Virtual Geographic Environments as Collective Constructions

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

Abstract:

Web 2.0 technologies are enabling the collective construction of virtual environments. Unlike the imagined worlds of sites such as Second Life, these are attempts at geographic realism, built as patchworks of individually volunteered information. As such they echo the objectives of national spatial data infrastructures, but with amateur individuals rather than professional agencies and corporations as the contributors. I give and assess examples of such projects, review the domains in which they appear most successful, and examine the issues that they raise concerning privacy, standards, verification, sustainability, and benefit.

1. INTRODUCTION

The early Web achieved its success by providing users with easy access to pages located on servers. Pages could contain simple text with embedded tags (html), or images in a range of formats. The resulting formatted displays had immediate appeal, particularly to users with very little in the way of computer experience or skills. Hyperlinks allowed rapid access to related material, and simple and universal standards for content made it possible to support user access with very "thin" clients. Later, functions were implemented to support forms, enabling users to contribute records to databases, as for example with on-line reservation systems, opening the potential for Internet-based commerce. The format constraints on such content were gradually removed, allowing the development of such popular sites as EBay. Most recently, the concept of user-generated content (UGC) has expanded to include wikis, blogs, and a range of other services that make it easy for the general user to contribute unstructured information to servers that in turn assemble, possibly filter, and render content. The popularity of UGC has led to the coining of the term *Web 2.0*, implying a radically revised conceptualization of the nature of the Web as a multi-way enterprise in which clients and servers are no longer as clearly distinguished.

This development fits well with a trend that has been apparent in the world of geographic information (GI), or information about the surface of the Earth, for the past two decades. In most developed countries the process of map-making has long been entrusted to national mapping agencies such as the U.S. Geological Survey and the National Geospatial-Intelligence Agency. Fifty years ago the costs of equipping a map-making enterprise, including field surveyors, photogrammetrists, cartographers, and printers, along with their expensive machines, justified the concentration of such efforts in a single agency funded by the national government. The high cost of map-making also contributed to its emphasis on the relatively static features of the Earth's surface rather than the relatively dynamic. But a series of developments has led to a substantial erosion of this role, the empowerment of lower levels of government and even individuals, and a new set of opportunities to capture dynamics and change. They include the Global Positioning System, which has partially replaced the need for costly surveying equipment; satellite remote sensing, with its abundant supplies of timely, high-resolution imagery; Web mapping and imagery services which make it possible for individuals to generate their own overlays; and a host of sensors of various kinds. By the early 1990s it was possible for many farmers to have better information about their soils and crop yields than the U.S. Department of Agriculture; for companies selling digital street maps to rely on local employees to provide updates; and for military commanders to receive and compile intelligence from personnel in the field to augment and in some cases replace information from central sources.

Moreover, governments have been less and less willing to shoulder the ever-increasing costs of acquiring, compiling, and disseminating geographic data [*]. In some countries they have resorted to various user-pay arrangements, in which the ultimate consumers of geographic information pay at least some of the cost of acquiring it, and more than the cost of reproduction. In other countries governments have simply abandoned the map-making enterprise entirely, or maintained it solely as support for national security and defense. Estes and Mooneyhan [*] have tracked this decline, arguing that mapping as a national government function peaked some decades ago.

In the U.S., the National Research Council's Mapping Science Committee saw the writing on the wall in the early 1990s, and outlined the concept of spatial data infrastructure [*]—the notion that map-making would evolve into a collective but distributed enterprise, involving agencies from the national to the local level, along with the private sector and citizenry. Maps and geographic information in general would be produced to nationally agreed standards, but as a patchwork whose pieces would be driven primarily by local needs, and thus would be updated more frequently, and produced to finer levels of detail, in areas of greatest activity, such as metropolitan areas.

Given a server with appropriate tools, the various pieces of the patchwork could be fitted together, removing any obvious inconsistencies, and distributed over the Web; and the quality and lineage of each piece of the patchwork could be documented in metadata. 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. Since then extensive activities have been initiated under the Federal Geographic Data Committee, data clearinghouses and geo-portals have been established, and numerous standards have been published. The Open Geospatial Consortium, a group of industrial, academic, and government partners, has been particularly instrumental in promoting interoperability and real-time access to data.

These developments made it possible at least partially to reverse the traditional radial, top-down flow of geographic information. Rather than a single national agency producing much of the nation's supply, it is now local governments, commercial firms, and even individuals who are most engaged in the production, as well as the consumption, of the NSDI. Spatial data infrastructure efforts have been initiated in almost all developed countries, often with the assistance of the Global Spatial Data Infrastructure community. Software for exploiting these resources is now commonly available, in the form of Web services as well as the more traditional stand-alone GIS.

This picture has recently been reinforced by a spate of efforts that fall under the general rubric of UGC and Web 2.0, and are directly related to the supply and consumption of geographic information. Among the hundreds of such sites are Wikimapia, a service that builds on the success of Wikipedia and attempts to "describe the Earth" by allowing contributors to post descriptions of any user-selected area, from an entire continent to a single house. Wikimapia attempts to mimic Wikipedia in its quality-control mechanisms: individuals can edit each other's contributions, and entries are reviewed by a hierarchy of editors. There are now some 5.5 million entries in Wikimapia. The Flickr site allows individuals to post and geo-reference photographs, and to add information to them that may include links to other sources, such as Wikipedia; it now holds on the order of 10^8 photographs, including over 2,000 of the Australian Uluru or Ayer's Rock alone. Somewhat more ambitious is OpenStreetMap, an effort to build a free, public digital map of the globe at sufficient spatial resolution to act as a way-finding aid—a remarkable effort particularly in areas of the globe where no such digital maps exist, or where they are available only at exorbitant cost.

Together, these sites and others like them constitute a remarkable voluntary effort to construct a virtual geographic environment. They rely on the apparent willingness of large numbers of individuals, distributed over the surface of the Earth, to spend significant amounts of their own time creating pieces of the patchwork based on their personal knowledge of local geography. It is particularly remarkable that those same individuals are increasingly reluctant to contribute to other public data gathering exercises, including the census and questionnaire surveys. A number of terms have become associated with the phenomenon in addition to Web 2.0 and UGC. The terms *crowdsourcing* and *collective intelligence* draw attention to the notion that the collective contributions of a number of individuals may be more reliable than those of any one individual. The term *volunteered geographic information* (VGI) refers specifically to geographic information, and to the contrast between the actions of amateurs and those of authoritative agencies, and I will use it here. The term *asserted* geographic information draws attention to the fact that such information is not subject to the normal checks and quality control mechanisms of those agencies, while the term *neo-geography* emphasizes the contrast between this grass-roots phenomenon and the current state of the academic discipline of geography.

The VGI phenomenon raises significant and far-reaching issues that call out for research. Moreover, the changes outlined in this section have occurred at a pace that far exceeds society's ability to rethink and reorganize its institutions, with the result that current legacy arrangements are already well out of touch with the emerging reality. The remainder of this paper examines the phenomenon, its drivers, and its implications in some detail. Section 2 identifies several basic concepts that underlie and may help to understand the nature of VGI. Section 3 attempts to develop the foundations of a research agenda, spanning issues of individual behavior that are normally considered part of social psychology, and the technical and domain-specific issues that would normally be identified as GIScience.

2. BASIC CONCEPTS

2.1 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 of the commonly cited 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, at least locally. The human senses may be augmented with imaging devices, and information may be uplinked using cellular, satellite, or other communication technology. Web 2.0 provides the necessary means to assemble, integrate, and redistribute this information. However, existing institutional arrangements make it unlikely that the information will ascend up the hierarchy to be incorporated into the authoritative products of traditional mapping agencies.

2.2 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. Recently there has been much interest in new efforts along these lines, enabled by GI acquisition tools such as GPS, and the services of Web 2.0.

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 local place-names, streets, and other well-defined geographic features that are commonly familiar to neighborhood residents.

2.3 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 to contribute geographic information 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 are 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, radial approach to the creation and dissemination of geographic information.

2.4 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 [*]. Earth-observing satellites may not pass over the area affected by the event 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 or Internet connections, using voice, text, or pictures. To date there has been very little use of such novel data sources in these situations, in part because of an almost complete lack of the tools needed to collect, synthesize, verify, and redistribute the information, and in part because of a mistrust of information asserted by volunteers. 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.

3. RESEARCH ISSUES

3.1 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 that became one of the foundations of the NSDI. 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 was being updated almost daily at the time of writing.

Given this evident reluctance to provide documentation, it is perhaps surprising that the opportunity to create UGC 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 Web 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 ones 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.

Another possible motivation stems from the ability of contributors of UGC to access and retrieve their own content. This is in sharp contrast to most government data-collection efforts, which result in personal data disappearing into the maw of government agencies, to reappear only in aggregated form. Contributors of UGC can check, edit, and monitor the information they provided, and thus feel a sense of personal ownership over the collective enterprise.

3.2 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. Yet there is ample evidence that the quality of Google's geographic products falls far short of those of the traditional mapping agencies. Google's imagery is often significantly mis-registered, and no information is provided on the quality, provenance, or semantics of the imagery or ancillary data. Thus services such as Google Earth fall far short of scientific norms, including replicability.

UGC is sometimes termed *asserted* 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 UGC efforts are driven by the kinds of altruism inherent in any 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 Web 2.0 servers? UGC 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.

3.3 The digital divide

Despite the apparent openness of UGC, 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, either as users or as contributors. Moreover issues of language and alphabet also affect access even for those with broadband connections, since many Web 2.0 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 UGC do not yet exist except in limited and isolated form. So while it was argued above that such limited tools were potentially significant in early warning, military intelligence, and emergency management, significant work still needs to be done to realize the potential.

3.4 Applications

Geographic information compiled as UGC has the potential to be a significant source of 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). With sites such as Wikimapia one can learn a great deal about remote places, acquiring the kinds of information needed for planned activities on the

ground. 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 mechanisms as blogs and UGC 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 UGC 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 UGC may offer the most interesting, lasting, and compelling value.

4. CONCLUSION

One of the greatest impediments to the construction of virtual environments has been lack of data. While parts of the world are now well endowed with digital geographic information, particularly of a topographic nature, and while satellites now provide abundant imagery that is often of high spatial resolution, the general state of world mapping is in decline. Moreover the richest legacy of geographic information describes phenomena that are relatively static, and describes only their two-dimensional form. At this point there seems little prospect of this situation improving, at least as far as traditional, authoritative production mechanisms are concerned. Only in those areas dominated by the private sector, such as the production of digital street centerlines, is there much evidence of improvement. A recent study estimated that a full three-dimensional representation of all built structures on Earth would amount to some 3 petabytes of digital information, not in itself an unmanageable amount even by today's standards—but that its acquisition, compilation, and quality assurance would consume approximately 10% of U.S. gross domestic product for 10 years given existing, intensively manual techniques.

By contrast, I have argued that the concept of VGI, and the mechanisms that support its production, provide an interesting and compelling alternative. It is local communities, not national mapping agencies, that are the ultimate custodians of geographic knowledge, and the evidence suggests that they are willing, under the right circumstances, to provide it in forms that can be readily compiled and distributed. Modernist ideas of central authority and a single truth are increasingly outdated in a world that is growing used to the notion of multiple viewpoints, and to technologies that make it easy to maintain large quantities of potentially inconsistent data. We now have the tools to achieve interoperability between alternative spellings and renderings of placenames and even multiple placenames, and are increasingly coming to expect our information technologies to reflect and support our multicultural, post-modern world. Web 2.0 provides the opportunity to rethink many of the most cherished concepts of geographic information production, and to engage with the richness of individual geographic knowledge.

5. ACKNOWLEDGMENTS

I am grateful to the many students and colleagues who share my interest in VGI, and particularly to the participants in the upcoming NCGIA Specialist Meeting on VGI, which will take place in Santa Barbara December 13–14, 2007. Further details on the meeting, which is funded by the U.S. Army Research Office and Los Alamos National Laboratory, are available at www.ncgia.ucsb.edu/projects/vgi.

6. REFERENCES

[1] J.E. Estes and W. Mooneyhan, 1994. Of maps and myths. *Photogrammetric Engineering and Remote Sensing* 60: 517–524.

[2] M.F. Goodchild, 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.

[3] National Research Council, 1993. *Toward a Coordinated Spatial Data Infrastructure for the Nation*. Washington, DC: National Academy Press.

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