registers the imagery at a future date that VGI will be clearly misplaced. In essence, Google has created a new datum or horizontal reference system that is substantially different from the current North American datum, but which is widely accepted because of the authority of Google. The shift is comparable in



Fig. 8 A swath of Santa Barbara approximately 60 m wide has disappeared because of misregistration of imagery in Google Earth. Note the blurring down the center of the image, and the break in Highway 101, one of the few features to cross the swath diagonally (since this paper was written Google has re-registered its imagery of Santa Barbara, reducing the problem substantially)

magnitude to that created when North American mapping agencies replaced the North American Datum of 1927 (NAD27) with the current NAD83.

VGI is sometimes termed *asserted* geographic information, in that its content is asserted by its creator without citation, reference, or other authority. The early days of the Internet were characterized by a certain altruism, a belief in the essential goodness of users, and there was little anticipation of the subversive phenomena of *spam, viruses, and denial-of-service* attacks that now pervade the network. Similarly many VGI efforts are driven by the kinds of altruism inherent in any kind of voluntary community effort. Can we expect, then, a similar pattern of disillusionment as *antisocial elements recognize and exploit the inevitable vulnerabilities*? Will there be efforts to create fictitious landscapes, or to attack and bring down VGI servers? VGI is currently a somewhat exotic domain, *but if and when users begin to rely on its services a growing pattern of efforts to undermine it seems inevitable.*

The digital divide

Despite the apparent openness of VGI, it remains largely the preserve of those fortunate to have access

to the Internet—and broadband access in particular. While a growing fraction of citizens in developed countries have such access, *it is largely unavailable to the majority of the world's population who live in developing countries.* Moreover *issues of language and alphabet also affect access* even for those with broadband connections, since many VGI servers support only the Roman alphabet and English. In principle, *much could be achieved through mobile phones*, which often have the ability to connect to the Internet and to capture images, but the tools needed to exploit this limited environment as a source for VGI do not yet exist. So while I argued above that such limited tools were potentially significant in early warning and emergency management, significant work still needs to be done to realize the potential.

The value of VGI

As I hope the examples in this paper illustrate, *VGI has the potential* to be a significant source of *geographers' understanding of the surface of the Earth.* It can be timely, a property that was particularly stressed in the discussion of early warning. By motivating individuals to act voluntarily, *it is far cheaper* than any alternative, and its products are *almost invariably available to all* (but see the earlier discussion of the digital divide).

In earlier sections I discussed why people might be motivated to create VGI, but not *why they might want to use it.* With sites such as Wikimapia one can learn a great deal about remote places, acquiring the kinds of information needed for planned tourist visits, or to provide background to travelogs. Sites such as OpenStreetMap *often provide the cheapest source of geographic information, and sometimes the only source, particularly in areas where access to geographic information is regarded as an issue of national security.*

It is already clear in many fields that such informal sources as blogs and VGI can act as very useful sources of military and commercial intelligence. The tools already exist to scan Web text searching for references to geographic places, and to geocode the results. *Thus the most important value of VGI may lie in what it can tell about local activities in various geographic locations that go unnoticed by the world's media,* and about life at a local level. It is in that area

that VGI may offer the most interesting, lasting, and compelling value to geographers.

References

- Brown, M. C. (2006). *Hacking Google Maps and Google Earth*. New York: Wiley.
- Butler, D. (2006). Virtual globes: The web-wide world. *Nature*, 439, 776–778.
- Estes, J. E., & Mooneyhan, W. (1994). Of maps and myths. *Photogrammetric Engineering and Remote Sensing*, 60, 517–524.
- Fernández-Armesto, F. (2006). *Amerigo: The man who gave his name to America*. London: Weidenfeld and Nicolson.
- Goodchild, M. F., Fu, P., & Rich, P. (2007). Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers*, 97(2), 249–265.
- Johnson, S. (2006). *The Ghost Map: The story of London's most terrifying epidemic, and how it changed science, cities, and the modern world*. New York: Riverhead.
- Monmonier, M. (2006). *From squaw tit to whorehouse meadow*. Chicago: University of Chicago Press.
- National Research Council (1993). *Toward a coordinated spatial data infrastructure for the nation*. Washington: National Academies Press.
- National Research Council (2007). *Successful response starts with a map: improving geospatial support for disaster management*. Washington: National Academies Press.